F

Fair Division*

STEVEN J. BRAMS
Department of Politics, New York University,
New York, USA

Article Outline

Glossary

Efficiency An allocation is efficient if there is no other allocation that is better for one player and at least as good for all the other players.

Envy-freeness An allocation is envy-free if each player thinks it receives at least a tied-for-largest portion and so does not envy the portion of any other player.

Equitability An allocation is equitable if each player values the portion that it receives the same as every other player values its portion.

Definition of the Subject

Cutting a cake, dividing up the property in an estate, determining the borders in an international dispute – such allocation problems are ubiquitous. *Fair division* treats all these problems and many more through a rigorous analysis of procedures for allocating goods, or deciding who wins on what issues, in a dispute.

Introduction

The literature on fair division has burgeoned in recent years, with five academic books [1,13,23,28,32] and one popular book [15] providing overviews. In this review, I will give a brief survey of three different literatures: (i) the division of a single heterogeneous good (e. g., a cake with different flavors or toppings); (ii) the division, in whole or part, of several divisible goods; and (iii) the allocation of several indivisible goods. In each case, I assume the different people, called *players*, may have different preferences for the items being divided.

For (i) and (ii), I will describe and illustrate procedures for dividing divisible goods fairly, based on different criteria of fairness. For (iii), I will discuss problems that arise in allocating indivisible goods, illustrating trade-offs that must be made when different criteria of fairness cannot all be satisfied simultaneously.

Single Heterogeneous Good

The metaphor I use for a single heterogeneous good is a cake, with different flavors or toppings, that cannot be cut into pieces that have exactly the same composition. Unlike a sponge or layer cake, different players may like different pieces – even if they have the same physical size – because they are not homogeneous.

Some of the cake-cutting procedures that have been proposed are discrete, whereby players make cuts with a knife – usually in a sequence of steps – but the knife is not allowed to move continuously over the cake. Moving-knife procedures, on the other hand, permit such continuous movement and allow players to call "stop" at any point at which they want to make a cut or mark.

There are now about a dozen procedures for dividing a cake among three players, and two procedures for dividing a cake among four players, such that each player is assured of getting a most valued or tied-for-most-valued piece, and there is an upper bound on the number of cuts that must be made [16]. When a cake is so divided, no

player will envy another player, resulting in an *envy-free division*.

In the literature on cake-cutting, two assumptions are commonly made:

1. The goal of each player is to maximize the minimum-size piece (*maximin piece*) that he or she can guarantee for himself or herself, regardless of what the other players do. To be sure, a player might do better by not following such a *maximin strategy*; this will depend on the strategy choices of the other players. However, all players are assumed to be *risk-averse*: They never choose strategies that might yield them more valued pieces if they entail the possibility of giving them less than their maximin pieces.

2. The preferences of the players over the cake are continuous. Consider a procedure in which a knife moves across a cake from left to right and, at any moment, the piece of the cake to the left of the knife is A and the piece to the right is B. The continuity assumption enables one to use the intermediate-value theorem to say the following: If, for some position of the knife, a player views piece A as being more valued than piece B, and for some other position he or she views piece B as being more valued than piece A, then there must be some intermediate position such that the player values the two pieces exactly the same.

Only two 3-person procedures [2,30], and no 4-person procedure, make an envy-free division with the minimal number of cuts ($n - 1$ cuts if there are n players). A cake so cut ensures that each player gets a single connected piece, which is especially desirable in certain applications (e. g., land division).

For two players, the well-known procedure of "I cut the cake, you choose a piece," or "cut-and-choose," leads to an envy-free division if the players choose maximin strategies. The cutter divides the cake 50-50 in terms of his or her preferences. (Physically, the two pieces may be of different size, but the cutter values them the same.) The chooser takes the piece he or she values more and leaves the other piece for the cutter (or chooses randomly if the two pieces are tied in his or her view). Clearly, these strategies ensure that each player gets at least half the cake, as he or she values it, proving that the division is envy-free.

But this procedure does not satisfy certain other desirable properties [7,22]. For example, if the cake is, say, half vanilla, which the cutter values at 75 percent, and half chocolate, which the chooser values at 75 percent, a "pure" vanilla-chocolate division would be better for the cutter than the divide-and-choose division, which gives him or her exactly 50% percent of the value of the cake.

The moving-knife equivalent of "I cut, you choose" is for a knife to move continuously across the cake, say from left to right. Assume that the cake is cut when one player calls "stop." If each of the players calls "stop" when he or she perceives the knife to be at a 50-50 point, then the first player to call "stop" will produce an envy-free division if he or she gets the left piece and the other player gets the right piece. (If both players call "stop" at the same time, the pieces can be randomly assigned to the two players.)

To be sure, if the player who would truthfully call "stop" first knows the other player's preference and delays calling "stop" until just before the knife would reach the other player's 50-50 point, the first player can obtain a greater-than-50-percent share on the left. However, the possession of such information by the cutter is not generally assumed in justifying cut-and-choose, though it does not undermine an envy-free division.

Surprisingly, to go from two players making one cut to three players making two cuts cannot be done by a discrete procedure if the division is to be envy-free.[1] The 3-person discrete procedure that uses the fewest cuts is one discovered independently by John L. Selfridge and John H. Conway about 1960; it is described in, among other places, Brams and Taylor (1996) and Robertson and Webb (1998) and requires up to five cuts.

Although there is no discrete 4-person envy-free procedure that uses a bounded number of cuts, Brams, Taylor, and Zwicker (1997) and Barbanel and Brams (2004) give moving-knife procedures that require up to 11 and 5 cuts, respectively. The Brams-Taylor-Zwicker (1997) procedure is arguably simpler because it requires fewer simultaneously moving knives. Peterson and Su (2002) give a 4-person envy-free moving-knife procedure for chore division, whereby each player thinks he or she receives the least undesirable chores, that requires up to 16 cuts.

To illustrate ideas, I describe next the Barbanel–Brams [2] 3-person, 2-cut envy-free procedure, which is based on the idea of squeezing a piece by moving two knives simultaneously. The Barbanel–Brams [2] 4-person, 5-cut envy-free procedure also uses this idea, but it is considerably more complicated and will not be described here.

The latter procedure, however, is not as complex as Brams and Taylor's [12] general *n*-person discrete procedure. Their procedure illustrates the price one must pay for an envy-free procedure that works for all *n*, because it places no upper bound on the number of cuts that are required to produce an envy-free division; this is also true of other *n*-person envy-free procedures [25,27]. While the

---

[1][28], pp. 28–29; additional information on the minimum numbers of cuts required to give envy-freeness is given in [19], and [29].

number of cuts needed depends on the players' preferences over the cake, it is worth noting that Su's [31] approximate envy-free procedure uses the minimal number of cuts at a cost of only small departures from envy-freeness.[2]

I next describe the Barbanel–Brams 3-person, 2-cut envy-free procedure, called the *squeezing procedure* [2]. I refer to players by number – player 1, player 2, and so on – calling even-numbered players "he" and odd-numbered players "she." Although cuts are made by two knives in the end, initially one player makes "marks," or virtual cuts, on the line segment defining the cake; these marks may subsequently be changed by another player before the real cuts are made.

Squeezing procedure. A referee moves a knife from left to right across a cake. The players are instructed to call "stop" when the knife reaches the 1/3 point for each. Let the first player to call "stop" be player 1. (If two or three players call "stop" at the same time, randomly choose one.) Have player 1 place a mark at the point where she calls "stop" (the right boundary of piece A in the diagram below), and a second mark to the right that bisects the remainder of the cake (the right boundary of piece B below). Thereby player 1 indicates the two points that, for her, trisect the cake into pieces A, B, and C, which will be assigned after possible modifications.

```
    A       B       C
 /–––––|–––––|–––––/
        1       1
```
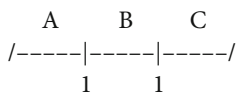
Because neither player 2 nor player 3 called "stop" before player 1 did, each of players 2 and 3 thinks that piece A is at most 1/3. They are then asked whether they prefer piece B or piece C. There are three cases to consider:

1. If players 2 and 3 each prefer a different piece – one player prefers piece B and the other piece C – we are done: Players 1, 2, and 3 can each be assigned a piece that they consider to be at least tied for largest.
2. Assume players 2 and 3 both prefer piece B. A referee places a knife at the right boundary of B and moves it to the left. At the same time, player 1 places a knife at the left boundary of B and moves it to the right in such a way that the value of the cake traversed on the left (by B's knife) and on the right (by the referee's knife) are equal for player 1. Thereby pieces A and C increase

equally in player 1's eyes. At some point, piece B will be diminished sufficiently to a new piece, labeled B′ – in either player 2's or player 3's eyes – to tie with either piece A′ or C′, the enlarged A and C pieces. Assume player 2 is the first, or tied for the first, to call "stop" when this happens; then give player 3 piece B′, which she still thinks is the most valued or the tied-for-most-valued piece. Give player 2 the piece he thinks ties for the most value with piece B′ (say, piece A′), and give player 1 the remaining piece (piece C′), which she thinks ties for the most value with the other enlarged piece (A′). Clearly, each player will think that he or she received at least a tied-for-most-valued piece.

3. Assume players 2 and 3 both prefer piece C. A referee places a knife at the right boundary of B and moves it to the right. Meanwhile, player 1 places a knife at the left boundary of B and moves it to the right in such a way as to maintain the equality, in her view, of pieces A and B as they increase. At some point, piece C will be diminished sufficiently to C′ – in either player 2's or player 3's eyes – to tie with either piece A′ or B′, the enlarged A and B pieces. Assume player 2 is the first, or the tied for the first, to call "stop" when this happens; then give player 3 piece C′, which she still thinks is the most valued or the tied-for-most-valued piece. Give player 2 the piece he thinks ties for the most value with piece C′ (say, piece A′), and give player 1 the remaining piece (piece B′), which she thinks ties for the most value with the other enlarged piece (A′). Clearly, each player will think that he or she received at least a tied-for-most-valued piece.

Note that who moves a knife or knives varies, depending on what stage is reached in the procedure. In the beginning, I assume a referee moves a single knife, and the first player to call "stop" (player 1) then trisects the cake. But, at the next stage of the procedure, in cases (2) and (3), it is a referee and player 1 that move two knives simultaneously, "squeezing" what players 2 and 3 consider to be the most-valued piece until it eventually ties, for one of them, with one of the two other pieces.

Several Divisible Goods

Most disputes – divorce, labor-management, merger-acquisition, and international – involve only two parties, but they frequently involve several homogeneous goods that must be divided, or several issues that must be resolved.[3]

---

[2]See [10,20,26] for other approaches, based on bidding, to the housemates problem discussed in [31]. On approximate solutions to envy-freeness, see [33]. For recent results on pie-cutting, in which radial cuts are made from the center of a pie to divide it into wedge-shaped pieces, see [3,8].

[3]Dividing several homogeneous goods is very different from cake-cutting. Cake-cutting is most applicable to a problem like land division, in which hills, dales, ponds, and trees form an incongruous mix, making it impossible to give all or one thing (e. g., trees) to one play-

As an example of the latter, consider an executive negotiating an employment contract with a company. The issues before them are (1) bonus on signing, (2) salary, (3) stock options, (4) title and responsibilities, (5) performance incentives, and (6) severance pay [14].

The procedure I describe next, called *adjusted winner* (AW), is a 2-player procedure that has been applied to disputes ranging from interpersonal to international ([15]).[4] It works as follows. Two parties in a dispute, after perhaps long and arduous bargaining, reach agreement on (i) what issues need to be settled and (ii) what winning and losing means for each side on each issue. For example, if the executive wins on the bonus, it will presumably be some amount that the company considers too high but, nonetheless, is willing to pay. On the other hand, if the executive loses on the bonus, the reverse will hold.

Thus, instead of trying to negotiate a specific compromise on the bonus, the company and the executive negotiate upper and lower bounds, the lower one favoring the company and the upper one favoring the executive. The same holds true on other issues being decided, including non-monetary ones like title and responsibilities.

Under AW, each side will always win on some issues. Moreover, the procedure guarantees that both the company and the executive will get at least 50% of what they desire, and often considerably more.

To implement AW, each side secretly distributes 100 points across the issues in the dispute according to the importance it attaches to winning on each. For example, suppose that the company and the executive distribute their points as follows, illustrating that the company cares more about the bonus than the executive (it would be a bad precedent for it to go too high), whereas the reverse is true for severance pay (the executive wants to have a cushion in the event of being fired):

	Issues	Company	Executive
1.	Bonus	*10*	5
2.	Salary	35	*40*
3.	Stock Options	15	*20*
4.	Title and Responsibilities	*15*	10
5.	Performance Incentives	*15*	5
6.	Severance Pay	10	*20*
	Total	100	100

er. By contrast, in property division it is possible to give all of one good to one player. Under certain conditions, 2-player cake division, and the procedure to be discussed next (adjusted winner), are equivalent [22].

[4]A website for AW can be found at http://www.nyu.edu/projects/adjustedwinner. Procedures applicable to more than two players are discussed in [13,15,23,32].

The italicized figures show the side that wins initially on each issue by placing more points on it. Notice that whereas the company wins a total of $10 + 15 + 15 = 40$ of its points, the executive wins a whopping $40 + 20 + 20 = 80$ of its points.

This outcome is obviously unfair to the company. Hence, a so-called *equitability adjustment* is necessary to equalize the points of the two sides. This adjustment transfers points from the initial winner (the executive) to the loser (the company).

The key to the success of AW – in terms of a mathematical guarantee that no win-win potential is lost – is to make the transfer in a certain order (for a proof, see [13], pp. 85–94). That is, of the issues initially won by the executive, look for the one on which the two sides are in closest agreement, as measured by the quotient of the winner's points to the loser's points. Because the winner-to-loser quotient on the issue of salary is $40/35 = 1.14$, and this is smaller than on any other issue on which the executive wins (the next-smallest quotient is $20/15 = 1.33$ on stock options), some of this issue must be transferred to the company.

But how much? The point totals of the company and the executive will be equal when the company's winning points on issues 1, 4, and 5, plus x percent of its points on salary (left side of equation below), equal the executive's winning points on issues 2, 3, and 6, minus x percent of its points on salary (right side of equation):

$$40 + 35x = 80 - 40x$$
$$75x = 40 .$$

Solving for x gives $x = 8/15 \approx 0.533$. This means that the executive will win about 53% on salary, and the company will lose about 53% (i. e., win about 47%), which is almost a 50-50 compromise between the low and high figures they negotiated earlier, only slightly favoring the executive.

This compromise ensures that both the company and the executive will end up with exactly the same total number of points after the equitability adjustment:

$$40 + 35(.533) = 80 - 40(.533) \approx 58.7 .$$

On all other issues, either the company or the executive gets its way completely (and its winning points), as it should since it valued these issues more than the other side.

Thus, AW is essentially a winner-take-all procedure, except on the one issue on which the two sides are closest and which, therefore, is the one subject to the equitability adjustment. On this issue a split will be necessary, which

will be easier if the issue is a quantitative one, like salary, than a more qualitative one like title and responsibilities.[5]

Still, it should be possible to reach a compromise on an issue like title and responsibilities that reflects the percentages the relative winner and relative loser receive (53% and 47% on salary in the example). This is certainly easier than trying to reach a compromise on each and every issue, which is also less efficient than resolving them all at once according to AW.[6]

In the example, each side ends up with, in toto, almost 59% of what it desires, which will surely foster greater satisfaction than would a 50-50 split down the middle on each issue. In fact, assuming the two sides are truthful, there is no better split for both, which makes the AW settlement *efficient*.

In addition, it is *equitable*, because each side gets exactly the same amount above 50%, with this figure increasing the greater the differences in the two sides' valuations of the issues. In effect, AW makes optimal trade-offs by awarding issues to the side that most values them, except as modified by the equitability adjustment that ensures that both sides do equally well (in their own subjective terms, which may not be monetary). On the other hand, if the two sides have unequal claims or entitlements – as specified, for example, in a contract – AW can be modified to give each side shares of the total proportional to its specified claims.

Can AW be manipulated to benefit one side? It turns out that exploitation of the procedure by one side is practically impossible unless that side knows exactly how the other side will allocate its points. In the absence of such information, attempts at manipulation can backfire miserably, with the manipulator ending up with less than the minimum 50 points its honesty guarantees it [13,15].

While AW offers a compelling resolution to a multi-issue dispute, it requires careful thought to delineate what the issues being divided are, and tough bargaining to determine what winning and losing means on each. More specifically, because the procedure is an additive point scheme, the issues need to be made as independent as possible, so that winning or losing on one does not substantially affect how much one wins or loses on others. To the degree that this is not the case, it becomes less meaningful to use the point totals to indicate how well each side does.

---

[5]AW may require the transfer of more than one issue, but at most one issue must be divided in the end.

[6]A procedure called *proportional allocation* (PA) awards issues to the players in proportion to the points they allocate to them. While inefficient, PA is less vulnerable to strategic manipulation than AW, with which it can be combined ([13], pp. 75–80).

The half dozen issues identified in the executive-compensation example overlap to an extent and hence may not be viewed as independent (after all, might not the bonus be considered part of salary?). On the other hand, they might be reasonably thought of as different parts of a compensation *package*, over which the disputants have different preferences that they express with points. In such a situation, losing on the issues you care less about than the other side will be tolerable if it is balanced by winning on the issues you care more about.

Indivisible Goods

The challenge of dividing up indivisible goods, such as a car, a boat, or a house in a divorce, is daunting, though sometimes such goods can be shared (usually at different times). The main criteria I invoke are *efficiency* (there is no other division better for everybody, or better for some players and not worse for the others) and *envy-freeness* (each player likes its allocation at least as much as those that the other players receive, so it does not envy anybody else). But because efficiency, by itself, is not a criterion of fairness (an efficient allocation could be one in which one player gets everything and the others nothing), I also consider other criteria of fairness besides envy-freeness, including Rawlsian and utilitarian measures of welfare (to be defined).

I present two paradoxes, from a longer list of eight in [4],[7] that highlight difficulties in creating "fair shares" for everybody. But they by no means render the task impossible. Rather, they show how dependent fair division is on the fairness criteria one deems important and the trade-offs one considers acceptable. Put another way, achieving fairness requires some consensus on the ground rules (i. e., criteria), and some delicacy in applying them (to facilitate trade-offs when the criteria conflict).

I make five assumptions. First, players rank indivisible items but do not attach cardinal utilities to them. Second, players cannot compensate each other with side payments (e. g., money) – the division is only of the indivisible items. Third, players cannot randomize among different allocations, which is a way that has been proposed for "smoothing out" inequalities caused by the indivisibility of items. Fourth, all players have positive values for every item. Fifth, a player prefers one set *S* of items to a different set *T* if (i) *S* has as many items as *T* and (ii) for every item *t* in *T* and not in *S*, there is a distinct item *s* in *S* and not *T* that the player prefers to *t*. For example, if a player

---

[7]For a more systematic treatment of conflicts in fairness criteria and trade-offs that are possible, see [5,6,9,11,18,21].

ranks four items in order of decreasing preference, 1 2 3 4, I assume that it prefers

- the set {1,2} to {2,3}, because {1} is preferred to {3}; and
- the set {1,3} to {2,4}, because {1} is preferred to {2} and {3} is preferred to {4},

whereas the comparison between sets {1,4} and {2,3} could go either way.

Paradox 1. A unique envy-free division may be inefficient.

Suppose there is a set of three players, {A, B, C}, who must divide a set of six indivisible items, {1, 2, 3, 4, 5, 6}. Assume the players rank the items from best to worst as follows:

    A : 1 2 3 4 5 6
    B : 4 3 2 1 5 6
    C : 5 1 2 6 3 4

The unique envy-free allocation to (A, B, C) is ({1,3}, {2,4}, {5,6}), or for simplicity (13, 24, 56), whereby A and B get their best and 3rd-best items, and C gets its best and 4th-best items. Clearly, A prefers its allocation to that of B (which are A's 2nd-best and 4th-best items) and that of C (which are A's two worst items). Likewise, B and C prefer their allocations to those of the other two players. Consequently, the division (13, 24, 56) is envy-free: All players prefer their allocations to those of the other two players, so no player is envious of any other.

Compare this division with (12, 34, 56), whereby A and B receive their two best items, and C receives, as before, its best and 4th-best items. This division *Pareto-dominates* (13, 24, 56), because two of the three players (A and B) prefer the former allocation, whereas both allocations give player C the same two items (56).

It is easy to see that (12, 34, 56) is Pareto-optimal or efficient: No player can do better with some other division without some other player or players doing worse, or at least not better. This is apparent from the fact that the only way A or B, which get their two best items, can do better is to receive an additional item from one of the two other players, but this will necessarily hurt the player who then receives fewer than its present two items. Whereas C can do better without receiving a third item if it receives item 1 or item 2 in place of item 6, this substitution would necessarily hurt A, which will do worse if it receives item 6 for item 1 or 2.

The problem with efficient allocation (12, 34, 56) is that it is not *assuredly* envy-free. In particular, C will envy A's allocation of 12 (2nd-best and 3rd-best items for C) if it prefers these two items to its present allocation of 56

(best and 4th-best items for C). In the absence of information about C's preferences for subsets of items, therefore, we cannot say that efficient allocation (12, 34, 56) is envy-free.[8]

But the real bite of this paradox stems from the fact that not only is inefficient division (13, 24, 56) envy-free, but it is uniquely so – there is no other division, including an efficient one, that guarantees envy-freeness. To show this in the example, note first that an envy-free division must give each player its best item; if not, then a player might prefer a division, like envy-free division (13, 24, 56) or efficient division (12, 34, 56), that does give each player its best item, rendering the division that does not do so envy-possible or envy-ensuring. Second, even if each player receives its best item, this allocation cannot be the only item it receives, because then the player might envy any player that receives two or more items, *whatever* these items are.

By this reasoning, then, the only possible envy-free divisions in the example are those in which each player receives two items, including its top choice. It is easy to check that no efficient division is envy-free. Similarly, one can check that no inefficient division, except (13, 24, 56), is envy-free, making this division uniquely envy-free.

Paradox 2. Neither the Rawlsian maximin criterion nor the utilitarian Borda-score criterion may choose a unique efficient and envy-free division.

Unlike the example illustrating paradox 1, efficiency and envy-freeness are compatible in the following example:

    A: 1 2 3 4 5 6
    B: 5 6 2 1 4 3
    C: 3 6 5 4 1 2

There are three efficient divisions in which (A, B, C) each get two items: (i) (12, 56, 34); (12, 45, 36); (iii) (14, 25, 36). Only (iii) is envy-free: Whereas C might prefer B's 56 allocation in (i), and B might prefer A's 12 allocation in (ii), no player prefers another player's allocation in (iii).

Now consider the following *Rawlsian maximin criterion* to distinguish among the efficient divisions: Choose

---

[8]Recall that an *envy-free* division of indivisible items is one in which, no matter how the players value subsets of items consistent with their rankings, no player prefers any other player's allocation to its own. If a division is not envy-free, it is *envy-possible* if a player's allocation *may* make it envious of another player, depending on how it values subsets of items, as illustrated for player C by division (12, 34, 56). It is *envy-ensuring* if it causes envy, independent of how the players value subsets of items. In effect, a division that is envy-possible has the potential to cause envy. By comparison, an envy-ensuring division always causes envy, and an envy-free division never causes envy.

a division that maximizes the minimum rank of items that players receive, making a worst-off player as well off as possible.[9] Because (ii) gives a 5th-best item to B, whereas (i) and (iii) give players, at worst, a 4th-best item, the latter two divisions satisfy the Rawlsian maximin criterion.

Between these two, (i), which is envy-possible, is arguably better than (iii), which is envy-free: (i) gives the two players that do not get a 4th-best item their two best items, whereas (iii) does not give B its two best items.[10]

Now consider what a modified Borda count would also give the players under each of the three efficient divisions. Awarding 6 points for obtaining a best item, 5 points for obtaining a 2nd-best item, . . . , 1 point for obtaining a worst item in the example, (ii) and (iii) give the players a total of 30 points, whereas (i) gives the players a total of 31 points.[11] This criterion, which I call the *utilitarian Borda-score criterion*, gives the nod to division (i); the Borda scores provide a measure of the overall utility or welfare of the players. Thus, neither the Rawlsian maximin criterion nor the utilitarian Borda-score criterion guarantees the selection of the unique efficient and envy-free division of (iii).

Conclusions

The squeezing procedure I illustrated for dividing up a cake among three players ensures efficiency and envy-freeness, but it does not satisfy equitability. Whereas adjusted winner satisfies efficiency, envy-freeness, and equitability for two players dividing up several divisible goods, all these properties cannot be guaranteed if there are more than two players. Finally, the two paradoxes relating to the fair division of indivisible good, which are independent of the procedure used, illustrate new difficulties – that no division may satisfy either maximin or utilitarian notions of welfare and, at the same time, be efficient and envy-free.

---

[9]This is somewhat different from Rawls's (1971) proposal to maximize the utility of the player with minimum utility, so it might be considered a modified Rawlsian criterion. I introduce a rough measure of utility next with a modified Borda count.

[10]This might be considered a second-order application of the maximin criterion: If, for two divisions, players rank the worst item any player receives the same, consider the player that receives a next-worst item in each, and choose the division in which this item is ranked higher. This is an example of a *lexicographic decision rule*, whereby alternatives are ordered on the basis of a most important criterion; if that is not determinative, a next-most important criterion is invoked, and so on, to narrow down the set of feasible alternatives.

[11]The standard scoring rules for the Borda count in this 6-item example would give 5 points to a best item, 4 points to a 2nd-best item, . . . , 0 points to a worst item. I depart slightly from this standard scoring rule to ensure that each player obtains some positive value for all items, including its worst choice, as assumed earlier.

Future Directions

Patently, fair division is a hard problem, whatever the things being divided are. While some conflicts are ineradicable, as the paradoxes demonstrate, the trade-offs that best resolve these conflicts are by no means evident. Understanding these may help to ameliorate, if not solve, practical problems of fair division, ranging from the splitting of the marital property in a divorce to determining who gets what in an international dispute.

Bibliography

1. Barbanel JB (2005) The Geometry of Efficient Fair Division. Cambridge University Press, New York
2. Barbanel JB, Brams SJ (2004) Cake Division with Minimal Cuts: Envy-Free Procedures for 3 Persons, 4 Persons, and Beyond. Math Soc Sci 48(3):251–269
3. Barbanel JB, Brams SJ (2007) Cutting a Pie Is Not a Piece of Cake. Am Math Month (forthcoming)
4. Brams SJ, Edelman PH, Fishburn PC (2001) Paradoxes of Fair Division. J Philos 98(6):300–314
5. Brams SJ, Edelman PH, Fishburn PC (2004) Fair Division of Indivisible Items. Theory Decis 55(2):147–180
6. Brams SJ, Fishburn PC (2000) Fair Division of Indivisible Items Between Two People with Identical Preferences: Envy-Freeness, Pareto-Optimality, and Equity. Soc Choice Welf 17(2):247–267
7. Brams SJ, Jones MA, Klamler C (2006) Better Ways to Cut a Cake. Not AMS 35(11):1314–1321
8. Brams SJ, Jones MA, Klamler C (2007) Proportional Pie Cutting. Int J Game Theory 36(3–4):353–367
9. Brams SJ, Kaplan TR (2004) Dividing the Indivisible: Procedures for Allocating Cabinet Ministries in a Parliamentary System. J Theor Politics 16(2):143–173
10. Brams SJ, Kilgour MD (2001) Competitive Fair Division. J Political Econ 109(2):418–443
11. Brams SJ, King DR (2004) Efficient Fair Division: Help the Worst Off or Avoid Envy? Ration Soc 17(4):387–421
12. Brams SJ, Taylor AD (1995) An Envy-Free Cake Division Protocol. Am Math Month 102(1):9–18
13. Brams SJ, Taylor AD (1996) Fair Division: From Cake-Cutting to Dispute Resolution. Cambridge University Press, New York
14. Brams SJ, Taylor AD (1999a) Calculating Consensus. Corp Couns 9(16):47–50
15. Brams SJ, Taylor AD (1999b) The Win-Win Solution: Guaranteeing Fair Shares to Everybody. W.W. Norton, New York
16. Brams SJ, Taylor AD, Zwicker SW (1995) Old and New Moving-Knife Schemes. Math Intell 17(4):30–35
17. Brams SJ, Taylor AD, Zwicker WS (1997) A Moving-Knife Solution to the Four-Person Envy-Free Cake Division Problem. Proc Am Math Soc 125(2):547–554
18. Edelman PH, Fishburn PC (2001) Fair Division of Indivisible Items Among People with Similar Preferences. Math Soc Sci 41(3):327–347
19. Even S, Paz A (1984) A Note on Cake Cutting. Discret Appl Math 7(3):285–296
20. Haake CJ, Raith MG, Su FE (2002) Bidding for Envy-Freeness: A Procedural Approach to *n*-Player Fair Division Problems. Soc Choice Welf 19(4):723–749

21. Herreiner D, Puppe C (2002) A Simple Procedure for Finding Equitable Allocations of Indivisible Goods. Soc Choice Welf 19(2):415–430
22. Jones MA (2002) Equitable, Envy-Free, and Efficient Cake Cutting for Two People and Its Application to Divisible Goods. Math Mag 75(4):275–283
23. Moulin HJ (2003) Fair Division and Collective Welfare. MIT Press, Cambridge
24. Peterson E, Su FE (2000) Four-Person Envy-Free Chore Division. Math Mag 75(2):117–122
25. Pikhurko O (2000) On Envy-Free Cake Division. Am Math Month 107(8):736–738
26. Potthoff RF (2002) Use of Linear Programming to Find an Envy-Free Solution Closest to the Brams-Kilgour Gap Solution for the Housemates Problem. Group Decis Negot 11(5):405–414
27. Robertson JM, Webb WA (1997) Near Exact and Envy-Free Cake Division. Ars Comb 45:97–108
28. Robertson J, Webb W (1998) Cake-Cutting Algorithms: Be Fair If You Can. AK Peters, Natick
29. Shishido H, Zeng DZ (1999) Mark-Choose-Cut Algorithms for Fair and Strongly Fair Division. Group Decis Negot 8(2): 125–137
30. Stromquist W (1980) How to Cut a Cake Fairly. Am Math Month 87(8):640–644
31. Su FE (1999) Rental Harmony: Sperner's Lemma in Fair Division. Am Math Month 106:922–934
32. Young HP (1994) Equity in Theory and Practice. Princeton University Press, Princeton
33. Zeng DZ (2000) Approximate Envy-Free Procedures. Game Practice: Contributions from Applied Game Theory. Kluwer Academic Publishers, Dordrecht, pp. 259–271

Field Computation in Natural and Artificial Intelligence

BRUCE J. MACLENNAN
Department of Electrical Engineering and Computer Science, University of Tennessee, Knoxville, USA

Article Outline

Glossary

Axon Nerve fiber adapted to the efficient, reliable, active transmission of neural impulses between locations in the brain or body.

Dendrite Nerve fibers adapted to the (primarily passive) sensing and integration of signals from other neurons, which are transmitted to the neuron cell body.

Dirac delta function A *distribution* or *generalized function* that is defined to be infinite at the origin, zero everywhere else, and to have unit area (or volume). More generally, such a function but with its infinite point located elsewhere than the origin. Dirac delta functions are idealized impulses and exist as limit objects in Hilbert spaces.

Eigenfield An eigenfield of a linear operator has the property of passing through the operator with its shape unchanged and only its amplitude possibly modified. Equivalent to an eigenfunction of the operator, but stresses the function's role as a field.

Field A continuous distribution of continuous quantity. Mathematically, an element of an appropriate space, such as a Hilbert space, of continuous-valued functions over a continuum. See also *phenomenological field* and *structural field*.

Field computation Computation in which data are represented by fields, or by other representations that can be mathematically modeled by fields.

Field space A suitably constrained set of fields. Generally field spaces are taken to be subspaces of Hilbert spaces.

Field transformation Functions between field spaces; more generally, functions whose input and/or outputs are fields. Synonymous with *operator* in this article.

Functional A scalar-valued function of functions, and in particular a scalar-valued field transformation.

Idempotency An operation is idempotent when repeating it several times has the same effect as doing it once.

Impulse response The response of a system to an input that is an idealized impulse (a *Dirac delta function*, q.v.).

Microfeature Features of a stimulus or representation that are much smaller and at a lower level than ordinary (macro-)features, which are the sort of properties for which natural languages have words. Typically microfeatures have meaning (are interpretable) only in the context of many other microfeatures. Pixels are examples of microfeatures of images.

Nullcline In a two-dimensional system of differential equations ($\dot{u}_k = f_k(u, v)$, $k = 1, 2$), the lines along which each of the derivatives is zero ($f_k(u, v) = 0$).

Operator A function of functions (i. e., a functions whose inputs and/or outputs are functions), and in particular a function whose inputs and/or outputs are fields. Operators may be linear or nonlinear. Synonymous, in this article, with *field transformation*.

Orthonormal (ON) A set of vectors, fields, or functions is orthonormal if they are: (1) mutually orthogonal (i. e., inner products of distinct elements are 0), and (2) individually normalized (i. e., inner products of elements with themselves are 1).

Phenomenological field A physical distribution of quantity that for practical purposes may be treated mathematically as a continuous distribution of continuous quantity, although it is not so in reality (cf. *structural field*).

Physical realizability A field is physically realizable if it can be represented in some physical medium.

Population coding Neural representation in which a population of neurons jointly represent a stimulus or other information. Each individual neuron is broadly tuned to a range of stimuli, but collectively they can represent a stimulus accurately.

Post–Moore's law computing Refers to computing paradigms that will be important after the expiration of Moore's Law [48], which predicts a doubling of digital logic density every two years.

Projection A systematic pattern of axonal connections from one region of a brain to another.

Radial basis function (RBF) One of a set of real-valued functions, each of whose value decreases with distance from a central point (different for each function). The set as a whole satisfies some appropriate criteria of completeness (ability to approximate a class of functions).

Receptive field The receptive field of a sensory neuron is the set of stimuli to which it responds. By extension, the receptive field of a non-sensory neuron is the set of inputs (from other neurons) to which it responds. Each neuron has a *receptive field profile* which describes the extent to which particular patterns of input stimulate or inhibit activity in the neuron (and so, in effect, its receptive field is fuzzy-boundaried).

Structural field A physical field that is in reality a continuous distribution of continuous quantity (cf. *phenomenological field*).

Synapse A connection between neurons, often from the axon of one to the dendrite of another. Electrical impulses in the pre-synaptic neuron cause neurotransmitter molecules to be secreted into the synapses between the neurons. These chemicals bind to receptors in the post-synaptic neuron membrane, and cause fluctuations in the membrane potential.

Transfer function A function expressing the effect of a linear system on its input, expressed in terms of its effect on the amplitude and phase of each component frequency.

Unit doublet A generalized function that is the derivative of the Dirac delta function (q.v.). It is zero except infinitesimally to the left of the origin, where it is $+\infty$, and infinitesimally to the right of the origin, where it is $-\infty$.

Definition of the Subject

A *field* may be defined as a spatially continuous distribution of continuous quantity. The term is intended to include physical fields, such as electromagnetic fields and potential fields, but also patterns of electrical activity over macroscopic regions of neural cortex. Fields include two-dimensional representations of information, such as optical images and their continuous Fourier transforms, and one-dimensional images, such as audio signals and their spectra, but, as will be explained below, fields are not limited to two or three dimensions. A *field transformation* is a mathematical operation or function that operates on one or more fields in parallel yielding one or more fields as results. Since, from a mathematical standpoint, fields are defined over a continuous domain, field transformations operate with *continuous parallelism*. Some examples of field transformations are point-wise summation and multiplication of fields, Fourier and wavelet transforms, and convolutions and correlations.

Field computation is a model of computation in which information is represented primarily in fields and in which information processing is primarily by means of field transformations. Thus is may be described as *continuously analog computing* (see ▶ Analog Computation). Field computation may be *feed-forward*, in which one or more fields progress through a series of field transformations from input to output, or it may be *recurrent*, in which there is feedback from later stages of the field computation back to earlier stages. Furthermore, field computations can proceed in discrete sequential steps (similar to digital program execution, but with each step applying a field transformation), or in continuous time according to partial differential equations.

A distinction is often made in science between *structural fields* and *phenomenological fields*. Structural fields are physically continuous distributions of continuous quantity, such as gravitational fields and electromagnetic fields. Phenomenological fields are distributions of quantity that can be treated mathematically as though they are continuous, even if they are not physically continuous. For example, the velocity field of a macroscopic volume of fluid is a phenomenological field, because it is not physically continuous (each discrete molecule has its own velocity), but can be treated as though it is. Similarly, a macro-

scopic charge distribution is a phenomenological field because charge is quantized but can be treated as a continuous quantity for many purposes. Although structural fields are sometimes used, often field computation is concerned with phenomenological fields, that is, with information that can be treated as a continuous distribution of continuous quantity, regardless of whether it is physically continuous. Practically, this means that quantization in both the distribution and the quantity must be sufficiently fine that they can be modeled mathematically by continua.

One of the goals of field computation is to provide a mathematical language for describing information processing in the brain and in future large artificial neural networks intended to exhibit brain-scale intelligence. Neural computation is qualitatively different from ordinary digital computation. Computation on an ordinary computer can be characterized as *deep but narrow*; that is, the processor operates on relatively few data values at a time, but the operations are very rapid, and so many millions of operations can be executed each second. Even on a modern parallel computer, the degree of parallelism is modest, on the order of thousands, whereas even a square millimeter of cortex has at least 146 000 neurons operating in parallel (see p. 51 in [11]). On the other hand, since neurons are quite slow (responding on the order of milliseconds), the "100-Step Rule" says that there can be at most about 100 sequential processing stages between sensory input and response [19]. Therefore, neural computation is *shallow but wide*; that is, it uses relatively few sequential stages, but each operates with a very high degree of parallelism (on the order of many millions). In addition to its speed, modern electronic digital arithmetic is relatively precise compared to the analog computation of neurons (at most about one digit of precision) (see p. 378 in [43]). Therefore we can conclude that neuronal information processing operates according to quite different principles to ordinary digital computing.

It is not unreasonable to suppose that achieving an artificial intelligence that is comparable to the natural intelligence of mammals will require a similar information processing architecture; in any case that seems to be a promising research direction. Therefore we should be aiming toward components with computational capabilities comparable to neurons and densities of at least 15 million per square centimeter. Fortunately, the brain demonstrates that these components do not have to be high-speed, high-precision devices, nor do they have to be precisely connected, for the detailed connections can be established through self-organization and learning. The theory of field computation can contribute in two ways: first, by providing a mathematical framework for understanding infor-

mation in massively parallel analog computation systems, such as the brain, and second, by suggesting how to exploit relatively homogeneous masses of computational materials (e. g., thin films, new nanostructured materials). For the same reasons, field computers may provide an attractive alternative for "post-Moore's law computing".

Introduction

The term "field computation" dates from 1987 [32], but examples of field computation are much older. For example, G Kirchhoff (1824–87) and others developed the *field analogy method* in which partial differential equation (PDE) problems are solved by setting up an analogous physical system and measuring it [26]. Thus a two-dimensional boundary value problem, for example determining a steady-state temperature or magnetic field distribution, could solved by setting up an analogous system with a conductive sheet or a shallow tank containing an electrolytic solution (see p. 34 in [63]). When the boundary conditions were applied, the system computed the steady-state solution field in parallel and at electronic speed. The resulting potential field could not be displayed directly at that time, and so it was necessary to probe the field at discrete points and plot the equipotential lines; later devices allowed the equipotentials to be traced more or less automatically (see pp. 2-6 in [70]). In either case, setting up the problem and reading out the results were much slower than the field computation, which was comparatively instantaneous. Three-dimensional PDEs were similarly solved with tanks containing electrolytic solutions (see pp. 2-5–6 in [70]). (For more on conductive sheet and electrolytic tanks methods, see Chap. 9 in [65].)

Non-electronic field computation methods were also developed in the nineteenth century, but continued to be used through the first half of the twentieth century to solve the complex PDEs that arise in practical engineering (see pp. 2-8–9 in [70]). For example, so called "rubber-sheet computers" were used to compute the complex electric fields in vacuum tubes. A thin elastic membrane represented the field, and rods or plates pushing the sheet down from above or up from below represented constant negative or positive potential sources. The sheet assumed the shape of the electrical potential field, which could be viewed directly. By altering the rods and plates and observing the effects on the sheet, the engineer could develop an intuitive understanding of the field's dependence on the potential sources. These simple mechanical devices used effectively instantaneous field computation to display the steady-state field's dependence on the boundary conditions.

Electrolytic tanks and conductive and elastic sheets are all examples of the use of structural fields in computation, but other mechanical field computers used discrete approximations of spatially continuous fields, and therefore made use of phenomenological fields. For example, "pin-and-rod systems", which were developed in the nineteenth century, exploited the fact that equipotential lines and flux (or stream) lines always cross at right angles (see pp. 2-9–11 in [70]). A (two-dimensional) field was represented by two arrays of flexible but stiff wires, representing the flux and equipotential lines. At each crossing point was a pin with two perpendicular holes drilled through it, through which the crossing wires passed. The pins were loose enough that they could move on the wires, while maintaining, of course, their relative position and the perpendicular crossings of the wires. To solve a PDE problem (for example, determining the pressure potentials and streamlines of a non-tubulent flow through a nozzle), the edges of the pin-and-rod system were bent to conform to the boundary conditions; the rest of the system adjusted itself to the steady-state solution field. Like the rubber-sheet computers, pin-and-rod systems allowed the solution field to be viewed directly and permitted exploration of the effects on the solution of changes in the boundary conditions.

Through the first half of the twentieth century, *network analyzers* were popular electronic analog computers, which were often used for field computation (see pp. 35–40 in [63]) This was similar to the field analogy method, but a discrete network of resistors or resistive elements replaced such continuous conducting media as the electrolytic tank and conductive sheet. Nevertheless, a sufficiently fine mesh of resistive elements may be treated as a phenomenological field, and network analyzers were used to solve PDE problems (see pp. 2-6–8 in [70]). Boundary conditions were defined by applying voltages to the appropriate locations in the network, and the resulting steady-state field values were determined by measuring the corresponding nodes in the network. As usual, it was possible to monitor the effects of boundary condition changes on particular locations in the field, and to plot them automatically or display them on an oscilloscope.

The field computers discussed so far were suited to determining the steady-state solution of a system of PDEs given specified boundary conditions; as a consequence they were sometimes called *field plotters* or *potential analyzers* (see pp. 2-3 in [70]). These are essentially static problems, although, as we have seen, it was possible to simulate and monitor changes in the (relatively) steady-state solution as a consequence of (relatively slowly) changing boundary conditions. On the other hand, truly dynamic problems, which simulated the evolution of a field in time, could be addressed by *reactive networks*, that is, networks incorporating capacitive and inductive elements as well as resistors (see pp. 2-11–13 in [70]). For example an *RC network analyzer*, which had capacitance at each of the nodes of the resistor network, could solve the diffusion equation, for the charge on the capacitors corresponded to the concentration of the diffusing substance at corresponding locations in the medium. An *RLC network analyzer* had inductance, as well as resistance and capacitance, at each node, and so it was able to address a wider class of PDEs, including wave equations.

Although these twentieth-century field computers were constructed from discrete resistors, capacitors, and inductors, which limited the size of feasible networks, analog VLSI and emerging fabrication technologies will permit the implementation of much denser devices incorporating these and similar field computation techniques (see Sect. "Field Computers").

The following section will present the mathematical foundations and notation for field computation; Hilbert spaces provide the basic mathematical framework. Next we discuss examples of field computation in the brain, especially in its computational maps. Fields appear in a number of contexts, including activity at the axon hillocks, in patterns of axonal connection between areas, and in patterns of synaptic connection to dendrites. The following section presents examples of field computation in the brain and in other natural and artificial systems, including fields for sensorimotor processing, excitable media, and diffusion processes. Next we consider special topics in field computation in cognition, including the separation of information (semantics) from pragmatics, and the analysis of discrete symbols as field excitations. We also consider the relevance of universal mutivariate approximation theorems to general-purpose field computation. Then we discuss hardware specifically oriented toward field computation, including electronic, optical, and chemical technologies. Finally, we consider future directions for field computation research.

Basic Principles

Mathematical Definitions

Mathematically, a field is (generally continuous) function $\phi \colon \Omega \to K$ defined over some bounded domain Ω (often a compact subset of a Euclidean space) and taking values in an algebraic field K. Typically K is the real numbers, but in some applications it is the complex numbers or a vector space (see Sects. "Neuronal Fields", "Diffusion Processes", "Motion in Direction Fields", and "Gabor Wavelets and Coherent States" below).

As usual, the value of a field ϕ at a point $u \in \Omega$ of its domain can be denoted by $\phi(u)$, but we more often use the notation ϕ_u with the same meaning. The latter is especially convenient for time-varying fields. For example, the value of a field ϕ at point u and time t can be denoted by $\phi_u(t)$ rather than $\phi(u, t)$. The entire field at a particular time t is then written $\phi(t)$. As is commonly done in mathematics, we may consider ϕ to be a variable implicitly defined over all $u \in \Omega$. (In this article lower-case Greek letters are usually used for fields. We occasionally use bold-face numbers, such as **0** and **1**, for constant-valued fields; thus $\mathbf{0}_u = 0$ for all $u \in \Omega$. When it is necessary to make the field's domain explicit, we write $\mathbf{0}_\Omega$, $\mathbf{1}_\Omega$, etc.)

For practical field computation (e. g., in natural and artificial intelligence) we are interested in fields that can be realized in some physical medium, which places restrictions on the space of allowable fields. These restrictions vary somewhat for different physical media (e. g., neural cortex or optical fields), but we can specify a few general conditions for *physical realizability*. Generally, fields are defined over a bounded domain, although sometimes we are interested in fields that are extended in time with no prespecified bound (e. g., an auditory signal). Furthermore, since most media cannot represent unbounded field amplitudes, it is reasonable to assume that a field's range of variation is also bounded (e. g., $|\phi_u| \leq B$ for all $u \in \Omega$). In addition, most media will not support unbounded gradients, so the field's derivatives are bounded. Indeed, physically realizable fields are band-limited in both the spatial and temporal domains. Although different assumptions apply in different applications, from a mathematical perspective it is generally convenient to assume that fields are *uniformly continuous square-integrable functions* (defined below), and therefore that they belong to a Hilbert function space. In any case we use the notation $\Phi_K(\Omega)$ for a physically realizable space of K'' valued fields over a domain Ω, and write $\Phi(\Omega)$ when the fields' values are clear from context.

The foregoing considerations suggest that the inner product of fields is an important concept, and indeed it is fundamental to Hilbert spaces Therefore, if ϕ and ψ are two real-valued fields with the same domain, $\phi, \psi \in \Phi(\Omega)$, we define their inner product in the usual way:

$$\langle \phi \mid \psi \rangle = \int_\Omega \phi_u \psi_u \mathrm{d}u \,.$$

If the fields are complex-valued, then we take the complex conjugate of one of the fields:

$$\langle \phi \mid \psi \rangle = \int_\Omega \phi_u^* \psi_u \mathrm{d}u \,.$$

For vector-valued fields $\boldsymbol{\phi}, \boldsymbol{\psi} \in \Phi_{\mathbb{R}^n}(\Omega)$ we may define

$$\langle \boldsymbol{\phi} \mid \boldsymbol{\psi} \rangle = \int_\Omega \boldsymbol{\phi}_u \cdot \boldsymbol{\psi}_u \mathrm{d}u \,,$$

where $\boldsymbol{\phi}_u \cdot \boldsymbol{\psi}_u$ is the ordinary scalar product on the vector space $\mathbb{R}^n$. Finally, the inner-product norm $\|\phi\|$ is defined in the usual way:

$$\|\phi\|^2 = \langle \phi \mid \phi \rangle \,.$$

As previously remarked, the elements of a Hilbert space are required to be square-integrable ("finite energy"): $\|\phi\| < \infty$.

Field Transformations

A *field transformation* or *operator* is any continuous function that maps one or more input fields into one or more output fields. In the simplest case a field transformation $F \colon \Phi(\Omega) \to \Phi(\Omega')$ maps a field in the input space $\Phi(\Omega)$ into a field in the output space $\Phi(\Omega')$.

We do not want to exclude degenerate field transformations, which operate on a field to produce a single real number, for example, or operate on a scalar value to produce a field. (An example of the former is the norm operation, $\|\cdot\|$, and an example of the latter is the operator that produces a constant-valued field over a domain.) In these cases we can consider the inputs or outputs to belong to a space $\Phi(\Omega)$ in which Ω is a singleton set. For example, the real numbers can be treated as fields in

$$\Phi^0 = \Phi_{\mathbb{R}}(\{0\}) \,.$$

Since $\mathbb{R}$ and Φ^0 are isomorphic, we will ignore the difference between them when no confusion can result.

Another class of simple field transformations are the *local transformations*, in which each point of the output field is a function of the corresponding point in the input field. In the simplest case, the same function is applied at each point. Suppose that for input field $\phi \in \Phi_J(\Omega)$, the output field $\psi \in \Phi_K(\Omega)$ is defined $\psi_u = f(\phi_u)$, where $f \colon J \to K$. Then we write $\overline{f} \colon \Phi_J(\Omega) \to \Phi_K(\Omega)$ for the local transformation $\psi = \overline{f}(\phi)$. For example, $\overline{\log}(\phi)$ applies the log function to every element of ϕ and returns field of the results. More generally, we may apply a different function (from a parameterized family) at each point of the input field. Suppose $F \colon \Omega \times J \to K$, then we define $\overline{F} \colon \Phi_J(\Omega) \to \Phi_K(\Omega)$ so that if $\psi = \overline{F}(\phi)$, then $\psi_u = F(u, \phi_u)$.

Field transformations may be linear or nonlinear. The most common linear transformations are *integral operators of Hilbert–Schmidt type*, which are the field analogs of

matrix-vector products. Let $\phi \in \Phi(\Omega)$ and $L \in \Phi(\Omega' \times \Omega)$ be square-integrable fields; then the product $L\phi = \psi \in \Phi(\Omega')$ is defined:

$$\psi_u = \int_\Omega L_{uv}\phi_v \mathrm{d}v \,.$$

L is called the *kernel* of the operator. It is easy to show that physically realizable linear operators have a Hilbert–Schmidt kernel, because physically realizable fields and the operators on them are band-limited [33]. Therefore they can be computed by field products of the form $L\phi$.

According to the Riesz Representation Theorem (e.g., Sect. 12.4 in [9]), a continuous linear functional (real-valued operator) $L: \Phi(\Omega) \to \mathbb{R}$ has a *representer*, which is a field $\rho \in \Phi(\Omega)$ such that $L\phi = \langle \rho \mid \phi \rangle$. However, since linear operators are continuous if and only if they are bounded, and since practical field transformations are bounded, all practical linear functionals have representers.

We define multilinear products in the same way. Suppose $\phi_k \in \Phi(\Omega_k)$, for $k = 1, \ldots, n$, and $M \in \Phi(\Omega' \times \Omega_n \times \cdots \Omega_2 \times \Omega_1)$. Then $M\phi_1\phi_2\cdots\phi_n = \psi \in \Phi(\Omega')$ is defined

$$\psi_u = \int_{\Omega_n} \cdots \int_{\Omega_2} \int_{\Omega_1} M_{uv_n \cdots v_2 v_1}\phi_1(v_1)\phi_2(v_2) \\ \cdots \phi_n(v_n)\mathrm{d}v_1\mathrm{d}v_2\cdots\mathrm{d}v_n \,.$$

Again, physically realizable multilinear operators are band limited, and so they can be computed by this kind of multilinear product [33].

Like the field analogs of inner products and matrix-vector products, it is also convenient to define an analog of the outer product. For $\phi \in \Phi(\Omega)$ and $\psi \in \Phi(\Omega')$ we define the outer product $\phi \wedge \psi \in \Phi(\Omega \times \Omega')$ so that $(\phi \wedge \psi)_{(u,v)} = \phi_u\psi_v$, for $u \in \Omega$, $v \in \Omega'$. Inner, outer, and field products are related as follows for $\phi, \chi \in \Phi(\Omega)$ and $\psi \in \Phi(\Omega')$:

$$\phi(\chi \wedge \psi) = \langle \phi \mid \chi \rangle\psi = (\psi \wedge \chi)\phi \,.$$

In the simplest kind of field computation (corresponding to a feed-forward neural network), an input field ϕ is processed through one or more field transformations $F_1, \ldots, F_n$ to yield an output field ψ:

$$\psi = F_n(\cdots F_1(\phi)\cdots) \,.$$

This includes cases in which the output field is the continuously-varying image of a time-varying input field,

$$\psi(t) = F_n(\cdots F_1(\phi(t))\cdots) \,.$$

More complex feed-forward computations may involve additional input, output, and intermediate fields, which might be variable or not.

In an ordinary artificial neural network, the activity y_i of neuron i in one layer is defined by the activities $x_1, \ldots, x_n$ of the neurons in the preceding layer by an equation such as

$$y_i = s\left(\sum_{j=1}^{N} W_{ij}x_j + b_i\right) \,, \tag{1}$$

where W_{ij} is the weight or strength of the connection from neuron j to neuron i, b_i is a *bias term*, and $s: \mathbb{R} \to \mathbb{R}$ is a *sigmoid function*, that is, a non-decreasing, bounded continuous function. (The hyperbolic tangent is a typical example.) The field computation analog is obtained by taking the number of neurons in each layer to the continuum limit. That is, the activities ψ_u in one neural field ($u \in \Omega'$) are defined by the values ϕ_v in the input field ($v \in \Omega$) by this equation:

$$\psi_u = \int_\Omega L_{uv}\phi_v \mathrm{d}v + \beta_u \,,$$

where $L \in \Phi(\Omega' \times \Omega)$ is an *interconnection field* and $\beta \in \Phi(\Omega')$ is a *bias field*. More compactly,

$$\psi = \bar{s}(L\phi + \beta) \,. \tag{2}$$

Typically, the input is processed through a series of layers, each with its own weights and biases. Analogously, in field computation we may have an N-layer neural field computation, $\phi_k = \bar{s}(L_k\phi_{k-1} + \beta_k)$, $k = 1, \ldots, N$, where $\phi_0 \in \Phi(\Omega_0)$ is the input, $\phi_N \in \Phi(\Omega_N)$ is the output, $L_k \in \Phi(\Omega_k \times \Omega_{k-1})$ are the interconnection fields, and $\beta_k \in \Phi(\Omega_k)$ are the bias fields. Other examples of neural-network style field computing are discussed later (Sect. "Examples of Field Computation").

Many important field computation algorithms are iterative, that is, they sequentially modify one or more fields at discrete moments of time. They are analogous to ordinary computer programs, except that the variables contain fields rather than scalar quantities (integers, floating-point numbers, characters, etc., and arrays of these). Since the current value of a field variable may depend on its previous values, iterative field computations involve *feedback*. Examples of iterative algorithms include field computation analogs of neural network algorithms that adapt in discrete steps (e.g., ordinary back-propagation), and re-

current neural networks, which have feedback from later layers to earlier layers.

Analog field computers, like ordinary analog computers, can operate in continuous time, defining the continuous evolution of field variable by differential equations. For instance, $\dot{\phi} = F(\phi)$ is a simple first-order field-valued differential equation, which can be written $\mathrm{d}\phi_u(t)/\mathrm{d}t = F_u[\phi(t)]$. An example is the familiar diffusion equation $\dot{\phi} = k^2 \nabla^2 \phi$.

Continuously varying fields arise in a number of contexts in natural and artificial intelligence. For example, sensorimotor control (in both animals and robots) depends on the processing of continuously varying input fields (e. g., visual images or auditory signals) and their transformation into continuously varying output signals (e. g., to control muscles or mechanical effectors). One of the advantages of field computing for these applications is that the fields are processed in parallel, as they are in the brain. Often we find continuous field computation in optimization problems, in adaptation and learning, and in the solution of other continuous problems. For example, a field representing the interpretation of perceptual data (such as stereo disparity) may be continuously converging to the optimal interpretation or representation of the data.

Optimization problems are sometimes solved by continuous gradient ascent or descent on a potential surface defined by a functional F over a field space ($F: \Phi(\Omega) \to \mathbb{R}$), where $F(\phi)$ defines the "goodness" of solution ϕ. Gradient ascent is implemented by $\dot{\phi} = r\nabla F(\phi)$, where r is the rate of ascent. This and other examples are discussed in Sect. "Gradient Processes", but the use of the gradient raises the issue of the derivatives of field transformations, such as F, which we now address.

Derivatives of Field Transformations

Since fields are functions, field spaces are function spaces (generally, Hilbert spaces), and therefore the derivatives of field transformations are the derivatives of operators over function spaces (see § 251G in [42]). There are two common definitions of the differentiation of operators on Hilbert spaces (more generally, on Banach spaces), the Fréchet and the Gâteaux derivatives, which turn out to be the same for field transformations [33]. Therefore suppose that $F: \Phi(\Omega) \to \Phi(\Omega')$ is a field transformation and that U is an open subset of $\Phi(\Omega)$. Then $D \in \mathcal{L}(\Phi(\Omega), \Phi(\Omega'))$, the space of bounded linear operators from $\Phi(\Omega)$ to $\Phi(\Omega')$, is called the *Fréchet differential* of F at $\phi \in U$ if for all $\alpha \in \Phi(\Omega)$ such that $\phi + \alpha \in U$ there is an $E: \Phi(\Omega) \to \Phi(\Omega')$ such that,

$$F(\phi + \alpha) = F(\phi) + D(\alpha) + E(\alpha)$$

and

$$\lim_{\|\alpha\| \to 0} \frac{\|E(\alpha)\|}{\|\alpha\|} = 0 \,.$$

The *Fréchet derivative* $F': \Phi(\Omega) \to \mathcal{L}(\Phi(\Omega), \Phi(\Omega'))$ is defined by $F'(\phi) = D$, which is the locally linear approximation to F at ϕ.

Similarly $\mathrm{d}F: \Phi(\Omega) \times \Phi(\Omega) \to \Phi(\Omega')$ is a *Gâteaux derivative* of F if for all $\alpha \in U$ the following limit exists:

$$\mathrm{d}F(\phi, \alpha) = \lim_{s \to 0} \frac{F(\phi + s\alpha) - F(\phi)}{s} = \left. \frac{\mathrm{d}F(\phi + s\alpha)}{\mathrm{d}s} \right|_{s=0} \,.$$

If the Fréchet derivative exists, then the two derivatives are identical, $\mathrm{d}F(\phi, \alpha) = F'(\phi)(\alpha)$ for all $\alpha \in \Phi(\Omega)$.

Based on the analogy with finite-dimensional spaces, we define $\nabla F(\phi)$, the gradient of F at ϕ, to be the field $K \in \Phi(\Omega' \times \Omega)$ satisfying $F'(\phi)(\alpha) = K\alpha$ for all α in $\Phi(\Omega)$. That is, $F'(\phi)$ is an integral operator with kernel $K = \nabla F(\phi)$; note that $F'(\phi)$ is an operator but $\nabla F(\phi)$ is a field. The field analog of a directional derivative is then defined:

$$\nabla_\alpha F(\phi) = [\nabla F(\phi)]\alpha = F'(\phi)(\alpha) \,.$$

Because of their importance, it is worth highlighting the gradients of functionals (real-valued operators on fields). According to the preceding definition, the gradient of a functional $F: \Phi(\Omega) \to \Phi^0$ will be a two-dimensional field $\nabla F(\phi) \in \Phi(\{0\} \times \Omega)$. (Recall $\Phi^0 = \Phi(\{0\})$.) However, when confusion is unlikely, it is more convenient to define $\nabla F(\phi) = \gamma \in \Phi(\Omega)$, where γ is the representer of $F'(\phi)$. Then $F'(\phi)(\alpha) = \langle \gamma \mid \alpha \rangle = \langle \nabla F(\phi) \mid \alpha \rangle$.

Higher order derivatives of field operators are defined in the obvious way, but it is important to note that each derivative is of "higher type" that the preceding. That is, we have seen that if $F: \Phi(\Omega) \to \Phi(\Omega')$, then $\mathrm{d}F: \Phi(\Omega)^2 \to \Phi(\Omega')$, where $\Phi(\Omega)^2 = \Phi(\Omega) \times \Phi(\Omega)$. Similarly, $\mathrm{d}^n F: \Phi(\Omega)^{n+1} \to \Phi(\Omega')$. Also, as $F': \Phi(\Omega) \to \mathcal{L}(\Phi(\Omega), \Phi(\Omega'))$, so $F'': \Phi(\Omega) \to \mathcal{L}(\Phi(\Omega), \mathcal{L}(\Phi(\Omega), \Phi(\Omega')))$ and, in general,

$$F^{(n)}: \Phi(\Omega) \to \overbrace{\mathcal{L}(\Phi(\Omega), \mathcal{L}(\Phi(\Omega), \cdots, \mathcal{L}(\Phi(\Omega),}^{n} \Phi(\Omega')) \cdots)) \,.$$

Corresponding to higher-order derivatives are higher-order gradients:

$$\begin{aligned}
\mathrm{d}F^n(\phi, \alpha_1, \ldots, \alpha_n) &= \nabla^n F(\phi)\alpha_1 \cdots \alpha_n \\
&= \nabla^n F(\phi)(\alpha_n \wedge \cdots \wedge \alpha_1) \\
&= \nabla_{\alpha_n} \cdots \nabla_{\alpha_1} F(\phi) \,.
\end{aligned}$$

For reference, we state the chain rules for Fréchet and Gâteaux derivatives:

$$(F \circ G)'(\phi)(\alpha) = F'[G(\phi)][G'(\phi)(\alpha)] , \qquad (3)$$

$$d(F \circ G)(\phi, \alpha) = dF[G(\phi), dG(\phi, \alpha)] . \qquad (4)$$

Just as a real function can be expanded in a Taylor series around a point to obtain a polynomial approximation, there is a corresponding theorem in functional analysis that allows the expansion of an operator around a fixed field. This suggests an approach to general-purpose computation based on *field polynomials* [32], but there are also other approaches suggested by neural networks (see Sect. "Universal Approximation" below). We begin with a formal statement of the theorem.

Theorem 1 (Taylor) *Suppose that U is any open subset of $\Phi(\Omega)$ and that $F \colon \Phi(\Omega) \to \Phi(\Omega')$ is a field transformation that is C^n in U (that is, its first n derivatives exist). Let $\phi \in U$ and $\alpha \in \Phi(\Omega)$ such that $\phi + s\alpha \in U$ for all $s \in [0, 1]$. Then:*

$$F(\phi + \alpha) = \sum_{k=0}^{n-1} \frac{d^k F(\phi, \overbrace{\alpha, \ldots, \alpha}^{k})}{k!} + R_n(\phi, \alpha) ,$$

where

$$R_n(\phi, \alpha) = \int_0^1 \frac{(1-s)^{n-1} d^n F(\phi + s\alpha, \overbrace{\alpha, \ldots, \alpha}^{n})}{(n-1)!} \, ds .$$

Here the "zeroth derivative" is defined in the obvious way: $d^0 F(\phi) = F(\phi)$.

If the first n gradients exist (as they will for physically realizable fields and transformations), then the Taylor approximation can be written:

$$F(\phi + \alpha) = F(\phi) + \sum_{k=1}^{n} \frac{\nabla_\alpha^k F(\phi)}{k!} + R_n(\phi, \alpha) .$$

However, $\nabla_\alpha^k F(\phi) = \nabla^k F(\phi) \alpha^{(k)}$, where $\alpha^{(k)}$ is the k-fold outer product:

$$\alpha^{(k)} = \overbrace{\alpha \wedge \alpha \wedge \cdots \wedge \alpha}^{k} .$$

If we define the fields $\Gamma_k = \nabla^k F(\phi)$, then we can see this approximation as a "field polynomial":

$$F(\phi + \alpha) \approx F(\phi) + \sum_{k=1}^{n} \frac{\Gamma_k \alpha^{(k)}}{k!} .$$

Such an approximation may be computed by a field analog of "Horner's rule", which is especially appropriate for computation in a series of layers similar to a neural network. Thus $F(\phi + \alpha) \approx G_0(\alpha)$, where

$$G_k(\alpha) = \Gamma_k + \frac{G_{k+1}(\alpha)}{k+1} \alpha ,$$

for $k = 0, \ldots, n$, $\Gamma_0 = F(\phi)$, and $G_{n+1}(\alpha) = 0$.

Field Computation in the Brain

There are a number of contexts in mammalian brains in which information representations are usefully treated as fields, and information processing as field computation. These include neuronal cell bodies, patterns of axonal projection, and synapses. Of course, all of these are discrete structures, but in many cases the numbers are sufficiently large (e. g., 146×10^3 neurons/mm^2: see p. 51 in [11]) that the representations are usefully treated as fields; that is, they are *phenomenological fields*). (We omit discussing the intriguing possibility that the brain's electromagnetic field may affect conscious experience [44,50].)

Neuronal Fields

Computational maps, in which significant information is mapped to cortical location, are found throughout the brain [27]. For example, *tonotopic maps* in auditory cortex have systematic arrangements of neurons that respond to particular pitches, and *retinotopic maps* in visual cortex respond systematically to patches of color, edges, and other visual features projected onto the retina. Other *topographic maps* in somatosensory cortex and motor cortex systematically reflect sensations at particular locations in the body, or control motor activity at those locations, respectively. The number of identified maps is very large and there are probably many that have not been identified. And while some are quite large and can be investigated by fMRI and other noninvasive imaging techniques, other are less than a square millimeter in size [27]. However, even a $0.1 \, \text{mm}^2$ map may have tens of thousands of neurons, and thus be analyzed reasonably as a field.

In mathematical terms, let $\mathcal{X}$ be a space of features represented by a cortical map. These might be microfeatures of a sensory stimulus (e. g., oriented edges at particular retinal locations) or motor neurons (e. g., controlling muscle fibers in particular locations). These examples are peripheral features, but $\mathcal{X}$ might represent patterns of activity in nonperipheral groups of neurons (e. g., in other cortical maps). Let Ω be a two-dimensional manifold corresponding to a cortical map representing $\mathcal{X}$. There will a piecewise continuous function $\mu \colon \mathcal{X} \to \Omega$ so that $\mu(x)$

is the cortical location corresponding to feature $x \in X$. The mapping μ may be only piecewise continuous since X may be of higher dimension that Ω. (This is the reason, for example, that we find stripes in striate cortex; it is a consequence of mapping a higher dimensional space into a lower one.)

Typically, the activity $\phi_{\mu(x)}$ at a cortical location $\mu(x)$ will reflect the degree of presence of the feature x in the map's input. Furthermore, the responses of neurons are often broadly tuned, therefore the response at a location $\mu(x')$ will generally be a decreasing function $r[d(x, x')]$ of the distance $d(x, x')$, where d is some appropriate metric on X. Therefore an input feature x will generate a response field $\phi = \xi(x)$ given by

$$\phi_{\mu(x')} = r[d(x, x')] .$$

If a number of features $x_1, \ldots, x_n$ are simultaneously present in the input, then the activity in the map may be a superposition of the activities due to the individual features:

$$\xi(x_1) + \cdots + \xi(x_n).$$

Furthermore, a sensory or other input, represented as a subset $X' \subset X$ of the feature space, generates a corresponding field,

$$\xi(X') = \int_{X'} \xi(x) \mathrm{d}x$$

(with an appropriate definition of integration for X, which usually can be taken to be a measure space). (See Sect. "Nonlinear Computation via Topographic Maps" for more on computation on superpositions of inputs via topographic maps.)

The preceding discussion of cortical maps refers somewhat vaguely to the "activity" of neurons, which requires clarification. In cortical maps the represented microfeatures are correlated most closely with the location of the neuronal cell body, which often interacts with nearby neurons. Therefore, when a cortical map is treated mathematically as a field, there are several physical quantities that can be interpreted as the field's value ϕ_u at a particular cortical location u. Although the choice depends somewhat on the purpose of the analysis, the most common interpretation of $\phi_u(t)$ will be the instantaneous spiking frequency at time t of the neuron at location u. We will refer to $\phi(t) \in \Phi(\Omega)$ as the *neuronal field* (at time t) associated with the neurons u in the map Ω.

The relative phase of neural impulses is sometimes relevant to neural information processing [25]. For example, the relative phase with which action potentials arrive

a neuron's dendrites can determine whether or not the induced post-synaptic potentials add. In these cases it may be convenient to treat neural activity as a complex-valued field, $\psi(t) \in \Phi_{\mathbb{C}}(\Omega)$, which can be written in polar form:

$$\psi(t) = \rho(t) \mathrm{e}^{i\phi(t)} .$$

Then the magnitude (or modulus) field $\rho(t)$ may represent the impulse rate and the phase (or argument) field $\phi(t)$ may represent the relative phase of the impulses. That is, $\rho_u(t)$ is the rate of neuron u (at time t) and $\phi_u(t)$ is its phase. For example, in a *bursting neuron* (which generates impulses in clusters), $\rho(t)$ can represent the impulse rate within the clusters and $\phi(t)$ the relative phase of the clusters. More generally, in a complex-valued neuronal field, the phase part may represent microfeatures of stimulus, while the magnitude part represent pragmatic characteristics of the microfeatures, such as their importance, confidence, or urgency. (Such dual representations, comprising semantics and pragmatics, are discussed in Sect. "Information Fields".)

Synaptic and Dendritic Fields

The surface of each neuron's dendritic tree and soma (cell body) is a complicated two-dimensional manifold Ω_{m}, and so the electrical field across the neuron's membrane is naturally treated as a two-dimensional potential field $\phi \in \Phi(\Omega_{\mathrm{m}})$. Synaptic inputs create electrical disturbances in this field, which, to a first approximation, propagate passively according to the cable equations (see pp. 25–31 in [3]). However, there are also nonlinear effects due to voltage-gated ion channels etc. (see p. 381 in [58]). Therefore the membrane field obeys a nonlinear PDE (partial differential equation) dependent on a synaptic input field ϵ:

$$M(\epsilon, \phi, \dot{\phi}, \ddot{\phi}, \dots) = 0 .$$

The electrical field ϕ on the membrane includes the field ϕ_a at the axon hillock $a \in \Omega_{\mathrm{m}}$. This voltage determines the rate at which the neuron generates action potentials (APs, nerve impulses), which constitute the neuron's contribution to a neuronal field. The dependence of the impulse rate r on the membrane field, $r(t) = A_r[\phi(t)]$, which is approximately linear (that is, the rate is proportional to the depolarization, relative to the resting potential, at the axon hillock). To a first approximation, the dendritic tree implements an approximately linear (adaptive) analog filter on its input field [35,36]. Some purposes require a more detailed analysis, which looks at the time-varying action potential $V(t)$, rather than at the instantaneous impulse rate, as a function of the membrane field, $V(t) = A_V[\phi(t)]$.

Many neurons have tens of thousands of synaptic inputs (see p. 304 in [3]), and so these quantitative properties can be treated as a field over a domain Ω, which is a subset of the dendritic membrane. The post-synaptic potential ϵ_s at synapse s is a result of the synaptic efficacy σ_s and the pre-synaptic axonal impulse rate ζ_s. The synaptic efficacy is the composite effect of the number of receptors for the neurotransmitter released by the incoming axon, as well as of other factors, such as the dependence of neurotransmitter flux on the impulse rate. Some learning processes (e. g., long-term potentiation) alter the synaptic efficacy field σ.

However, because synaptic transmission involves the diffusion of neurotransmitter molecules across the synaptic cleft, the subsequent binding to and unbinding from receptors, and the opening and closing of ion channels, the post-synaptic potential is not a simple product, $\epsilon_s = \sigma_s \zeta_s$. Rather, the synaptic system filters the input field. To a first approximation we may analyze it as a linear system S:

$$\left(\begin{array}{c} \epsilon \\ \dot{\psi} \end{array} \right) = \mathrm{S}(\sigma) \left(\begin{array}{c} \zeta \\ \psi \end{array} \right) ,$$

where ψ represents the internal state of the synaptic system (concentrations of neurotransmitter in the clefts, receptor and ion channel states, etc.). The parameter σ shows the system's dependence on the synaptic efficacies. The preceding equation is an abbreviation for the following system (in which we suppress the σ parameter):

$$\epsilon = S_{11}\zeta + S_{12}\psi ,$$
$$\dot{\psi} = S_{21}\zeta + S_{22}\psi ,$$

in which the products are Hilbert–Schmidt integral operators (that is, the S_{ij} are fields operating on the input and state fields).

Axonal Projection Fields

Bundles of axons form *projections* from one brain region to another; through the pattern of their connections they may effect certain field transformations (explained below). The input is typically a neuronal field $\phi \in \Phi(\Omega)$ defined over the source region Ω. At their distal ends the axons branch and form synapses with the dendritic trees of the neurons in the destination region. Since each axon may form synapses with many destination neurons, and each neuron may receive synapses from many axons, it is convenient to treat all the synapses of the destination neurons as forming one large synaptic system S, where the synaptic efficacies σ_u range over all the synapses u in the destination region, $u \in \Omega'$. Correspondingly we can consider the field $\zeta \in \Phi(\Omega')$ of pre-synaptic inputs ζ_u to all of

these synapses. The axons and their synapses define an *axonal projection system* P, which is, to a first approximation, a linear system:

$$\left(\begin{array}{c} \zeta \\ \dot{\alpha} \end{array} \right) = \mathrm{P} \left(\begin{array}{c} \phi \\ \alpha \end{array} \right) ,$$

where α represents the internal state of the axonal projection system.

The function of axons is to transmit nerve impulses over relatively long distances with no change of amplitude or waveform. However, there is a transmission delay, and different axons in a projection may introduce different delays. Thus an axonal projection may change the phase relationships of the input field, in addition to introducing an overall delay. On the basis of our analysis the axonal projection as a linear system, we can express the Laplace transform Z of the pre-synaptic field $\zeta(t)$ in terms of the transfer function H^S of the projection and the Laplace transform Φ of the input field $\phi(t)$:

$$Z(s) = H^\mathrm{S}(s)\Phi(s)$$

(where s is the conjugate variable of time). Note that all the variables refer to fields, and so this equation means

$$Z_u(s) = \int_\Omega H^\mathrm{S}_{uv}(s)\Phi_v(s)\mathrm{d}v ,$$

where $H^\mathrm{S}_{uv}(s) \in \Phi_\mathbb{C}(\Omega' \times \Omega)$ is the (complex-valued) transfer function to synapse u from input neuron v. Since the effects of the axons are pure delays, the transfer function is imaginary:

$$H^\mathrm{S}_{uv}(s) = \exp(-i\Delta_{uv}s) ,$$

where Δ_{uv} is the delay imposed by the axon from neuron v to synapse u. Thus the delay field $\Delta \in \Phi(\Omega' \times \Omega)$ defines the effect of the axonal projection on the input field.

The system S comprising all the synapses of the destination neurons is also characterized by a transfer function $H^\mathrm{S}(s)$; that is, $E(s) = H^\mathrm{S}(s)Z(s)$, where $E(s)$ is the Laplace transform of the post-synaptic field $\epsilon(t)$. Therefore the combined effect of the axonal projection and the synapses is $E(s) = H^\mathrm{SP}(s)\Phi(s)$, where the composite transfer function is $H^\mathrm{SP}(s) = H^\mathrm{S}(s)H^\mathrm{P}(s)$. Note that this a field equation, which abbreviates

$$H^\mathrm{SP}_{uv}(s) = \int_\Omega pH^\mathrm{S}_{uw}(s)H^\mathrm{P}_{wv}\mathrm{d}w .$$

The transfer function $H^\mathrm{SP}_{uv}(s)$ has a corresponding *impulse response* $\eta^\mathrm{SP}_{uv}(t)$, which represents the post-synaptic response at u to a mathematical impulse (Dirac delta function) injected at v. (For Dirac delta functions, see Glossary

and Sect. "Approximation of Spatial Integral and Differential Operators".) The impulse response characterizes the effect of signal transmission to u from v as follows:

$$\epsilon_u(t) = \int_{\Omega'} \eta_{uv}^{\mathrm{SP}}(t) \otimes \phi_v(t)\mathrm{d}v \, ,$$

where "$\otimes$" represents convolution in the time domain. This may be abbreviated as a field equation, $\epsilon(t) = \eta^{\mathrm{SP}}(t) \otimes \phi(t)$.

Since axonal projections largely determine the *receptive fields* of the destination neurons, it will be worthwhile to consider the relation of the projection field to the neuronal field at the destination region. Therefore, let ψ_w represent the output signal of a destination neuron w in response to an input field ϕ. We may write

$$\psi_w(t) = F_w[\eta^{\mathrm{SP}}(t) \otimes \phi(t)],$$

where F_w represents the (possibly nonlinear) function computed by neuron w on the subset of the post-synaptic signal $\eta^{\mathrm{SP}}(t) \otimes \phi(t)$ in its dendritic tree. Therefore, the destination neuronal field is given by the field equation $\psi = F[\eta^{\mathrm{SP}} \otimes \phi]$. Many neurons behave as "leaky integrators" (see pp. 52–54 in [3]), which are approximately linear, and in these cases the combined effect of the axonal projection, synaptic field, and destination neurons is a linear operator applied to the input signal, $\psi(t) = L\phi(t)$.

Examples of Field Computation

Neural-Network-like Computation

Many neural network approaches to artificial intelligence can be adapted easily to field computation, effectively by taking the number of neurons in a layer to the continuum limit. For example, as discussed in Sect. "Field Transformations", $\psi = \bar{s}(L\phi + \beta)$ (Eq. 2) is the field analog of one layer of a neural net, that is, a continuum neural net, with interconnection field L and bias field β.

Discrete Basis Function Networks   *Radial basis function (RBF) networks* are a familiar and useful class of artificial neural networks, which have similarities to neural networks in the brain [29,52]. Indeed, RBF networks are inspired by the observation that many sensory neurons are tuned to a point in sensory space and that their response falls off continuously with distance from that central point (recall Sect. "Neuronal Fields"). RBFs are usually defined over finite-dimensional spaces, but the extension to fields is straight-forward. Therefore we will consider a set of functionals $r_1, r_2, \ldots$, where $r_j \colon \Phi(\Omega) \to [0, 1]$. Typically we restrict our attention to finite sets of basis functionals,

but we include the infinite case for generality. The intent is that each r_j is tuned to a different field input η_j, its "focal field", and that $r_j(\phi)$ represents the closeness of ϕ to the focal field η_j.

If all the RBFs have the same *receptive field profile*, that is, the same fall-off of response with increasing distance from the focal field, then we can write $r_j(\phi) = r(\|\phi - \eta_j\|)$, where the receptive field profile is defined by a $r \colon [0, \infty) \to [0, 1]$ that is monotonically decreasing with $r(0) = 1$ and $r(x) \longrightarrow 0$ as $x \longrightarrow \infty$.

As is well known, the inner product is frequently used as a measure of similarity. Expanding the difference in terms of the inner product yields:

$$\|\phi - \eta_j\|^2 = \|\phi\|^2 - 2\langle \phi \mid \eta_j \rangle + \|\eta_j\|^2 \, .$$

The inverse relation between the inner product and distance is especially obvious if, as is often the case (see Sect. "Information Fields"), the input and focal fields are normalized ($\|\phi\| = 1 = \|\eta_j\|$); then:

$$\|\phi - \eta_j\|^2 = 2 - 2\langle \phi \mid \eta_j \rangle \, .$$

Therefore, RBFs with an identical fall-off of response can be defined in terms of a fixed function $c \colon [-1, 1] \to [0, 1]$ applied to the inner product, $r_j(\phi) = c(\langle \phi \mid \eta_j \rangle)$, where the monotonically increasing function c equals 1 when $\phi = \eta_j$ and equals 0 when the fields are maximally different ($\phi = -\eta_j$). That is, for normalized fields $\langle \phi \mid \eta_j \rangle \in [-1, 1]$, and so $c(-1) = 0$, $c(1) = 1$.

Such RBFs are closely related to familiar artificial neurons (Eq. 1). Indeed, we may define $r_j(\phi) = c(\langle \eta_j \mid \phi \rangle + b_j)$, where $c \colon \mathbb{R} \to [0, 1]$ is a sigmoidal activation function and b_j is the bias term. Here the input ϕ to the neuron is a field, as is its receptive field profile η_j, which is the focal field defined by the neuron's interconnection field.

Generally, neurons are quite broadly tuned, and so individual RBFs do not characterize the input very precisely, but with an appropriate distribution of focal fields the collection of RBFs can characterize the input accurately, a process known as *coarse coding* (e. g., pp. 91–96 in [58]; [59]). Therefore the discrete ensemble of RBFs compute a representation $\mathbf{p}(\phi)$ of the input given by $p_j(\phi) = r_j(\phi)$.

When information is represented in some way we must consider the adequacy of the representation for our information processing goals. In general, it is not necessary that a representation function $\mathbf{p}$ preserve all characteristics and distinctions of the input space; indeed often the function of representation is to extract the relevant features of the input for subsequent processing. Nevertheless it will be worthwhile to consider briefly RBF-like rep-

resentations that do not lose any information. A Hilbert function space is isomorphic (indeed, isometric) to the space ℓ_2 of square-summable sequences; that is, there is a one-to-one correspondence between fields and the infinite sequences of their generalized Fourier coefficients. Therefore let $\beta_1, \beta_2, \ldots$ be any orthonormal (ON) basis of $\Phi(\Omega)$ and define $p_j(\phi) = \langle \beta_j \mid \phi \rangle$. Define $\mathbf{p}: \Phi(\Omega) \to \ell_2$ so that $\mathbf{p}(\phi)$ is the infinite sequence of generalized Fourier coefficients, $(p_1(\phi), p_2(\phi), \ldots)$. Mathematically, we can always find an m such that the first m coefficients approximate the fields as closely as we like; practically, physically realizable fields are band-limited, and so they have only a finite number of nonzero Fourier coefficients. Therefore, we may use $\mathbf{p}^m: \Phi(\Omega) \to \mathbb{R}^m$ to compute the m-dimensional representations (relative to an understood ON basis):

$$\mathbf{p}^m(\phi) = (p_1(\phi), p_2(\phi), \ldots, p_m(\phi))^{\mathrm{T}}.$$

Continua of Basis Functions In the preceding section we looked at the field computation of a discrete, typically finite, set of basis functionals. This is appropriate when the basis elements are relatively few in number and there is no significant topological relation among them. In the brain, however, large masses of neurons typically have a significant topological relation (e. g., they may form a topographic map (Sect. "Neuronal Fields")), and so we are interested in cases in which each point in an output field ψ is a result of applying a different basis function to the input field. Suppose $\phi \in \Phi(\Omega)$ and $\psi \in \Phi(\Omega')$. For all $u \in \Omega'$ we want $\psi_u = R(u, \phi)$, where $R: \Omega' \times \Phi(\Omega) \to \Phi(\Omega')$. That is, R defines a family of functionals in which, for each u, $R(u, -)$ has a different focal field, which varies continuously with u.

For example, suppose we want ψ_u to be an innerproduct comparison of ϕ with the focal field η^u: $\psi_u = c(\langle \eta^u \mid \phi \rangle)$. Since $\langle \eta^u \mid \phi \rangle = \int_\Omega \eta^u_v \phi_v \mathrm{d}v$, define the field $H \in \Phi(\Omega' \times \Omega)$ by $H_{uv} = \eta^u_v$. Then a point in the output field is given by $\psi_u = c[(H\phi)_u]$, and the entire field is computed by:

$$\psi = \bar{c}(H\phi).$$

This is, of course, the field analog of one layer of a neural net (Eq. 2), but with no bias field. In a similar way we can define a continuum of RBFs: $\psi_u = r(\|\phi - \eta^u\|)$.

Spatial Correlation and Convolution A special case of Eq. (5) rises when all the focal fields η^u are the same shape but centered on different points $u \in \Omega$. That is, $\eta^u_v = \varrho(v - u)$, where $\varrho \in \Phi(\Omega)$ is the common shape of the

focal fields (their *receptive field profile*). In this case,

$$\langle \eta^u \mid \phi \rangle = \int_\Omega \varrho(v - u)\phi(v)\mathrm{d}v.$$

This is simply the *cross-correlation* of ϱ and ϕ, which we may write $\varrho \star \phi$. In general,

$$(\psi \star \phi)_u = \int_\Omega \psi(v - u)\phi(v)\mathrm{d}v, \tag{6}$$

which gives the correlation of ψ and ϕ at a relative displacement u. Therefore in this case the RBF field is given by $\psi = \bar{c}(\varrho \star \phi)$. If the receptive field ϱ is symmetric, $\varrho(-x) = \varrho(x)$, then

$$\langle \eta^u \mid \phi \rangle = \int_\Omega \varrho(u - v)\phi(v)\mathrm{d}v,$$

which is $\varrho \otimes \phi$, the *convolution* of ϱ and ϕ. In general,

$$(\psi \otimes \phi)_u = \int_\Omega \psi(u - v)\phi(v)\mathrm{d}v. \tag{7}$$

Hence $\psi = \bar{s}(\varrho \otimes \phi)$ when ϱ is symmetric. Computation of these fields by means of convolution or correlation rather than by the integral operator (Eq. 5) may be more convenient on field computers that implement convolution or correlation directly.

Approximation of Spatial Integral and Differential Operators Correlation and convolution (Eqs. 6, 7) can be used to implement many useful linear operators, in particular spatial integral and differential operators. Of course these linear operations can be implemented by a field product with the appropriate Hilbert–Schmidt kernel, but convolution and correlation make use of lower dimensional fields than the kernel.

For example, suppose we want to compute the indefinite spatial integral of a field $\phi \in \Phi(\mathbb{R})$. That is, we want to compute $\psi = \int \phi$ defined by $\psi_x = \int_{-\infty}^x \phi_y \mathrm{d}y$. This can be computed by the convolution $\psi = \upsilon \otimes \phi$ where υ is the *Heaviside* or *unit step field* on $\mathbb{R}$:

$$\upsilon_x = \begin{cases} 1 & \text{if } x \geq 0 \\ 0 & \text{if } x < 0 \end{cases}.$$

The Heaviside field is discontinuous, and therefore it may not be physically realizable, but obviously it may be approximated arbitrarily closely by a continuous field.

Spatial differentiation is important in image processing in nervous systems and artificial intelligence systems. In the one-dimesional case, for $\phi \in \Phi(\mathbb{R})$ we want $\phi' \in$

$\Phi(\mathbb{R})$, where $\phi'_u = \mathrm{d}\phi_u/\mathrm{d}u$. To express this as a convolution we may begin by considering the *Dirac delta function* or *unit impulse function* δ, which is the derivative of the unit step function, $\delta(x) = \upsilon'(x)$. This is a *generalized function* or *distribution* with the following properties:

$$\delta(0) = +\infty,$$
$$\delta(x) = 0, x \neq 0,$$
$$\int_{-\infty}^{+\infty} \delta(x)\mathrm{d}x = 1 .$$

Obviously such a function is not physically realizable (more on that shortly), but such functions exist as limit objects in Hilbert spaces. The Dirac delta satisfies the following "sifting property":

$$\phi_x = \int_{-\infty}^{+\infty} \delta(x - y)\phi(y)\mathrm{d}y ;$$

that is, the Dirac delta is an identity for convolution, $\phi = \delta \otimes \phi$. Now observe:

$$\phi'_x = \mathrm{D}_x \int_{-\infty}^{+\infty} \delta(x - y)\phi(y)\mathrm{d}y$$
$$= \int_{-\infty}^{+\infty} \delta'(x - y)\phi(y)\mathrm{d}y ,$$

where δ' is the derivative of the Dirac delta. It is called the *unit doublet* and has the property of being zero everywhere except infinitesimally to the left of the origin, where it is $+\infty$, and infinitesimally to the right of the origin, where it is $-\infty$. Thus the spatial derivative of a field can be computed by convolution with the unit doublet: $\phi' = \delta' \otimes \phi$.

Obviously, neither the unit impulse (Dirac delta) nor the unit doublet is physically realizable, but both may be approximated arbitrarily closely by physically realizable fields. For example, the delta function can be approximated by a sufficiently sharp Gaussian field γ (i. e., $\gamma_x = \sqrt{r/\pi} \exp(-rx^2)$ for sufficiently large r). Corresponding to the sifting property $\phi = \delta \otimes \phi$ we have Gaussian smoothing $\phi \approx \gamma \otimes \phi$, which is a typical effect of the limited bandwidth of physically realizable fields in cortex and other physical media. Similarly, the unit doublet can be approximated by a *derivative of Gaussian (DoG)* field γ', where $\gamma'_x = \mathrm{d}\gamma_x/\mathrm{d}x$. Thus, the spatial derivative can be approximated the convolution $\phi' \approx \gamma' \otimes \phi$. Indeed, in the nervous system we find neurons with approximately DoG receptive field profiles. (These derivative formulas are perhaps more intuitively expressed in terms of correlation, $\phi' = (-\delta') \star \phi \approx (-\gamma') \star \phi$, since this is more easily related to the difference, $\phi_{x+\epsilon} - \phi_{x-\epsilon}$.)

If ψ is a two-dimensional field, $\psi \in \Phi(\mathbb{R}^2)$, it is easy to show that the partial derivative along the first dimension can be computed by convolution with $\delta' \wedge \delta$, and along the second by convolution with $\delta \wedge \delta'$. The partial derivatives may be approximated by convolutions with $\gamma' \wedge \gamma$ and $\gamma \wedge \gamma'$. The divergence of a field can be computed by a two-dimensional convolution with the sum of these fields:

$$\nabla \cdot \psi = (\delta' \wedge \delta + \delta \wedge \delta') \otimes \psi \approx (\gamma' \wedge \gamma + \gamma \wedge \gamma') \otimes \psi .$$

Similarly the gradient is

$$\nabla \psi = [(\delta' \wedge \delta) \otimes \psi]\mathbf{i} + [(\delta \wedge \delta') \otimes \psi]\mathbf{j} ,$$

where $\mathbf{i} \in \Phi_{\mathbb{R}^2}(\mathbb{R}^2)$ is a constant vector field of unit vectors in the x direction, $\mathbf{i}_{(x,y)} = (1, 0)$, and $\mathbf{j}$ is a similar field in the y direction. It is approximated by

$$\nabla \psi \approx [(\gamma' \wedge \gamma) \otimes \psi]\mathbf{i} + [(\gamma \wedge \gamma') \otimes \psi]\mathbf{j} . \tag{8}$$

To compute the Laplacian we need the second partial derivatives, but note that for a one-dimensional field $\phi'' = \delta' \otimes (\delta' \otimes \phi) = (\delta' \otimes \delta') \otimes \phi = \delta'' \otimes \phi$, where δ'' is the second derivative of the Dirac function (a "unit triplet"). Hence, for two-dimensional ψ

$$\nabla^2 \psi = (\delta'' \wedge \delta + \delta \wedge \delta'') \otimes \psi \approx (\gamma'' \wedge \gamma + \gamma \wedge \gamma'') \otimes \psi, \tag{9}$$

where γ'' is the second derivative of the Gaussian, a typical (inverted) "Mexican hat function" with the center-surround receptive-field profile often found in the nervous system. These formulas extend in the obvious way to higher-dimensional fields.

Change of Field Domain

We have seen that physically realizable linear operators are integral operators, and therefore can be computed by field products of the form $K\phi$. However, the kernel K might not be physically realizable if its dimension is too high. For example, suppose $L: \Phi(\Omega) \to \Phi(\Omega)$ is a linear operator on two-dimensional visual images; that is, Ω is a bounded subset of two-dimensional Euclidean space. Its kernel K, satisfying $K\phi = L(\phi)$, will be a four-dimensional field $K \in \Phi(\Omega \times \Omega)$, and therefore physically unrealizable. Therefore we need means for realizing or approximating high-dimensional fields in three or fewer spatial dimensions.

The simplest way to accomplish this is to represent fields of higher-dimensional spaces by corresponding fields over lower dimensional spaces. For example, to represent $\phi \in \Phi(\Omega)$ by $\psi \in \Phi(\Omega')$, suppose $\beta_1, \beta_2, \ldots$ is an ON basis for $\Phi(\Omega)$, as $\eta_1, \eta_2, \ldots$ is for $\Phi(\Omega')$. Then,

let the generalized Fourier coefficients of ϕ be used as the coefficients to compute a corresponding ψ. Observe:

$$\psi = \sum_k \eta_k \langle \beta_k \mid \phi \rangle = \sum_k (\eta_k \wedge \beta_k) \phi \; .$$

(Of course, a finite sum is sufficient for physically realizable fields.) Therefore the change of basis can be implemented by the kernel $K = \sum_k \eta_k \wedge \beta_k$. By this means, any Hilbert–Schmidt operator on two-dimensional fields can be implemented by a physically realizable field product: represent the input by a one-dimensional field, generate the one-dimensional representation of the output by a product with a two-dimensional kernel, and convert this representation to the output field. Specifically, suppose $\phi \in \Phi(\Omega)$, $\psi \in \Phi(\Omega')$, and $L: \Phi(\Omega) \to \Phi(\Omega')$ is a Hilbert–Schmidt linear operator. The three-dimensional kernel $H = \sum_k \eta_k \wedge \beta_k \in \Phi([0,1] \times \Omega)$ will be used to generate a one-dimensional representation of the two-dimensional input, $H\phi \in \Phi([0,1])$. Similarly, the two-dimensional output will be generated by $\Theta = \sum_j \zeta_j \wedge \eta_j \in \Phi(\Omega' \times [0,1])$, where $\zeta_1, \zeta_2, \dots$ is an ON basis for $\Phi(\Omega')$. It is easy to show that the required two-dimensional kernel $K \in \Phi([0,1]^2)$ such that $L = \Theta K H$ is just

$$K = \sum_{jk} \langle \zeta_j \mid L\beta_k \rangle (\eta_j \wedge \eta_k) \; .$$

We have seen (see Sect. "Neural-Network-Like Computation") that field computation can often be implemented by neural-network-style computation on finite-dimensional spaces. For example, a linear field transformation (of Hilbert–Schmidt type) can be factored through the *eigenfield basis*. Let $\epsilon_1, \epsilon_2, \dots$ be the eigenfields of L with corresponding eigenvalues $e_1, e_2, \dots$: $L\epsilon_k = e_k \epsilon_k$. The eigenfields can be chosen to be orthonormal (ON), and, since $\Phi(\Omega)$ is a Hilbert space, only a finite number of the eigenvalues are greater than any fixed bound, so ϕ can be approximated arbitrarily closely by a finite sum $\phi \approx \sum_{k=1}^m c_k \epsilon_k$, where $c_k = \langle \epsilon_k \mid \phi \rangle$; that is, ϕ is represented by the finite-dimensional vector $\mathbf{c}$. The discrete set of coefficients $c_1, \dots, c_m$ is not a field because there is no significant topological relationship among them; also, typically, m is relatively small.

The output ψ is computed by a finite sum, $\psi \approx \sum_{k=1}^m \epsilon_k e_k c_k$. In terms of neural computation, we have a finite set of neurons $k = 1, \dots, m$ whose receptive field profiles are the eigenfields, so that they compute $e_k c_k = e_k \langle \epsilon_k \mid \phi \rangle$. The outputs of these neurons amplitude-modulate the generation of the individual eigenfields ϵ_k, whose superposition yields the output ψ.

It is not necessary to factor the operator through the eigenfield basis. To see this, suppose $L: \Phi(\Omega) \to \Phi(\Omega')$

and that the fields β_k are an ON basis for $\Phi(\Omega)$ and that the fields ζ_j are an ON basis for $\Phi(\Omega')$. Represent the input by a finite-dimensional vector $\mathbf{c}$, where $c_k = \langle \beta_k \mid \phi \rangle$. Then the output ψ can be represented by the finite-dimensional vector $\mathbf{d}$, where $d_j = \langle \zeta_j \mid \psi \rangle$. (Since the input and output spaces are both Hilbert spaces, only a finite number of these coefficients are greater than any fixed bound.) It is easy to show $\mathbf{d} = M\mathbf{c}$, where $M_{jk} = \langle \zeta_j \mid L\beta_k \rangle$ (the Hilbert–Schmidt theorem). In neural terms, a first layer of neurons with receptive field profiles β_k compute the discrete representation $c_k = \langle \beta_k \mid \phi \rangle$. Next, a layer of linear neurons computes the linear combinations $d_j = \sum_{k=1}^m M_{jk} c_k$ in order to control the amplitudes of the output basis fields in the output superposition $\psi \approx \sum_{j=1}^n d_j \zeta_j$. In this way, an arbitrary linear field transformation may be computed through a neural representation of relatively low dimension.

If a kernel has too high dimension to be physically realizable, it is not necessary to completely factor the product through a discrete space; rather, one or more dimensions can be replaced by a discrete set of basis functions and the others performed by field computation. To see the procedure, suppose we have a linear operator $L: \Phi(\Omega) \to \Phi(\Omega')$ with kernel $K \in \Phi(\Omega' \times \Omega)$, where $\Omega = \Omega_1 \times \Omega_2$ is of too great dimension. Let $\psi = K\phi$ and observe

$$\psi_u = \int_\Omega K_{uv} \phi_v \mathrm{d}v = \int_{\Omega_1} \int_{\Omega_2} K_{uxy} \phi_x(y) \mathrm{d}y \mathrm{d}x \; ,$$

where we consider $\phi_v = \phi_{xy}$ as a function of y, $\phi_x(y)$. Expand ϕ_x in terms of an ON basis of $\Phi(\Omega_2)$, $\beta_1, \beta_2, \dots$:

$$\phi_x = \sum_k \langle \phi_x \mid \beta_k \rangle \beta_k \; .$$

Note that

$$\langle \phi_x \mid \beta_k \rangle = \int_{\Omega_2} \phi_{xy} \beta_k(y) \mathrm{d}y = (\phi\beta_k)_x \; ,$$

where $\phi\beta_k \in \Phi(\Omega_1)$. Rearranging the order of summation and integration,

$$\begin{aligned}
\psi_u &= \sum_k \int_{\Omega_1} \int_{\Omega_2} K_{uxy} \beta_k(y) (\phi\beta_k)_x \mathrm{d}y \mathrm{d}x \\
&= \sum_k [K\beta_k(\phi\beta_k)]_u \; .
\end{aligned}$$

Hence, $\psi = \sum_k K\beta_k(\phi\beta_k)$. Let $J_k = K\beta_k$ to obtain a lower-dimensional field computation:

$$L(\phi) = \sum_k J_k(\phi\beta_k) \; .$$

Note that $J_k \in \Phi(\Omega' \times \Omega_1)$ and all the other fields are of lower dimension than $K \in \Phi(\Omega' \times \Omega)$. As usual, for physically realizable fields, a finite summation is sufficient.

We can discretize $\Phi(\Omega_1)$ by a similar process, which also can be extended straightforwardly to cases where several dimensions must be discretized. Normally we will discretize the dimension that will have the fewest generalized Fourier coefficients, given the bandwidth of the input fields.

The foregoing example discretized one dimension of the input space, but it is also possible to discretize dimensions of the output space. Therefore suppose $L\colon \Phi(\Omega) \to \Phi(\Omega')$ with kernel $K \in \Phi(\Omega' \times \Omega)$, where $\Omega' = \Omega_1 \times \Omega_2$ is of too great dimension. Suppose $\zeta_1, \zeta_2, \ldots$ are an ON basis for $\Phi(\Omega_1)$. Consider $\psi_u = \psi_{xy}$ as a function of x, expand, and rearrange:

$$
\begin{aligned}
\psi_{xy} &= \sum_k \zeta_k(x) \int_{\Omega_1} \zeta_k(x') \psi_{x'y} \, dx' \\
&= \sum_k \zeta_k(x) \int_{\Omega} \int_{\Omega_1} \zeta_k(x') K_{x'yv} \, dx' \phi_v \, dv \\
&= \sum_k \zeta_k(x) [(\zeta_k K)\phi]_y .
\end{aligned}
$$

Hence $\psi = \sum_k \zeta_k \wedge [(\zeta_k K)\phi]$. Let $J_k = \zeta_k K \in \Phi(\Omega_2 \times \Omega)$ and we can express the computation with lower dimensional fields:

$$
L(\phi) = \sum_k \zeta_k \wedge J_k \phi .
$$

Other approaches to reducing the dimension of fields are described elsewhere [33].

The converse procedure, using field computation to implement a matrix vector product, is also useful, since a field computer may have better facilities for field computation than for computing with vectors. Therefore suppose M is an $m \times n$ matrix, $\mathbf{c} \in \mathbb{R}^n$, and that we want to compute $\mathbf{d} = \mathrm{M}\mathbf{c}$ by a field product $\psi = K\phi$. The input vector will be represented by $\phi \in \Phi(\Omega)$, where we choose a field space $\Phi(\Omega)$ for which the first n ON basis elements $\beta_1, \ldots, \beta_n$ are physically realizable. The field representation is given by $\phi = \sum_{k=1}^n c_k \beta_k$. Analogously, the output is represented by a field $\psi \in \Phi(\Omega')$ given by $\psi = \sum_{j=1}^m d_k \zeta_k$, for ON basis fields $\zeta_1, \ldots, \zeta_m$. The required kernel $K \in \Phi(\Omega' \times \Omega)$ is given by

$$
K = \sum_{j=1}^m \sum_{k=1}^n M_{ij}(\zeta_j \wedge \beta_k) .
$$

To see this, observe:

$$
\begin{aligned}
K\phi &= \sum_{jk} M_{jk}(\zeta_j \wedge \beta_k)\phi \\
&= \sum_{jk} M_{jk} \zeta_j \langle \beta_k \mid \phi \rangle \\
&= \sum_j \zeta_j \sum_k M_{jk} c_k \\
&= \sum_j \zeta_j d_j .
\end{aligned}
$$

Diffusion Processes

Diffusion processes are useful in both natural and artificial intelligence. For example, it has been applied to path planning through a maze [66] and to optimization and constraint-satisfaction problems, such as occur in image processing and motion estimation [45,67]. Natural systems, such as developing embryos and colonies of organisms, use diffusion as a means of massively parallel search and communication.

A simple diffusion equation has the form $\dot{\phi} = d\nabla^2\phi$ with $d > 0$. On a continuous-time field computer that provides the Laplacian operator (∇^2) diffusion can be implemented directly by this equation. With sequential computation, the field will be iteratively updated in discrete steps:

$$
\phi := \phi + d\nabla^2\phi .
$$

If the Laplacian is not provided as a primitive operation, then its effect can be approximated by a spatial convolution with a suitable field ϱ (see Sect. "Spatial Correlation and Convolution"). In sequential computation we may use $\phi := (1 - d)\phi + d\varrho \otimes \phi$, where ϱ is an appropriate two-dimensional Gaussian or similarly shaped field. In continuous time, we may use $\dot{\phi} = d\varrho \otimes \phi$, where $\varrho = \gamma'' \wedge \gamma + \gamma \wedge \gamma''$ (Eq. (9), Sect. "Approximation of Spatial Integral and Differential Operators"), where γ is an appropriate one-dimensional Gaussian and γ'' is its second derivative (or similarly shaped fields).

Reaction-diffusion systems combine diffusion in two or more fields with local nonlinear reactions among the fields ▶ Reaction-Diffusion Computing. A typical reaction-diffusion system over fields $\phi^1, \ldots, \phi^n \in \Phi(\Omega)$ has the form:

$$
\begin{aligned}
\dot{\phi}^1 &= \overline{F_1}(\phi^1, \ldots, \phi^n) + d_1\nabla^2\phi^1, \\
\dot{\phi}^2 &= \overline{F_2}(\phi^1, \ldots, \phi^n) + d_2\nabla^2\phi^2, \\
&\;\;\vdots \\
\dot{\phi}^n &= \overline{F_n}(\phi^1, \ldots, \phi^n) + d_n\nabla^2\phi^n ,
\end{aligned}
$$

where the $d_k > 0$, and the local reactions $\overline{F_k}$ apply at each point $u \in \Omega$ of the fields: $F_k(\phi_u^1, \ldots, \phi_u^n)$. With obvious extension of the notation, this can be written as a differential equation on a vector field:

$$\dot{\boldsymbol{\phi}} = \overline{\mathbf{F}}(\boldsymbol{\phi}) + \mathrm{D}\nabla^2\boldsymbol{\phi},$$

where $\mathrm{D} = \mathrm{diag}(d_1, \ldots, d_n)$ is a diagonal matrix of diffusion rates.

Embryological development and many other biological processes of self-organization are controlled by local reaction to multiple diffusing chemicals (e. g., Chap. 7 in Bar-Yam [5], Chap. 3 in Solé and Goodwin [64]); these are examples of natural field computation, a subject pioneered by AM Turing [71]. For example, simple *activator-inhibitor systems* can generate *Turing patterns*, which are reminiscent of animal skin and hair-coat pigmentation patterns (e. g., Chap. 7 in Bar-Yam [5]). In the simplest case, these involve an activator (α) and an inhibitor (β), which diffuse at different rates, and a nonlinear interaction which increases both when $\alpha > \beta$, and decreases them otherwise (for example p. 668 in [5]):

$$\dot{\alpha} = \frac{k_1\alpha^2}{\beta(1 + k_5\alpha^2)} - k_2\alpha + d_\alpha\nabla^2\alpha,$$
$$\dot{\beta} = k_3\alpha^2 - k_4\beta + d_\beta\nabla^2\beta.$$

Reaction-diffusion systems have been applied experimentally in several image-processing applications, where they have been used to restore broken contours, detect edges, and improve contrast (pp. 26–31 in [1]). In general, diffusion accomplishes (high-frequency) noise filtering and the reaction is used for contrast enhancement.

A Adamatzky and his colleagues have used chemical implementation of reaction-diffusion systems to construct Voronoi diagrams around points and other two-dimensional objects (see Chap. 2 in [2]). Voronoi diagrams have been applied to collision-free path planning, nearest-neighbor pattern classification, and many other problems (see pp. 32–33 in [2]). They also demonstrated a chemical field computer on a mobile robot to implement a reaction-diffusion path planning (see Chap. 4 in [2]).

Excitable media are an important class of reaction-diffusion system, which are found, for example, in the brain, cardiac tissue, slime mold aggregation, and many other natural systems. In the simplest cases these comprise an *excitation field* ϵ and a *recovery field* ρ coupled by local nonlinear reactions:

$$\dot{\epsilon} = \overline{F}(\epsilon, \rho) + d_\epsilon\nabla^2\epsilon,$$
$$\dot{\rho} = \overline{G}(\epsilon, \rho) + d_\rho\nabla^2\rho.$$

Typically $G(e, r)$ is positive for large e and negative for large r, while along the *nullcline* $F(e, r) = 0$, r has a roughly cubic dependence on e, with $F(e, r) < 0$ for large values of r and > 0 for small ones. The intersection of the nullclines defines the system's stable state, and small perturbations return to the stable state. However excitation above a threshold will cause the excitation to increase to a maximum, after which the system becomes first refractory (unexcitable), then partially excitable with an elevated threshold, and finally back to its excitable, resting state. Excitation spreads to adjacent regions, but the refractory property assures that propagation takes the form of a unidirectional wave of constant amplitude. Characteristic circular and spiral waves appear in two-dimensional media. Excitable media are useful for rapid, efficient communication. For example, masses of slime mold amoebas (*Dictyostelium discoideum*) act as an excitable medium in which the propagating waves accelerate aggregation of the amoebas into a mound (see pp. 12–24 in [64]).

Many self-organizing systems and structures in biological systems involve reaction-diffusion processes, chemical gradients, excitable media, and other instances of field computation.

For example, J-L Deneubourg [15] has described the construction of equally-spaced pillars in termite nests in terms of three interrelated two-dimensional fields: ϕ, the concentration of cement pheromone in the air, σ, the amount of deposited cement with active pheromone, and τ the density of termites carrying cement (see also pp. 188–193 in [8], pp. 399–400 in [10], and pp. 151–157 in [64]). The amount of deposited cement with pheromone increases as it is deposited by the termites and decreases as the pheromone evaporates into the air: $\dot{\sigma} = k_1\tau - k_2\sigma$. The pheromone in the air is increased by this evaporation, but also decays and diffuses at specified rates: $\dot{\phi} = k_2\sigma - k_4\phi + d_\phi\nabla^2\phi$. Laden termites enter the system at a uniform rate r, deposit their cement (k_1), wander a certain amount (modeled by diffusion at rate d_τ), but also exhibit *chemotaxis*, that is, motion up the gradient of pheromone concentration:

$$\dot{\tau} = r - k_1\tau + d_\tau\nabla^2\tau - k_5\nabla \cdot (\tau\overline{\times}\nabla\phi),$$

where $\overline{\times}$ represents the point-wise (local) product, $(\phi\overline{\times}\psi)_u = \phi_u\psi_u$. See Fig. 1 for this model expressed as a field computation.

In addition to reaction-diffusion systems, chemical gradients, chemotaxis, and other field processes are essential to self-organization in morphogenesis, which can be understood in terms of field computation [14].
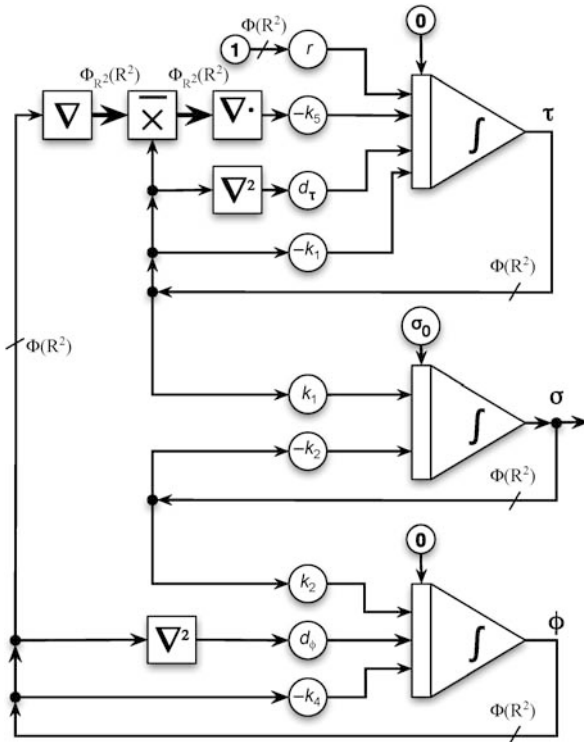
Field Computation in Natural and Artificial Intelligence, Figure 1
Field computation of Deneubourg's model of pillar construction by termites

Motion in Direction Fields

For an example of field computation in motor control, we may consider Georgopoulos' [21] explanation of the population coding of direction. In this case the feature space $\mathcal{D}$ represents directions in three-dimensional space, which we may identify with normalized three-dimensional vectors $\mathbf{d} \in \mathcal{D}$. Each neuron $u \in \Omega$ has a preferred direction $\boldsymbol{\eta}_u \in \mathcal{D}$ to which it responds most strongly, and it is natural to define u as the location in the map corresponding to this direction, $u = \mu(\boldsymbol{\eta}_u)$. However, Georgopoulos has shown that the direction is represented (more accurately and robustly) by a population code, in which the direction is represented by a neuronal field. Specifically, the activity ϕ_u of a neuron (above a base level) is proportional to the cosine of the angle between its preferred direction $\boldsymbol{\eta}_u$ and the direction $\mathbf{d}$ to be encoded. In particular, since the cosine of the angle between normalized vectors is equal to their scalar product, $\phi_u \propto \mathbf{d} \cdot \boldsymbol{\eta}_u$. A neurally plausible way of generating such a field is with a layer of radial basis functions (see Sect. "Continua of Basis Functions"), $\phi_u = r(\|\mathbf{d} - \boldsymbol{\eta}_u\|)$, where $r(x) = 1 - x^2/2$; then $\phi_u = \mathbf{d} \cdot \boldsymbol{\eta}_u$ [39].

Field computation is also used to update direction fields in the brain. For example, a remembered two-dimensional location, relative to the retina, must be updated when the eye moves [17,18]. In particular, if the direction field ϕ has a peak representing the remembered direction, and the eye moves in the direction $\mathbf{v}$, then this peak has to move in the direction $-\mathbf{v}$ in compensation. More specifically, if $\mathbf{v}$ is a two-dimensional vector defining the direction of eye motion, then the change in the direction field is given by the differential field equation, $\dot{\phi} = \mathbf{v} \cdot \nabla\phi$, where the gradient is a two-dimensional vector field (retinal coordinates). (That is, $\partial\phi(\mathbf{d}, t)/\partial t = \mathbf{v} \cdot \nabla_{\mathbf{d}}\phi(\mathbf{d}, t)$.) To see this, note that *behind* the moving peak $\nabla\phi$ and $-\mathbf{v}$ point in the same direction, and therefore $(-\mathbf{v}) \cdot \nabla\phi$ is positive; hence $\dot{\phi}$ is negative. Conversely, $\dot{\phi}$ is positive in front of the peak. Each component of the gradient may be approximated by convolution with a derivative-of-Gaussian (DoG) field, in accord with Eq. 8, which can be computed by neurons with DoG receptive field profiles. (Additional detail can be found elsewhere [39].)

RA Anderson [4] describes how transformations between retinal coordinates and head- or body-centered coordinates can be understood as transformations between field representations in area 7a of the posterior parietal cortex. For example, a minimum in a field may represent the destination of a motion (such as a saccade) in head-centered space, and then the gradient represents paths from other locations to that destination [39]. Further, the effects of motor neurons often correspond to vector fields [6,22].

Nonlinear Computation via Topographic Maps

As discussed in Sects. "Neuronal Fields" and "Motion in Direction Fields", the brain often represents scalar or vector quantities by topographic or computational maps, in which fields are defined over the range of possible values and a particular value is represented by a field with a peak of activity at the corresponding location. That is, a value $x \in \Omega$ is represented by a field $\phi_x \in \Phi(\Omega)$ that is distinctly peaked at x. For mathematical convenience we can idealize ϕ_x as a Dirac delta function (unit impulse) centered at x: δ_x, where $\delta_x(u) = \delta(u - x)$. That is, δ_x is an idealized topographic representation of x.

For every function $f: \Omega \to \Omega'$, with $y = f(x)$, there is a corresponding linear transformation of a topographic representation of its input, $\delta_x \in \Phi(\Omega)$, into a topographic representation of its output, $\delta_y \in \Phi(\Omega')$. It is easy to show that the kernel $K \in \Phi(\Omega' \times \Omega)$ of this operation is

$$K = \int_\Omega \delta_{f(x)} \wedge \delta_x \, dx \, ,$$

which is essentially a *graph* of the function f. That is, we can compute an arbitrary, possibly *nonlinear* function $y = f(x)$ by a *linear* operation on the corresponding computational maps, $\delta_y = K\delta_x$.

To avoid the use of Dirac delta functions, we can expand them into generalized Fourier series; for example, $\delta_x = \sum_k \beta_k \langle \beta_k \mid \delta_x \rangle = \sum_k \beta_k \beta_k(x)$. This expansion yields

$$K = \int_\Omega \left(\sum_j \zeta_j \zeta_j[f(x)] \right) \wedge \left(\sum_k \beta_k \beta_k(x) \right) dx$$

$$= \sum_{j,k} \zeta_j \wedge \beta_k \int_\Omega \zeta_j[f(x)]\beta_k(x)dx$$

$$= \sum_{j,k} \zeta_j \wedge \beta_k \langle \zeta_j \circ f \mid \beta_k \rangle,$$

where $\zeta_j \circ f$ is the composition of ζ_j and f: $(\zeta_j \circ f)(x) = \zeta_j[f(x)]$. A physically realizable approximation to K is obtained by limiting the summations to finite sets of physically realizable basis functions. (This has the effect of blurring the graph of f.)

Computation on topographic maps has a number attractive advantages. These are simple mathematical consequences of the linearity of topographic computation, but it will be informative to look at their applications in neural information processing. For example, transformation of input superpositions compute superpositions of the corresponding outputs in parallel: $K(\delta_x + \delta_{x'}) = \delta_{f(x)} + \delta_{f(x')}$ (recall Sect. "Neuronal Fields").

Since an input value is encoded by the position of the peak of a field rather than by its amplitude, the amplitude can be used for pragmatic characteristics of the input, such as its importance or certainty (see Sect. "Information Fields"). These pragmatic characteristics are preserved by topographic computation, $K(p\delta_x) = p\delta_{f(x)}$. Therefore if we have two (or more) inputs $x, x' \in \Omega$ with corresponding pragmatic scale factors $p, p' \in \mathbb{R}$, then the corresponding outputs carry the same factors, $K(p\delta_x + p'\delta_{x'}) = p\delta_{f(x)} + p'\delta_{f(x')}$. For example, if the inputs are weighted by confidence or importance, then the corresponding outputs will be similarly weighted. Further, if several inputs generate the same output, then their pragmatic scale factors will sum; for example if $f(x) = f(x')$, then $K(p\delta_x + p'\delta_{x'}) = (p + p')\delta_{f(x)}$. Thus, a number of inputs that are individually relatively unimportant (or uncertain) could contribute to a single output that is relatively important (or certain).

Finite superpositions of inputs are easily extended to the continuum case. For example, suppose that ϕ_x is the pragmatic scale factor associated with x, for all $x \in \Omega$

(for example, ϕ_x might be the probability of input x). We can think of the field ϕ as a continuum of weighted delta functions, $\phi = \int_\Omega \phi_x \delta_x dx$. Applying the kernel to this field yields a corresponding continuum of weighted outputs, $K\phi = \int_\Omega \phi_x \delta_{f(x)} dx \in \Phi(\Omega')$, where each point of the output field gives the total of the pragmatic scale factors (e. g., probabilities) of the inputs leading to the corresponding output value:

$$(K\phi)_y = \int\limits_{\{x \mid y = f(x)\}} \phi_x dx .$$

Therefore, by topographic computation, a transformation of an input probability distribution yields the corresponding output probability distribution.

We have remarked that the brain often uses *coarse coding*, in which a population of broadly-tuned neurons collectively represent a value with high precision (see Sect. "Discrete Basis Function Networks"). If ϕ is the coarse coding of input x, then its maximum will be at x and its amplitude will decrease with distance from x, $\phi_u = r(\|u - x\|)$. Similarly, $K\phi$ will be a coarse coding of the output $f(x)$ induced by the coarse coding of the input. As discussed in Sect. "Spatial Correlation and Convolution", if all the neurons have the same receptive field profile ϱ, then the effect of coarse coding is a convolution or correlation of ϱ with the input map.

Gabor Wavelets and Coherent States

In 1946 D Gabor presented a theory of information based on application to arbitrary signals of the Heisenberg–Weyl derivation of the quantum mechanical Uncertainty Principle [20]. Although he derived it for functions of time, it is easily generalizable to fields (square-integrable functions) over any finite-dimensional Euclidean space (reviewed elsewhere [34]). Therefore, for $\Omega \subset \mathbb{R}^n$, let $\psi \in \Phi(\Omega)$ be an arbitrary (possible complex-valued) field (assumed, as usual, to have a finite norm, that is, to be square-integrable; see Sect. "Mathematical Definitions"). To characterize this field's locality in space, we can measure its spread (or uncertainty) along each of the n spatial dimensions x_k by the root mean square deviation of x_k (assumed to have 0 mean):

$$\Delta x_k = \|x_k \psi(\mathbf{x})\| = \sqrt{\int_\Omega \psi_\mathbf{x}^* x_k^2 \psi_\mathbf{x} d\mathbf{x}} ,$$

where $\mathbf{x} = (x_1, \ldots, x_n)^\mathrm{T} \in \mathbb{R}^n$. Consider also the Fourier transform $\Psi(\mathbf{u})$ of $\psi(\mathbf{x})$, the spread or uncertainty of which, in the frequency domain, can be quantified in

a similar way:

$$\Delta u_k = \|(u_k - \bar{u})\Psi(\mathbf{u})\| = \sqrt{\int_\Omega \Psi_\mathbf{u}^* u_k^2 \Psi_\mathbf{u} d\mathbf{u}} \ .$$

It is straight-forward to show that the joint localization in any two conjugate variables (i. e., x_k in the space domain and u_k in the spatial-frequency domain) is limited by the *Gabor Uncertainty Principle*: $\Delta x_k \Delta u_k \geq 1/4\pi$.

This principle limits the information carrying capacity of any physically-realizable signal, so it is natural to ask if any function achieves the theoretical minimum, $\Delta x_k \Delta u_k = 1/4\pi$. Gabor showed that this minimum is achieved by what we may call the *Gabor elementary fields*, which have the form:

$$\Gamma_{\mathbf{pu}}(\mathbf{x}) = \exp\left[-\pi\|A(\mathbf{x}-\mathbf{p})\|^2\right]\exp[2\pi i\mathbf{u}\cdot(\mathbf{x}-\mathbf{p})] \ .$$

The second, imaginary exponential defines a plane wave originating at $\mathbf{p}$ with a frequency and direction determined by the *wave vector* $\mathbf{u}$. The first exponential defines a Gaussian envelope centered at $\mathbf{p}$ with a shape determined by the diagonal *aspect matrix* $A = \text{diag}(\alpha_1, \ldots, \alpha_n)$, which determines the spread of the function along each of the space and frequency axes:

$$\Delta x_k = \frac{\alpha_k}{2\sqrt{\pi}} \ , \quad \Delta u_k = \frac{\alpha_k^{-1}}{2\sqrt{\pi}} \ .$$

Gaussian-modulated complex exponentials of this form correspond to the *coherent states* of quantum mechanics.

Each Gabor elementary field defines a cell in $2n$-dimensional "Gabor space" with volume $(4\pi)^{-n}$. He explained that these correspond to elementary units of information, which he called *logons*, since a field of finite spatial extent and bandwidth occupies a finite region in Gabor space, which determines its *logon content*. It may be computed by

$$N = \prod_{k=1}^n \frac{X_k}{\Delta x_k}\frac{U_k}{\Delta u_k} = (4\pi)^n \prod_{k=1}^n X_k U_k \ ,$$

where X_k is the width of the field along the kth axis, and U_k its bandwidth on that axis, that is, a field's logon content is $(4\pi)^n$ times its Gabor-space volume.

The set of Gabor elementary functions are complete, and so any finite-energy function can be expanded into a series (see pp. 656–657 in [24]): $\psi = \sum_{k=1}^N c_k \Gamma_k$, where $\Gamma_1, \ldots, \Gamma_N$ are the Gabor fields corresponding to the cells occupied by ψ, and the c_k are complex coefficients. These N complex coefficients are the information conveyed by ψ, each corresponding to a logon or degree of freedom in the signal.

The Gabor elementary functions are not orthogonal, and so the coefficients cannot be computed by the inner product, $\langle \Gamma_k \mid \psi \rangle$. (They do form a *tight frame*, a very useful but weaker condition, under some conditions (see p. 1275 in [12]); see MacLennan [34] for additional discussion of the non-orthogonality issue.) On the other hand, it is easy to find the coefficients by minimization of the approximation error [13]. Let $\hat{\psi}(\mathbf{c}) = \sum_{k=1}^N c_k \Gamma_k$ and define the error $\mathcal{E} = \|\hat{\psi}(\mathbf{c}) - \psi\|^2$. This is a standard least-squares problem (cf. Sect. "Universal Approximation"), which can be solved by matrix calculation or by gradient descent on the error surface. It is easy to show that $\partial\mathcal{E}/\partial c_k = 2\langle \Gamma_k \mid \hat{\psi}(\mathbf{c}) - \psi \rangle$, and therefore gradient descent is given by $\dot{c}_k = r\langle \Gamma_k \mid \psi - \hat{\psi}(\mathbf{c}) \rangle$ for some rate $r > 0$.

There is considerable evidence (reviewed elsewhere [34]) that approximate Gabor representations are used in primary visual cortex, and there is also evidence that Gabor representations are used for generating motor signals (see pp. 139–144 in Pribram [53], Pribram et al. [54]).

Information Fields

J.J. Hopfield [25] observed that in some cases a neural impulse train can be understood as transmitting two signals: (1) the information content, encoded in the *phase* of the impulses relative to some global or local "clock", and (2) some other pragmatic characteristic of the information (such as importance, urgency, or confidence), encoded in the *rate* of the impulses. Such a combination of phase-encoded semantics and rate-encoded pragmatics may be common in the nervous system. Already in his *Laws of Thought* (1854), George Boole recognized idempotency as characteristic of information: repeating a message does not change its meaning, but it may affect its pragmatic import. The distinction is implicit in our typographic conventions; consider:

<div align="center">

YES   NO

YES   **NO**

</div>

The horizontal distinction is semantic, but the vertical is pragmatic. More generally, following a distinction that has been made in quantum mechanics (see pp. 35–36 in [7]), we may say that the *form* of the signal *guides* the resulting action, but its *magnitude* determines the *amount* of action.

Similarly in field computation it may be useful to represent information by a field's shape and pragmatics by its magnitude; that is, pragmatics depends on the total amount of "stuff", semantics on it disposition (also a holistic property). The magnitude of such an *information field*

is given by its norm $\|\psi\|$, where we normally mean the inner-product norm of the Hilbert space, $\|\psi\|^2 = \langle \psi \mid \psi \rangle$ (which we can think of as "energy"), but other norms may be appropriate, depending on the relevant sense of the "amount" of action. The semantics of such fields is determined by their form, which we may identify with the normalization of the field, $N(\psi) = \psi / \|\psi\|$ (for nonzero fields). Idempotency is expressed by the identity $N(z\psi) = N(\psi)$ for all $z \neq 0$.

Therefore, it is reasonable that the entropy of a field depends on its shape, but not its magnitude:

$$
S(\psi) = \int_\Omega \frac{\psi_u}{\|\psi\|} \log \frac{\psi_u}{\|\psi\|} du
$$
$$
= \int_\Omega N(\psi)_u \log N(\psi)_u du = \langle N(\psi) \mid \overline{\log} N(\psi) \rangle .
$$

It is perhaps unsurprising that similar issues arise in quantum mechanics and field computation, for they are both formulated in the language of Hilbert spaces. For example, a quantum mechanical state ψ is taken to be undetermined with respect to magnitude, so that $z\psi$ is the same state as ψ for any nonzero complex number z (see p. 17 in [16]). Therefore, the state is conventionally taken to the normalized, $\|\psi\| = 1$, so that its square is a probability density function, $\rho_x = |\psi_x|^2$.

Independence of magnitude is also characteristic of the quantum potential, which led Bohm and Hiley [7] to characterize this field as *active information*. For example, if we write the wave function in polar form, $\psi_x = R_x e^{iS_x/h}$, then the motion of a single particle is given by (see pp. 28–29 in [7]):

$$
\frac{\partial S_x}{\partial t} + \frac{(\nabla S_x)^2}{2m} + V_x + Q_x = 0 ,
$$

where the quantum potential is defined:

$$
Q_x = -\frac{\hbar^2}{2m} \frac{\nabla^2 R_x}{R_x} .
$$

Since the Laplacian $\nabla^2 R_x$ is scaled by R_x, the quantum potential depends only on the *local form* of the wavefunction ψ, not on its magnitude. From this perspective, the particle moves under its own energy, but the quantum potential controls the energy.

Field Representations of Discrete Symbols

Quantum field theory treats discrete particles as quantized excitations of a field. This observation suggests analogous means by which field computation can represent and manipulate discrete symbols and structures, such as those employed in symbolic AI. It also provides potential models for neural representation of words and categories, especially in computational maps, which may illuminate how discrete symbol processing interacts with continuous image processing. From this perspective, discrete symbol manipulation is an emergent property of continuous field computation, which may help to explain the flexibility of human symbolic processes, such as language use and reasoning [36,37,38].

Mathematically, discrete symbols have the *discrete topology*, which is defined by the *discrete metric*, for which the distance between any two distinct objects is 1: $d(x, x) = 0$ and $d(x, y) = 1$ for $x \neq y$. Therefore we will consider various field representations of symbols that have this property. For example, discrete symbols could be represented by localized, non-overlapping patterns of activity in a computational map. In particular, symbols could be represented by Dirac delta functions, for which $\langle \delta_x \mid \delta_x \rangle = 1$ and $\langle \delta_x \mid \delta_y \rangle = 0$ for $x \neq y$. Here we may let $d(x, y) = 1 - \langle \delta_x \mid \delta_y \rangle$. More realistically, symbols could be represented by physically realizable normalized fields ϕ_x with little or no overlap between the representations of different symbols: $\langle \phi_x \mid \phi_y \rangle \approx 0$ for $x \neq y$. Indeed, any sufficiently large set of orthonormal fields may be used to represent discrete symbols. Fields may seem like an inefficient way to represent discrete symbols, and so it is worth observing that with at least 146 000 neurons per square millimeter, a one hundred thousand-word vocabulary could be represented in a few square millimeters of cortex.

Since the meaning of these fields is conveyed by the location of activity peak in the map, that is, by the shape of the field rather than its amplitude, the field's amplitude can be used for pragmatic scale factors, as previously discussed (see Sect. "Nonlinear Computation via Topographic Maps"). This could be used, for example, to convey the confidence or probability of a word or verbal category, or another pragmatic factor, such as loudness (cf. Sect. "Information Fields").

Wave packets (coherent states, Gabor elementary functions) are localized patterns of oscillation resulting from the superposition of a number of nonlocal oscillators with a Gaussian distribution of frequencies [34]. The relative phase of these oscillators determines the position of the wave packet within its field of activity. Therefore different phase relationships may determine field representations for different discrete symbols. The amplitude of the wave packet could represent pragmatic information, and frequency could be used for other purposes, for example for *symbol binding*, with bound symbols having the same frequency. Continuous phase control could be used to control the motion of wave packets in other representa-

tions, such as direction fields (Sect. "Motion in Direction Fields").

Gradient Processes

Many optimization algorithms and adaptive processes are implemented by gradient ascent or gradient descent. Because of its physical analogies, it is more convenient to think of optimization as decreasing a *cost function* rather than increasing some *figure of merit*. For example, the function might represent the difficulty of a motor plan or the incoherence in an interpretation of sensory data (such as stereo disparity).

Therefore suppose that $U \colon \Phi(\Omega) \to \mathbb{R}$ is a functional that defines the undesirability of a field; the goal is to vary ϕ so that $U(\phi)$ decreases down a path of "steepest descent". (By analogy with physical systems, we may call U a *potential function* and think of gradient descent as a *relaxation process* that decreases the potential.) The change in the potential U is given by the chain rule for field transformations (Eq. 3):

$$
\begin{aligned}
\dot{U}(t) &= (U \circ \phi)'(t, 1) \\
&= U'[\phi(t)][\phi'(t)(1)] \\
&= \langle \nabla U[\phi(t)] \mid \dot{\phi}(t) \rangle .
\end{aligned}
$$

More briefly, suppressing the dependence on time, $\dot{U} = \langle \nabla U(\phi) \mid \dot{\phi} \rangle$. To guarantee $\dot{U} \leq 0$ we let $\dot{\phi} = -r\nabla U(\phi)$ with a rate $r > 0$ for gradient descent. Then,

$$
\begin{aligned}
\dot{U} &= \langle \nabla U(\phi) \mid \dot{\phi} \rangle \\
&= \langle \nabla U(\phi) \mid -r\nabla U(\phi) \rangle \\
&= -r\|\nabla U(\phi)\|^2 \leq 0 .
\end{aligned}
$$

Therefore, gradient descent decreases U so long as the gradient is nonzero. (More generally, of course, so long as the trajectory satisfies $\langle \nabla U(\phi) \mid \dot{\phi} \rangle < 0$ the potential will decrease.)

Often the potential takes the form of a *quadratic functional*:

$$
U(\phi) = \phi K \phi + L\phi + c ,
$$

where $K \in \Phi(\Omega \times \Omega)$, $\phi K \phi = \int_{\Omega} \int_{\Omega} \phi_u K_{uv} \phi_v \mathrm{d}u \mathrm{d}v$, L is a linear functional, and $c \in \mathbb{R}$. We require the *coupling field K* to be symmetric: $K_{uv} = K_{vu}$ for all $u, v \in \Omega$; typically it reflects the importance of correlated activity between any two locations u and v in ϕ. By the Riesz Representation Theorem (Sect. "Field Transformations") this quadratic functional may be written

$$
U(\phi) = \phi K \phi + \langle \rho \mid \phi \rangle + c ,
$$

where $\rho \in \Phi(\Omega)$. The field gradient of such a functional is especially simple:

$$
\nabla U(\phi) = 2K\phi + \rho .
$$

In many cases $\rho = \mathbf{0}$ and then gradient descent is a linear process: $\dot{\phi} = -rK\phi$.

This process can be understood as follows. Notice that $-K_{uv}$ decreases with the *coupling* between locations u and v in a field and reflects the inverse variation of the potential with the coherence of the activity at those sites (i. e., the potential measures lack of coherence). That is, if $K_{uv} > 0$ then the potential will be lower to the extent that activity at u *covaries* with activity at v (since then $-\phi_u K_{uv} \phi_v \leq 0$), and if $K_{uv} < 0$, the potential will be lower to the extent they *contravary*. Therefore, the gradient descent process $\dot{\phi} = -rK\phi$ changes ϕ_u to maximally decrease the potential in accord with the covariances and contravariances with other areas as defined by K: $\dot{\phi}_u = -r \int_{\Omega} K_{uv} \phi_v \mathrm{d}v$. The gradient descent will stop when it produces a field ϕ^* for which $-rK\phi^* = \mathbf{0}$, that is, a field in the *null space* of K (the set of all $\phi \in \Phi(\Omega)$ such that $K\phi = \mathbf{0}$).

Universal Approximation

A system of *universal computation* provides a limited range of facilities that can be programmed or otherwise set up to implement any computation in a large and interesting class. The most familiar example is the Universal Turing Machine (UTM), which can be programmed to emulate any Turing machine, and therefore can implement any (Church–Turing) computable function. While this model of universal computation has been important in the theory of digital computation, other models may be more relevant in for other computing paradigms [40,41] (see also ▶ Analog Computation).

Models of universal computation are important for both theory and practice. First, they allow the theoretical power of a computing paradigm to be established. For example, what cannot be computed by a UTM cannot be computed by a Turing machine or by any computer equivalent to a Turing machine. Conversely, if a function in Church–Turing computable, then it can be computed on a UTM or any equivalent machine (such as a programmable, general-purpose digital computer). Second, a model of universal computation for a computing paradigm provides a starting point for designing a general-purpose computer for that paradigm. Of course, there are many engineering problems that must be solved to design a practical general-purpose computer, but a model of universal computation establishes a theoretical foundation.

In the context of field computing there are several approaches to universal computation. One approach to universal field computation is based on a kind of field polynomial approximation based on the Taylor series for field transformations (see Sect. "Derivatives of Field Transformations") [32,33]. Another approach relies on a variety of "universal approximation theorems" for real functions, which are themselves generalizations of Fourier-series approximation (see pp. 208–209, 249–250, 264-265, 274–278, 290–294 in [23]). To explain this approach we will begin with the problem of interpolating a field transformation $F: \Phi(\Omega) \to \Phi(\Omega')$ specified by the samples $F(\phi_k) = \psi^k$, $k = 1, \dots, P$. Further, we require the interpolating function to have the form

$$\hat{\psi} = \sum_{j=1}^{H} r_j(\phi)\alpha_j ,$$

for some H, where the $r_j: \Phi(\Omega) \to \mathbb{R}$ are fixed nonlinear functionals (real-valued field transformation), and the $\alpha_j \in \Phi(\Omega')$ are determined by the samples so as to minimize the sum-of-squares error defined by $\mathcal{E} = \sum_{k=1}^{P} \|\hat{\psi}^k - \psi^k\|^2$, where $\hat{\psi}^k = \sum_{j=1}^{H} r_j(\phi_k)\alpha_j$. (A regularization term can be added if desired (see Chap. 5 in [23]).)

A field, as an element of a Hilbert space, has the same norm as the (infinite) sequence of its generalized Fourier coefficients (with respect to some ON basis). Let $\zeta_1, \zeta_2, \dots$ be a basis for $\Phi(\Omega')$, and we can compute the Fourier coefficients of $\hat{\psi}^k - \psi^k$ as follows:

$$\langle \zeta_i \mid \hat{\psi}^k - \psi^k \rangle = \left\langle \zeta_i \left| \sum_{j=1}^{H} r_j(\phi_k)\alpha_j - \psi^k \right. \right\rangle$$

$$= \left[\sum_{j=1}^{H} r_j(\phi_k)\langle \zeta_i \mid \alpha_j \rangle \right] - \langle \zeta_i \mid \psi^k \rangle .$$

Let $R_{kj} = r_j(\phi_k)$, $A_{ji} = \langle \zeta_i \mid \alpha_j \rangle$, and $Y_{ki} = \langle \zeta_i \mid \psi^k \rangle$. Then, $\langle \zeta_i \mid \hat{\psi}^k - \psi^k \rangle = \sum_{j=1}^{H} R_{kj}A_{ji} - Y_{ki}$. The fields may approximated arbitrarily closely by the first N Fourier coefficients, in which case R, A, and Y are ordinary matrices. Then $\|\hat{\psi}^k - \psi^k\|^2 \approx \sum_{i=1}^{N} E_{ki}^2$, where E = RA − Y. Therefore the approximate total error is $\hat{\mathcal{E}} = \sum_{k=1}^{P} \sum_{i=1}^{N} E_{ki}^2$, or $\hat{\mathcal{E}} = \|E\|_F^2$ (the squared Frobenius norm).

This is a standard least-squares minimization problem, and, as is well known (see pp. 371–373 in [28]), the error is minimized by A = R$^+$Y, where R$^+$ is the *Moore–Penrose pseudoinverse* of the interpolation matrix R: R$^+$ = (R^{T}R)$^{-1}$R^T. From A we can compute the required fields to approximate F: $\alpha_j = \sum_{i=1}^{N} A_{ji}\zeta_i$.

For universality, we require that the approximation error can be made arbitrarily small, which depends on the choice of the basis functionals r_j, as can be learned from multivariable interpolation theory. Therefore, we represent the input fields by their first M generalized Fourier coefficients, an approximation that can be made as accurate as we like. Let $\beta_1, \beta_2, \dots$ be an ON basis for $\Phi(\Omega)$ and let $\mathbf{p}^M: \Phi(\Omega) \to \mathbb{R}^M$ compute this finite-dimensional representation: $p_j^M(\phi) = \langle \beta_j \mid \phi \rangle$. We will approximate $r_j(\phi) \approx s_j[\mathbf{p}^M(\phi)]$, for appropriate functions $s_j: \mathbb{R}^M \to \mathbb{R}$, $j = 1, \dots, H$. That is, we are approximating the field transformation F by

$$F(\phi) \approx \sum_{j=1}^{H} s_j[\mathbf{p}^M(\phi)]\alpha_j .$$

Now let $S_{kj} = s_j[\mathbf{p}^M(\phi_k)]$, and we have corresponding finite-dimensional interpolation conditions Y = SA with the best least-square solution A = S$^+$Y.

Various universal approximation theorems tell us that, given an appropriate choice of basis functions $s_1, \dots, s_H$, any continuous function $\mathbf{f}: \mathbb{R}^M \to \mathbb{R}^N$ can be approximated arbitrarily closely by a linear combination of these functions (see pp. 208–209 in [23]). That is, the error $\hat{\mathcal{E}} = \|SA - Y\|_F^2$ can be made as small as we like. Therefore, appropriate choices for the s_j imply corresponding choices for the basis functionals r_j.

For example, one universal class of basis functions has the form $s_j(\mathbf{x}) = c(\mathbf{w}_j \cdot \mathbf{x} + b_j)$, for any nonconstant, bounded, monotone-increasing continuous function c (see p. 208 in [23]). This form is common in artificial neural networks, where $\mathbf{w}_j$ is a vector of neuron j's input weights (connection strengths) and b_j is its bias. To find the corresponding basis functional, $r_j(\phi) = s_j[\mathbf{p}^M(\phi)]$, observe

$$\mathbf{w}_j \cdot \mathbf{p}^M(\phi) + b_j = \sum_{k=1}^{M} w_{jk}p_k^M(\phi) + b_j$$

$$= \sum_{k=1}^{M} w_{jk}\langle \beta_k \mid \phi \rangle + b_j$$

$$= \left\langle \sum_{k=1}^{M} w_{jk}\beta_k \left| \phi \right. \right\rangle + b_j .$$

Therefore, let $\varpi_j = \sum_{k=1}^{M} w_{jk}\beta_k$, and we see that a universal class of functionals has the form:

$$r_j(\phi) = c(\langle \varpi_j \mid \phi \rangle + b_j) . \tag{10}$$

Thus, in this field analog of an artificial neuron, the input field ϕ is matched to the neuron's interconnection field ϖ_j.

Another universal class is the *radial basis functions*, $s_j(\mathbf{x}) = r(\|\mathbf{x} - \mathbf{c}^j\|)$, where the radial function r is monotonically decreasing, and the centers $\mathbf{c}^j$ are either fixed or dependent on the function to be approximated. A corresponding universal class of field functions has the form:

$$r_j(\phi) = r(\|\phi - \eta_j\|), \tag{11}$$

where each field $\eta_j = \sum_i c_i^j \zeta_i$ causes the maximal response of the corresponding basis function r_j. Furthermore, if we set $H = P$ and $\eta_j = \phi_j$, then the matrix R is invertible for a wide variety of radial functions r (see pp. 264–265 in [23]).

Thus familiar methods of universal approximation can be transferred to field computation, which reveals simple classes of field transformations that are universal. This implies that universal field computers can be designed around a small number of simple functions (e. g., field summation, inner product, monotonic real functions).

Field Computers

Structure

As previously explained (see Sect. "Definition"), fields do not have to be physically continuous in either variation or spatial extension (that is, in range or domain), so long as the discretization is sufficiently fine that a continuum is a practical approximation. Therefore, field computation can be implemented with ordinary serial or parallel digital computing systems (as it has been in the past). However, field computation has a distinctively different approach to information representation and processing; computation tends to be shallow (in terms of operations applied), but very wide, "massively parallel" in the literal sense of computing with an effectively continuous *mass* of processors. Therefore field computation provides opportunities for the exploitation of novel computing media that may not be suitable for digital computation. For example, as the brain illustrates how relatively slow, low precision analog computing devices can be used to implement intelligent information processing via field computation, so electronic field computers may exploit massive assemblages of low-precision analog devices, which may be imprecisely fabricated, located, and interconnected. Other possibilities are optical computing in which fields are represented by optical wavefronts, molecular computation based on films of bacteriorhodopsin or similar materials, chemical computers based on reaction-diffusion systems, and "free space computing" based on the interactions of charge carriers and electrical fields in homogeneous semiconductors (see Sect. "Field Computing Hardware").

Field computation is a kind of analog computation, and so there are two principal time domains in which field computation can take place, sequential time and continuous time (see ▶ Analog Computation). In *sequential* computation, operations take place in discrete steps in an order prescribed by a program. Therefore, sequential field computation is similar to ordinary digital computation, except that the individual program steps may perform massively parallel analog field operations. For example, a field assignment statement, such as:

$$\psi := \phi + \psi ;$$

updates that field variable ψ to contain the sum of ϕ and the previous value of ψ.

In *continuous-time computation* the fields vary continuously in time, generally according to differential equations in which time is the independent variable; this has been the mode of operation of most analog computers in the past. In this case, a simple dependence, such as $\psi = \phi + \chi$, is assumed to have an implicit time parameter, $\psi(t) = \phi(t) + \chi(t)$, which represents the real time of computation. Since continuous-time programs are often expressed by differential equations, these computers usually provide hardware for definite integration of functions with respect to time:

$$\psi(t) = \psi_0 + \int_0^t F[\phi(\tau)]\mathrm{d}\tau . \tag{12}$$

Continuous-time programs are expressed by circuit diagrams (variable-dependency diagrams) rather than by textual programs such as used in digital computer programming (see Fig. 1 for an example). Although special-purpose analog and digital computers are appropriate for many purposes, already in the first half of the twentieth century the value of general-purpose (programmable) digital and analog computers had been recognized (see ▶ Analog Computation). Therefore it will be worthwhile to consider briefly the sort of facilities we may expect to find in a general-purpose field computer (whether operating in sequential or continuous time).

We have seen that the following facilities are sufficient for universal computation (Sect. "Universal Approximation"): multiplication of fields by scalars, local (point-wise) addition of fields ($\psi_u = \phi_u + \chi_u$), and some means of computing appropriate basis functionals. Neural-net style functionals (Eq. 10) require inner product and any nonconstant, bounded, monotone-increasing scalar function (i. e., a sigmoid function). Radial basis functionals (Eq. 11) require the norm (which can be computed with the in-

ner product) and any non-constant, bounded, monotone-decreasing scalar function. (Point-wise subtraction can be implemented, of course, by scalar multiplication and point-wise addition.) These are modest requirements, and we can expect practical field computers to have additional facilities.

In addition, continuous-time field computers will implement definite integration with respect to time (Eq. 12), which is used to implement field processes defined by differential equations. The equations are implemented in terms of the operations required for universal computation or in terms of others, discussed next.

Additional useful operations for general-purpose field computing include matrix-vector style field products (Hilbert–Schmidt integral operators), outer product, convolution, cross-correlation, normalization, local (point-wise) product and quotient ($\psi_u = \phi_u \chi_u$, $\psi_u = \phi_u/\chi_u$), and various other local operations ($\overline{\log}$, $\overline{\exp}$, etc.). Operations on vector fields can be implemented by scalar field operations on the vector components (Cartesian or polar); in this manner, vector fields of any finite dimension can be processed. If vector fields and operations on them are provided by the hardware, then it is useful if these operations include conversions between scalar and vector fields (e. g., between vector fields and their Cartesian or polar coordinate fields). Other useful vector field operations include point-wise scalar products between vector fields ($\boldsymbol{\psi}_u = \boldsymbol{\phi}_u \cdot \boldsymbol{\chi}_u$), gradient ($\nabla$), Laplacian ($\nabla^2$), divergence ($\nabla\cdot$), and point-wise scalar-vector multiplication ($\boldsymbol{\psi}_u = \phi_u \boldsymbol{\chi}_u$). Scalar analog computation is a degenerate case of field computation (since scalars correspond to fields in $\Phi(\{0\})$), and so practical general-purpose field computers will include the facilities typical of analog computers (see ▶ Analog Computation).

The Extended Analog Computer

One interesting proposal for a general-purpose field computer is the *Extended Analog Computer* (EAC) of LA Rubel, which was a consequence of his conviction that the brain is an analog computer [55]. However, Rubel and others had shown that the existing model of a general-purpose analog computer (GPAC), the abstract *differential analyzer* defined by CE Shannon, had relatively severe theoretical limitations, and so it did not seem adequate as a model of the brain (see ▶ Analog Computation) [30,51,56,60,61]. Like Shannon's differential analyzer, the EAC is an abstract machine intended for theoretical investigation of the power of analog computation, not a proposal for a practical computer [57]; nevertheless, some actual computing devices have been based on it.

The EAC is structured in a series of levels, each building on those below it, taking outputs from the lower layers and applying analog operations to them to produce its own outputs. The inputs to the lowest layer are a finite number of "settings", which can be thought of as real-numbers (e. g., set by a continuously adjustable knob). This layer is able to combine the inputs with real constants to compute polynomials over which it can integrate to generate differentially algebraic functions; this layer is effectively equivalent to Shannon's GPAC. Each layer provides a number of analog devices, including "boundary-value-problem boxes", which can solve systems of PDEs subject to boundary conditions and other constraints. That is, these conceptual devices solve field computation problems. Although for his purposes Rubel was not interested in implementation, he did remark that PDE solvers might be implemented by physical processes that obeyed the same class of PDEs as the problem (e. g., using physical diffusion to solve diffusion problems). This of course is precisely the old field analogy method, which was also used in network analyzers (recall Sect. "Introduction"). Rubel was able to show that the EAC is able to solve an extremely large class of problems, but the extent of its power has not been determined (see ▶ Analog Computation).

Field Computing Hardware

Research in field computing hardware is ongoing and a comprehensive survey is beyond the scope of this article; a few examples must suffice.

Although the EAC was intended as a conceptual machine (for investigating the limits of analog computing), JW Mills has demonstrated several hardware devices inspired by it [46,47]. In these the diffusion of electrons in bulk silicon or conductive gels is used to solve diffusion equations subject to given boundary conditions, an technique he describes as "computing with empty space". This approach, in which a physical system satisfying certain PDEs is used to solve problems involving similar PDEs, is a contemporary version of the "field analogy method" developed by Kirchhoff and others (see Sect. "Introduction").

Adamatzky and his colleagues have investigated chemical field computers for implementing reaction-diffusion equations [1,2]; see Sect. "Diffusion Processes" and ▶ Reaction-Diffusion Computing. These use variants of the Belousov–Zhabotinsky Reaction and similar chemical reactions. Although the chemical reactions proceed relatively slowly, they are massively parallel: at the molecular level ("molar parallelism"). Also, Chaps. 6–8 in Adamatzky et al. [2] have designed both analog and digital electronic reaction-diffusion computers. M Perŭs and his col-

leagues have investigated the use of quantum holography to implement field analogues of neural-network algorithms [31,49].

Several investigators have explored optical implementations of field computers. For example, Skinner et al. [62] used self-lensing media, which respond nonlinearly to applied irradience, to implement feed-forward neural networks trained by back-propagation. Tõkés et al. [68,69] have been developing an optical field computer using bacteriorhodopsin as a medium.

Future Directions

In the future field computation can be expected to provide an increasingly important analytical and intuitive framework for understanding massively parallel analog computation in natural and artificial intelligence.

First, field computation will provide a theoretical framework for understanding information processing in the brain in terms of cortical maps and, more generally, at a level between anatomical structures and individual neurons or small neural circuits. This will require improved understanding of information processing in terms of field computation, which will benefit from cognitive neuroscience research, but also contribute new computational concepts to it. Increased understanding of neural field computation will improve our ability to design very large artificial neural networks, which will be more attractive as massively parallel neurocomputing hardware is developed.

Traditionally, artificial intelligence has approached knowledge representation from the perspective of discrete, language-like structures, which are difficult to reconcile with the massively parallel analog representations found in the cortex. Therefore field computation will provide an alternative framework for understanding knowledge representation and inference, which will be more compatible with neuroscience but also provide a basis for understanding cognitive phenomena such as context sensitivity, perception, sensorimotor coordination, image-based cognition, analogical and metaphorical thinking, and nonverbal intelligences (kinesthetic, emotional, aesthetic, etc.).

As we have seen, concepts from field computation may be applied to understanding the collective intelligence of large groups of organisms. This approach permits separating the abstract computational principles from the specifics of their realization by particular organisms, and therefore permits their application to other organisms or artificial systems. For example, principles of field computation governing the self-organization of groups of organisms are applicable to distributed robotics; in particular,

they will provide a foundation for controlling very large population of microrobots or nanobots.

Embryological morphogenesis is naturally expressed in terms of field computation, since the differentiation and self-organization of an (initially homogeneous) cell mass is governed by continuous distributions of continuous quantity. Therefore, field computation provides a vehicle for rising above the specifics of particular signaling molecules, mechanisms of cell migration, etc. in order to understand development in abstract or formal terms. Understanding morphogenesis in terms of field computation will facilitate applying its principles to other systems in which matter self-organizes into complex structures. In particular, field computation will suggest means for programming the reorganization of matter for nanotechnological applications and for describing the behavior of adaptive "smart" materials.

As we approach the end of Moore's Law [48], future improvements in computing performance will depend on developing new computing paradigms not based in sequential digital computation (see also ▶ Analog Computation). Improvements in both speed and density can be achieved by matching data representations and computational operations to the physical processes that realize them, which are primarily continuous and parallel in operation. Indeed, many of these processes are described in terms of fields or involve physical fields (i. e., phenomenological or structural fields). Therefore field computation points toward many structural processes that might be used for computation and provides a framework for understanding how best to use them. Thus we anticipate that field computation will play an important role in post-Moore's Law computing.

Bibliography

Primary Literature

1. Adamatzky A (2001) Computing in nonlinear media and automata collectives. Institute of Physics Publishing, Bristol
2. Adamatzky A, De Lacy Costello B, Asai T (2005) Reaction-diffusion computers. Elsevier, Amsterdam
3. Anderson JA (1995) An introduction to neural networks. MIT Press, Cambridge
4. Anderson RA (1995) Coordinate transformations and motor planning in posterior parietal cortex. In: Gazzaniga MS (ed) The cognitive neurosciences. MIT Press, Cambridge, pp 519–32
5. Bar-Yam Y (1997) Dynamics of complex systems. Perseus Books, Reading
6. Bizzi E, Mussa-Ivaldi FA (1995) Toward a neurobiology of coordinate transformation. In: Gazzaniga MS (ed) The cognitive neurosciences. MIT Press, Cambridge, pp 495–506
7. Bohm D, Hiley BJ (1993) The undivided universe: an ontological interpretation of quantum theory. Routledge, New York

8. Bonabeau E, Dorigo M, Theraulaz G (1999) Swarm intelligence: from natural to artificial systems. Santa Fe Institute Studies in the Sciences of Complexity. Oxford University Press, New York

9. Brachman G, Narici L (1966) Functional analysis. Academic Press, New York

10. Camazine S, Deneubourg J-L, Franks NR, Sneyd G, Theraulaz J, Bonabeau E (2001) Self-organization in biological systems. Princeton University Press, Princeton

11. Changeux J-P (1985) Neuronal man: the biology of mind. Oxford University Press, Oxford, tr. by L. Garey

12. Daubechies I, Grossman A, Meyer Y (1986) Painless non-orthogonal expansions. J Math Phys 27:1271–1283

13. Daugman JG (1993) An information-theoretic view of analog representation in striate cortex. In: Schwartz EL (ed) Computational neuroscience. MIT Press, Cambridge, pp 403–423

14. Davies JA (2005) Mechanisms of morphogensis. Elsevier, Amsterdam

15. Deneubourg JL (1977) Application de l'ordre par fluctuation à la description de certaines étapes de la construction du nid chez les termites. Insectes Sociaux 24:117–130

16. Dirac PAM (1958) The principles of quantum mechanics, 4th edn. Oxford University Press, Oxford

17. Droulez J, Berthoz A (1991) The concept of dynamic memory in sensorimotor control. In: Humphrey DR, Freund H-J (eds) Motor control: concepts and issues. Wiley, New York, pp 137–161

18. Droulez J, Berthoz A (1991) A neural network model of sensoritopic maps with predictive short-term memory properties. Proc Natl Acad Sci USA 88:9653–9657

19. Feldman JA, Ballard DH (1982) Connectionist models and their properties. Cogn Sci 6(3):205–254

20. Gabor D (1946) Theory of communication. J Inst Electr Eng 93(III):429–457

21. Georgopoulos AP (1995) Motor cortex and cognitive processing. In: The Cognitive Neurosciences. MIT Press, Cambridge, pp 507–517

22. Goodman SJ, Anderson RA (1989) Microstimulation of a neural-network model for visually guided saccades. J Cogn Neurosci 1:317–326

23. Haykin S (1999) Neural networks: a comprehensive foundation, 2nd edn. Prentice-Hall, Upper Saddle River

24. Heil CE, Walnut DF (1989) Continuous and discrete wavelet transforms. SIAM Rev 31(4):628–666

25. Hopfield JJ (1995) Pattern recognition computation using action potential timing for stimulus response. Nature 376:33–36

26. Kirchhoff G (1845) Ueber den Durchgang eines elektrischen Stromes durch eine Ebene, insbesondere durch eine kreisförmige. Ann Phys Chemie 140/64(4):497–514

27. Knudsen EJ, du Lac S, Esterly SD (1987) Computational maps in the brain. Ann Rev Neurosc 10:41–65

28. Leon SJ (1986) Linear algebra with applications, 2nd edn. Macmillan, New York

29. Light WA (1992) Ridge functions, sigmoidal functions and neural networks. In: Cheney EW, Chui CK, Schumaker LL (eds) Approximation theory VII. Academic Press, Boston, pp 163–206

30. Lipshitz L, Rubel LA (1987) A differentially algebraic replacment theorem. Proc Am Math Soc 99(2):367–72

31. Loo CK, Peruš M, Bischof H (2004) Associative memory based image and object recognition by quantum holography. Open Syst Inf Dyn 11(3):277–289

32. MacLennan BJ (1987) Technology-independent design of neurocomputers: the universal field computer. In: Caudill M, Butler C (eds) Proceedings of the IEEE First International Conference on Neural Networks, vol 3. IEEE Press, Piscataway, pp 39–49

33. MacLennan BJ (1990) Field computation: a theoretical framework for massively parallel analog computation, parts I–IV. Technical Report CS-90-100. Department of Computer Science, University of Tennessee, Knoxville, Available from www.cs.utk.edu/~mclennan

34. MacLennan BJ (1991) Gabor representations of spatiotemporal visual images. Technical Report CS-91-144. Department of Computer Science, University of Tennessee, Knoxville, Available from www.cs.utk.edu/~mclennan

35. MacLennan BJ (1993) Information processing in the dendritic net. In: Karl HP (ed) Rethinking neural networks: quantum fields and biological data. Lawrence Erlbaum, Hillsdale, pp 161–197

36. MacLennan BJ (1994) Continuous computation and the emergence of the discrete. In: Karl HP (ed) Origins: brain and self-organization. Lawrence Erlbaum, Hillsdale, pp 121–151

37. MacLennan BJ (1994) Image and symbol: continuous computation and the emergence of the discrete. In: Honavar V, Uhr L (eds) Artificial intelligence and neural networks: steps toward principled integration. Academic Press, New York, pp 207–224

38. MacLennan BJ (1995) Continuous formal systems: a unifying model in language and cognition. In: Proc. of the IEEE Workshop on Architectures for Semiotic Modeling and Situation Analysis in Large Complex Systems. IEEE Press, Piscataway, pp 161–172

39. MacLennan BJ (1997) Field computation in motor control. In: Morasso PG, Sanguineti V (eds) Self-organization, computational maps and motor control. Elsevier, Amsterdam, pp 37–73

40. MacLennan BJ (2003) Transcending Turing computability. Minds Mach 13:3–22

41. MacLennan BJ (2004) Natural computation and non-Turing models of computation. Theor Comput Sci 317:115–145

42. Mathematical Society of Japan (1980) Encyclopedic dictionary of mathematics. MIT Press, Cambridge

43. McClelland JL, Rumelhart DE, PDP Research Group (1986) Parallel distributed processing: explorations in the microstructure of cognition, vol 2. Psychological and biological models. MIT Press, Cambridge

44. McFadden J (2002) Synchronous firing and its influence on the brain's electromagnetic field: evidence for an electromagnetic field theory of consciousness. J Conscious Stud 9(4):23–50

45. Miller MI, Roysam B, Smith KR, O'Sullivan JA (1991) Representing and computing regular languages on massively parallel networks. IEEE Trans Neural Netw 2:56–72

46. Mills JW (1996) The continuous retina: Image processing with a single-sensor artificial neural field network. In: Proc. IEEE Conference on Neural Networks. IEEE Press, Piscataway

47. Mills JW, Himebaugh B, Kopecky B, Parker M, Shue C, Weilemann C (2006) "Empty space" computes: the evolution of an unconventional supercomputer. In: Proc. of the 3rd Conference on Computing Frontiers. ACM Press, New York, pp 115–126

48. Moore GE (1965) Cramming more components onto integrated circuits. Electronics 38(8):114–117

49. Peruš M (1998) A quantum information-processing "algorithm" based on neural nets. In: Wang P, Georgiou G, Janikow C, Yao Y (eds) Joint conference on information sci-

ences, vol II. Association for Intelligent Machinery, New York, pp 197–200

50. Pockett S (2000) The nature of consciousness: a hypothesis. Writers Club Press, San Jose

51. Pour-El MB (1974) Abstract computability and its relation to the general purpose analog computer (some connections between logic, differential equations and analog computers). Trans Am Math Soc 199:1–29

52. Powell MJD (1987) Radial basis functions for multivariable interpolation: a review. In: Algorithms for approximation. Clarendon, New York, pp 143–167

53. Pribram KH (1991) Brain and perception: holonomy and structural in figural processing. Lawrence Erlbaum, Hillsdale

54. Pribram KH, Sharafat A, Beekman GJ (1984) Frequency encoding in motor systems. In: Whiting HTA (ed) Human motor actions: Bernstein reassessed. Elsevier, New York, pp 121–156

55. Rubel LA (1985) The brain as an analog computer. J Theor Neurobiol 4:73–81

56. Rubel LA (1988) Some mathematical limitations of the general-purpose analog computer. Adv Appl Math 9:22–34

57. Rubel LA (1993) The extended analog computer. Adv Appl Math 14:39–50

58. Rumelhart DE, McClelland JL, PDP Research Group (1986) Parallel distributed processing: explorations in the microstructure of cognition, vol 1. Foundations. MIT Press, Cambridge

59. Sanger TD (1996) Probability density estimation for the interpretation of neural population codes. J Neurophysiol 76:2790–2793

60. Shannon CE (1941) Mathematical theory of the differential analyzer. J Math Phys Mass Inst Tech 20:337–354

61. Shannon CE (1993) Mathematical theory of the differential analyzer. In: Sloane NJA, Wyner AD (eds) Claude Elwood Shannon: collected papers. IEEE Press, New York, pp 496–513

62. Skinner SR, Behrman EC, Cruz-Cabrera AA, Steck JE (1995) Neural network implementation using self-lensing media. Appl Opt 34:4129–4135

63. Small JS (2001) The analogue alternative: the electronic analogue computer in Britain and the USA, 1930–1975. Routledge, London, New York

64. Solé R, Goodwin B (2000) Signs of life: how complexity pervades biology. Basic Books, New York

65. Soroka WW (1954) Analog methods in computation and simulation. McGraw-Hill, New York

66. Steinbeck O, Tóth A, Showalter K (1995) Navigating complex labyrinths: optimal paths from chemical waves. Science 267:868–871

67. Ting P-Y, Iltis RA (1994) Diffusion network architectures for implementation of Gibbs samplers with applications to assignment problems. IEEE Trans Neural Netw 5:622–638

68. Tőkés S, Orzó L, Ayoub A (2003) Two-wavelength POAC (programmable opto-electronic analogic computer) using bacteriorhodopsin as dynamic holographic material. In: Proc. of EC-CTD '03 Conference, vol 3. Krakow, pp 97–100

69. Tőkés S, Orzó L, Váró G, Dér A, Ormos P, Roska T (2001) Programmable analogic cellular optical computer using bacteriorhodopsin as analog rewritable image memory. In: Dér A, Keszthelyi L (eds) Bioelectronic applications of photochromic pigments. IOS Press, Amsterdam, pp 54–73

70. Truitt TD, Rogers AE (1960) Basics of analog computers. John F. Rider, New York

71. Turing AM (1952) The chemical basis of morphogenesis. Philos Trans Royal Soc B 237:37–72

Books and Reviews

Bachman G, Narici L (1966) Functional analysis. Academic Press, New York

Berberian SK (1961) Introduction to Hilbert space. Oxford, New York

MacLennan BJ (1991) Field computation: a theoretical framework for massively parallel analog computation, parts I–IV. Technical Report CS-90-100, Dept. of Computer Science, University of Tennessee, Knoxville. Available from http:www.cs.utk.edu/~mclennan

MacLennan BJ (2008) Foundations of Field Computation. In preparation. Available from http:www.cs.utk.edu/~mclennan

Field Theoretic Methods

UWE CLAUS TÄUBER

Department of Physics, Center for Stochastic Processes in Science and Engineering, Virginia Polytechnic Institute and State University, Blacksburg, USA

Article Outline

Glossary

Absorbing state State from which, once reached, an interacting many-particle system cannot depart, not even through the aid of stochastic fluctuations.

Correlation function Quantitative measure of the correlation of random variables; usually set to vanish for statistically independent variables.

Critical dimension Borderline dimension d_c above which mean-field theory yields reliable results, while for $d \leq d_c$ fluctuations crucially affect the system's large scale behavior.

External noise Stochastic forcing of a macroscopic system induced by random external perturbations, such as thermal noise from a coupling to a heat bath.

Field theory A representation of physical processes through continuous variables, typically governed by an exponential probability distribution.

Generating function Laplace transform of the probability distribution; all moments and correlation functions follow through appropriate partial derivatives.

Internal noise Random fluctuations in a stochastic macroscopic system originating from its internal kinetics.

Langevin equation Stochastic differential equation describing time evolution that is subject to fast random forcing.

Master equation Evolution equation for a configurational probability obtained by balancing gain and loss terms through transitions into and away from each state.

Mean-field approximation Approximative analytical approach to an interacting system with many degrees of freedom wherein spatial and temporal fluctuations as well as correlations between the constituents are neglected.

Order parameter A macroscopic density corresponding to an extensive variable that captures the symmetry and thereby characterizes the ordered state of a thermodynamic phase in thermal equilibrium. Nonequilibrium generalizations typically address appropriate stationary values in the long-time limit.

Perturbation expansion
Systematic approximation scheme for an interacting and/or nonlinear system that involves a formal expansion about an exactly solvable simplification by means of a power series with respect to a small coupling.

Definition of the Subject

Traditionally, complex macroscopic systems are often described in terms of ordinary differential equations for the temporal evolution of the relevant (usually collective) variables. Some natural examples are particle or population densities, chemical reactant concentrations, and magnetization or polarization densities; others involve more abstract concepts such as an apt measure of activity, etc. Complex behavior often entails (diffusive) spreading, front propagation, and spontaneous or induced pattern formation. In order to capture these intriguing phenomena, a more detailed level of description is required, namely the inclusion of spatial degrees of freedom, whereupon the above quantities all become local density fields. Stochasticity, i. e., randomly occuring propagation, interactions, or reactions, frequently represents another important feature of complex systems. Such stochastic processes generate *in-ternal noise* that may crucially affect even long-time and large-scale properties. In addition, other system variables, provided they fluctuate on time scales that are fast compared to the characteristic evolution times for the relevant quantities of interest, can be (approximately) accounted for within a Langevin description in the form of *external* additive or multiplicative noise.

A quantitative mathematical analysis of complex spatio-temporal structures and more generally cooperative behavior in stochastic interacting systems with many degrees of freedom typically relies on the study of appropriate *correlation functions*. *Field-theoretic*, i. e., spatially continuous, representations both for random processes defined through a master equation and Langevin-type stochastic differential equations have been developed since the 1970s. They provide a general framework for the computation of correlation functions, utilizing powerful tools that were originally developed in quantum many-body as well as quantum and statistical field theory. These methods allow us to construct systematic approximation schemes, e. g., *perturbative expansions* with respect to some parameter (presumed small) that measures the strength of fluctuations. They also form the basis of more sophisticated renormalization group methods which represent an especially potent device to investigate scale-invariant phenomena.

Introduction

Stochastic Complex Systems

Complex systems consist of many interacting components. As a consequence of either these interactions and/or the kinetics governing the system's temporal evolution, correlations between the constituents emerge that may induce cooperative phenomena such as (quasi-)periodic oscillations, the formation of spatio-temporal patterns, and phase transitions between different macroscopic states. These are characterized in terms of some appropriate collective variables, often termed *order parameters*, which describe the large-scale and long-time system properties. The time evolution of complex systems typically entails random components: either, the kinetics itself follows stochastic rules (certain processes occur with given probabilities per unit time); or, we project our ignorance of various fast microscopic degrees of freedom (or our lack of interest in their detailed dynamics) into their treatment as stochastic noise.

An exact mathematical analysis of nonlinear stochastic systems with many interacting degrees of freedom is usually not feasible. One therefore has to resort to either computer simulations of corresponding stochastic cellu-

lar automata, or approximative treatments. A first step, which is widely used and often provides useful qualitative insights, consists of ignoring spatial and temporal fluctuations, and just studying equations of motion for ensemble-averaged order parameters. In order to arrive at closed equations, additional simplifications tend to be necessary, namely the factorization of correlations into powers of the mean order parameter densities. Such approximations are called *mean-field* theories; familiar examples are rate equations for chemical reaction kinetics or Landau–Ginzburg theory for phase transitions in thermal equilibrium. Yet in some situations mean-field approximations are insufficient to obtain a satisfactory quantitative description (see, e. g., the recent work collected in [1,2]). Let us consider an illuminating example.

Example: Lotka–Volterra Model

In the 1920s, Lotka and Volterra independently formulated a mathematical model to describe emerging periodic oscillations respectively in coupled autocatalytic chemical reactions, and in the Adriatic fish population (see, e. g., [3]). We shall formulate the model in the language of population dynamics, and treat it as a stochastic system with two species A (the 'predators') and B (the 'prey'), subject to the following reactions: predator death $A \rightarrow \emptyset$, with rate μ; prey proliferation $B \rightarrow B + B$, with rate σ; predation interaction $A + B \rightarrow A + A$, with rate λ. Obviously, for $\lambda = 0$ the two populations decouple; while the predators face extinction, the prey population will explode. The average predator and prey population densities $a(t)$ and $b(t)$ are governed by the linear differential equations $\dot{a}(t) = -\mu a(t)$ and $\dot{b}(t) = \sigma b(t)$, whose solutions are exponentials. Interesting competition arises as a consequence of the nonlinear process governed by the

rate λ. In an exact representation of the system's temporal evolution, we would now need to know the probability of finding an $A - B$ pair at time t. Moreover, in a spatial Lotka–Volterra model, defined on a d-dimensional lattice, say, on which the individual particles can move via nearest-neighbor hopping, the predation reaction should occur only if both predators and prey occupy the same or adjacent sites. The evolution equations for the mean densities $a(t)$ and $b(t)$ would then have to be respectively amended by the terms $\pm \lambda \langle a(x,t)b(x,t) \rangle$. Here $a(x,t)$ and $b(x,t)$ represent local concentrations, the brackets denote the ensemble average, and $\langle a(x,t)b(x,t) \rangle$ represents $A - B$ cross correlations.

In the rate equation approximation, it is assumed that the local densities are uncorrelated, whereupon $\langle a(x,t)b(x,t) \rangle$ factorizes to $\langle a(x,t) \rangle \langle b(x,t) \rangle = a(t)b(t)$. This yields the famous deterministic Lotka–Volterra equations

$$\dot{a}(t) = \lambda a(t)b(t) - \mu a(t), \quad \dot{b}(t) = \sigma b(t) - \lambda a(t)b(t). \quad (1)$$

Within this mean-field approximation, the quantity $K(t) = \lambda[a(t) + b(t)] - \sigma \ln a(t) - \mu \ln b(t)$ (essentially the system's Lyapunov function) is a constant of motion, $\dot{K}(t) = 0$. This results in regular nonlinear population oscillations, whose frequency and amplitude are fully determined by the initial conditions, a rather unrealistic feature. Moreover Eqs. (1) are known to be unstable with respect to various model modifications (as discussed in [3]).

In contrast with the rate equation predictions, the original stochastic spatial Lotka–Volterra system displays much richer behavior (a recent overview is presented in [4]): The predator–prey coexistence phase is governed, for sufficiently large values of the predation rate, by an incessant sequence of 'pursuit and evasion' wave fronts
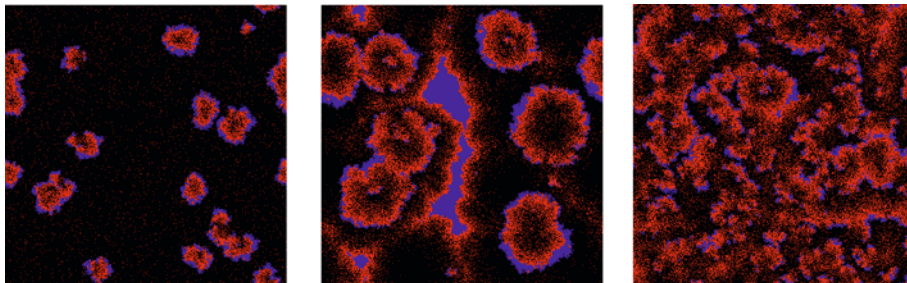
Field Theoretic Methods, Figure 1
Snapshots of the time evolution (*left* to *right*) of activity fronts emerging in a stochastic Lotka–Volterra model simulated on a 512 × 512 lattice, with periodic boundary conditions and site occupation numbers restricted to 0 or 1. For the chosen reaction rates, the system is in the species coexistence phase (with rates $\sigma = 4.0$, $\mu = 0.1$, and $\lambda = 2.2$), and the corresponding mean-field fixed point a focus. The *red*, *blue*, and *black* dots respectively represent predators A, prey B, and empty sites $\emptyset$. Reproduced with permission from [4]
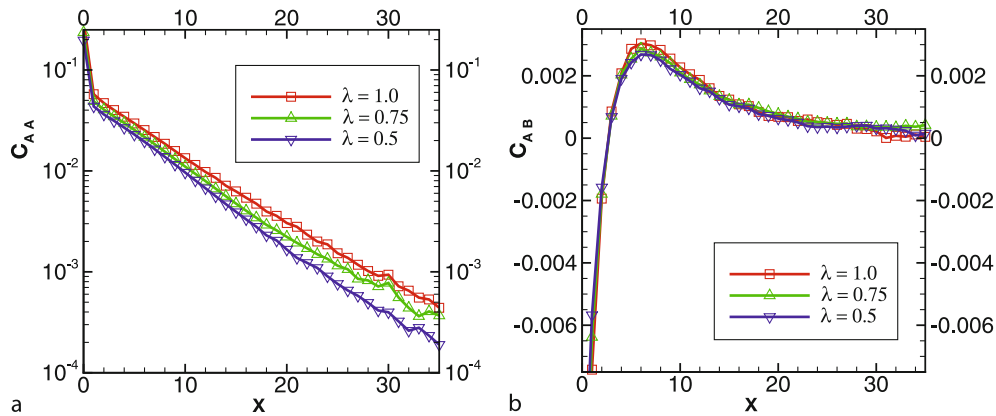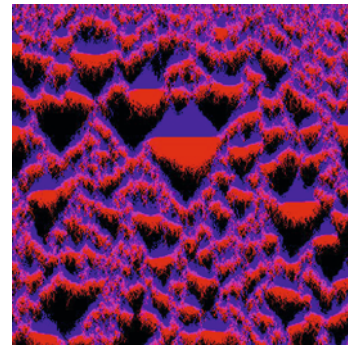
Field Theoretic Methods, Figure 2
Static correlation functions **a** $C_{AA}(x)$ (note the logarithmic scale), and **b** $C_{AB}(x)$, measured in simulations on a 1024 × 1024 lattice without any restrictions on the site occupations. The reaction rates were $\sigma = 0.1$, $\mu = 0.1$, and λ was varied from 0.5 (*blue triangles, upside down*), 0.75 (*green triangles*), to 1.0 (*red squares*). Reproduced with permission from [5]

that form quite complex dynamical patterns, as depicted in Fig. 1, which shows snapshots taken in a two-dimensional lattice Monte Carlo simulation where each site could at most be occupied by a single particle. In finite systems, these correlated structures induce erratic population oscillations whose features are independent of the initial configuration. Moreover, if locally the prey 'carrying capacity' is limited (corresponding to restricting the maximum site occupation number per lattice site), there appears an extinction threshold for the predator population that separates the active coexistence regime through a continuous phase transition from a state wherein at long times $t \to \infty$ only prey survive. With respect to the predator population, this represents an *absorbing state*: Once all A particles have vanished, they cannot be produced by the stochastic kinetics.

A quantitative characterization of the emerging spatial structures utilizes equal-time correlation functions such as $C_{AA}(x - x', t) = \langle a(x, t)a(x', t) \rangle - a(t)^2$ and $C_{AB}(x - x', t) = \langle a(x, t)b(x', t) \rangle - a(t)b(t)$, computed at some large time t in the (quasi-)stationary state. These are shown in Fig. 2 as measured in computer simulations for a stochastic Lotka–Volterra model (but here no restrictions on the site occupation numbers of the A or B particles were implemented). The $A - A$ (and $B - B$) correlations obviously decay essentially exponentially with distance x, $C_{AA}(x) \propto C_{BB}(x) \propto e^{-|x|/\xi}$, with roughly equal correlation lengths ξ for the predators and prey. The cross-correlation function $C_{AB}(x)$ displays a maximum at six lattice spacings; these positive correlations indicate the spatial extent of the emerging activity fronts (prey followed by the predators). At closer distance, the A and B parti-

Field Theoretic Methods, Figure 3
Space-time plot (space horizontal, with periodic boundary conditions; time vertical, proceeding downward) showing the temporal evolution of a one-dimensional stochastic Lotka–Volterra model on 512 lattice sites, but without any restrictions on the site occupation numbers (*red*: predators, *blue*: prey, *magenta*: sites occupied by both species; rates: $\sigma = 0.1$, $\mu = 0.1$, $\lambda = 0.1$). Reproduced with permission from [5]

cles become *anti*-correlated ($C_{AB}(x) < 0$ for $|x| < 3$): prey would not survive close encounters with the predators. In a similar manner, one can address temporal correlations. These appear prominently in the space-time plot of Fig. 3 obtained for a Monte Carlo run on a one-dimensional lattice (no site occupation restrictions), indicating localized population explosion and extinction events.

Correlation Functions and Field Theory

The above example demonstrates that stochastic fluctuations and correlations induced by the dynamical inter-

actions may lead to important features that are not adequately described by mean-field approaches. We thus require tools that allow us to systematically account for fluctuations in the mathematical description of stochastic complex systems and evaluate characteristic correlations. Such a toolbox is provided through *field theory* representations that are conducive to the identification of underlying symmetries and have proven useful starting points for the construction of various approximation schemes. These methods were originally devised and elaborated in the theory of (quantum and classical) many-particle systems and quantum fields ([6,7,8,9,10,11,12,13] represent a sample of recent textbooks).

Generating Functions

The basic structure of these field theories rests in a (normalized) exponential probability distribution $\mathcal{P}[S_i]$ for the N relevant variables S_i, $i = 1, \ldots, N$: $\int \prod_{i=1}^{N} dS_i$ $\mathcal{P}[S_i] = 1$, where the integration extends over the allowed range of values for the S_i; i.e.,

$$
\begin{aligned}
\mathcal{P}[S_i] &= \frac{1}{\mathcal{Z}} \exp\left(-\mathcal{A}[S_i]\right) , \\
\mathcal{Z} &= \int \prod_{i=1}^{N} dS_i \exp\left(-\mathcal{A}[S_i]\right) .
\end{aligned}
\tag{2}
$$

In canonical equilibrium statistical mechanics, $\mathcal{A}[S_i] = \mathcal{H}[S_i]/k_B T$ is essentially the Hamiltonian, and the normalization is the partition function $\mathcal{Z}$. In Euclidean quantum field theory, the action $\mathcal{A}[S_i]$ is given by the Langrangian.

All observables $\mathcal{O}$ should be functions of the basic degrees of freedom S_i; their ensemble average thus becomes

$$
\begin{aligned}
\langle \mathcal{O}[S_i] \rangle &= \int \prod_{i=1}^{N} dS_i \mathcal{O}[S_i] \mathcal{P}[S_i] \\
&= \frac{1}{\mathcal{Z}} \int \prod_{i=1}^{N} dS_i \mathcal{O}[S_i] \exp\left(-\mathcal{A}[S_i]\right) .
\end{aligned}
\tag{3}
$$

If we are interested in n-point correlations, i.e., expectation values of the products of the variables S_i, it is useful to define a *generating function*

$$
\mathcal{W}[j_i] = \left\langle \exp \sum_{i=1}^{N} j_i S_i \right\rangle ,
\tag{4}
$$

with $\mathcal{W}[j_i = 0] = 1$. Notice that $\mathcal{W}[j_i]$ formally is just the Laplace transform of the probability distribution $\mathcal{P}[S_i]$. The correlation functions can now be obtained via

partial derivatives of $\mathcal{W}[j_i]$ with respect to the sources j_i:

$$
\langle S_{i_1} \ldots S_{i_n} \rangle = \frac{\partial}{\partial j_{i_1}} \ldots \frac{\partial}{\partial j_{i_n}} \mathcal{W}[j_i] \Big|_{j_i = 0} .
\tag{5}
$$

Connected correlation functions or cumulants can be found by similar partial derivatives of the logarithm of the generating function:

$$
\langle S_{i_1} \ldots S_{i_n} \rangle_c = \frac{\partial}{\partial j_{i_1}} \ldots \frac{\partial}{\partial j_{i_n}} \ln \mathcal{W}[j_i] \Big|_{j_i = 0} ,
\tag{6}
$$

e.g., $\langle S_i \rangle_c = \langle S_i \rangle$, and $\langle S_i S_j \rangle_c = \langle S_i S_j \rangle - \langle S_i \rangle \langle S_j \rangle = \langle (S_i - \langle S_i \rangle)(S_j - \langle S_j \rangle) \rangle$.

Perturbation Expansion

For a Gaussian action, i.e., a quadratic form $\mathcal{A}_0[S_i] = \frac{1}{2} \sum_{ij} S_i A_{ij} S_j$ (for simplicity we assume real variables S_i), one may readily compute the corresponding generating function $\mathcal{W}_0[j_i]$. After diagonalizing the symmetric $N \times N$ matrix A_{ij}, completing the squares, and evaluating the ensuing Gaussian integrals, one obtains

$$
Z_0 = \frac{(2\pi)^{N/2}}{\sqrt{\det A}} ,
$$

$$
\mathcal{W}_0[j_i] = \exp\left(\frac{1}{2} \sum_{i,j=1}^{N} j_i A_{ij}^{-1} j_j \right) , \quad \langle S_i S_j \rangle_0 = A_{ij}^{-1} .
\tag{7}
$$

Thus, the two-point correlation functions in the Gaussian ensemble are given by the elements of the inverse harmonic coupling matrix. An important special property of the Gaussian ensemble is that all n-point functions with odd n vanish, whereas those with even n factorize into sums of all possible permutations of products of two-point functions A_{ij}^{-1} that can be constructed by pairing up the variables S_i (Wick's theorem). For example, the four-point function reads $\langle S_i S_j S_k S_l \rangle_0 = A_{ij}^{-1} A_{kl}^{-1} + A_{ik}^{-1} A_{jl}^{-1} + A_{il}^{-1} A_{jk}^{-1}$.

Let us now consider a general action, isolate the Gaussian contribution, and label the remainder as the nonlinear, anharmonic, or interacting part, $\mathcal{A}[S_i] = \mathcal{A}_0[S_i] + \mathcal{A}_f[S_i]$. We then observe that

$$
\mathcal{Z} = \mathcal{Z}_0 \left\langle \exp\left(-\mathcal{A}_f[S_i]\right) \right\rangle_0 ,
$$

$$
\langle \mathcal{O}[S_i] \rangle = \frac{\left\langle \mathcal{O}[S_i] \exp\left(-\mathcal{A}_f[S_i]\right) \right\rangle_0}{\left\langle \exp\left(-\mathcal{A}_f[S_i]\right) \right\rangle_0} ,
\tag{8}
$$

where the index 0 indicates that the expectation values are computed in the Gaussian ensemble. The nonlinear terms

in Eq. (8) may now be treated perturbatively by expanding the exponentials in the numerator and denominator with respect to the interacting part $\mathcal{A}_f[S_i]$:

$$\langle \mathcal{O}[S_i] \rangle = \frac{\left\langle \mathcal{O}[S_i] \sum_{\ell=0}^{\infty} \frac{1}{\ell!} \left(-\mathcal{A}_f[S_i] \right)^{\ell} \right\rangle_0}{\left\langle \sum_{\ell=0}^{\infty} \frac{1}{\ell!} \left(-\mathcal{A}_f[S_i] \right)^{\ell} \right\rangle_0} . \qquad (9)$$

If the interaction terms are polynomial in the variables S_i, Wick's theorem reduces the calculation of n-point functions to a summation of products of Gaussian two-point functions. Since the number of contributing terms grows factorially with the order ℓ of the perturbation expansion, graphical representations in terms of Feynman diagrams become very useful for the classification and evaluation of the different contributions to the perturbation series. Basically, they consist of lines representing the Gaussian two-point functions ('propagators') that are connected to vertices that stem from the (polynomial) interaction terms; for details, see, e. g., [6,7,8,9,10,11,12,13].

Continuum Limit and Functional Integrals

Discrete spatial degrees of freedom are already contained in the above formal description: for example, on a d-dimensional lattice with N_d sites the index i for the fields S_i merely needs to entail the site labels, and the total number of degrees of freedom is just $N = N_d$ times the number of independent relevant quantities. Upon discretizing time, these prescriptions can be extended in effectively an additional dimension to systems with temporal evolution. We may at last take the *continuum limit* by letting $N \to \infty$, while the lattice constant and elementary time step tend to zero in such a manner that macroscopic dynamical features are preserved. Formally, this replaces sums over lattice sites and time steps with spatial and temporal integrations; the action $\mathcal{A}[S_i]$ becomes a functional of the fields $S_i(x, t)$; partial derivatives turn into functional derivatives; and functional integrations $\int \prod_{i=1}^{N} dS_i \to \int \mathcal{D}[S_i]$ are to be inserted in the previous expressions. For example, Eqs. (3), (4) and (6) become

$$\langle \mathcal{O}[S_i] \rangle = \frac{1}{\mathcal{Z}} \int \mathcal{D}[S_i] \mathcal{O}[S_i] \exp\left(-\mathcal{A}[S_i] \right) , \qquad (10)$$

$$\mathcal{W}[j_i] = \left\langle \exp \int d^d x \int dt \sum_i j_i(x, t) S_i(x, t) \right\rangle , \quad (11)$$

$$\left\langle \prod_{j=1}^{n} S_{i_j}(x_j, t_j) \right\rangle_c = \prod_{j=1}^{n} \frac{\delta}{\delta j_{i_j}(x_j, t_j)} \ln \mathcal{W}[j_i] \Big|_{j_i=0} . \quad (12)$$

Thus we have arrived at a continuum field theory. Nevertheless, we may follow the procedures outlined above;

specifically, the perturbation expansion expressions (8) and (9) still hold, yet with arguments $S_i(x, t)$ that are now fields depending on continuous space-time parameters.

More than thirty years ago, Janssen and De Dominicis independently derived a mapping of the stochastic kinetics defined through nonlinear Langevin equations onto a field theory action ([14,15]; reviewed in [16]). Almost simultaneously, Doi constructed a Fock space representation and therefrom a stochastic field theory for classical interacting particle systems from the master equation describing the corresponding stochastic processes [17,18]. His approach was further developed by several authors into a powerful method for the study of internal noise and correlation effects in reaction-diffusion systems ([19,20,21,22,23]; for recent reviews, see [24,25]). We shall see below that the field-theoretic representations of both classical master and Langevin equations require *two* independent fields for each stochastic variable. Otherwise, the computation of correlation functions and the construction of perturbative expansions fundamentally works precisely as sketched above. But the underlying causal temporal structure induces important specific features such as the absence of 'vacuum diagrams' (closed response loops): the denominator in Eq. (2) is simply $\mathcal{Z} = 1$. (For unified and more detailed descriptions of both versions of dynamic stochastic field theories, see [26,27].)

Discrete Stochastic Interacting Particle Systems

We first outline the mapping of stochastic interacting particle dynamics as defined through a master equation onto a field theory action [17,18,19,20,21,22,23]. Let us denote the configurational probability for a stochastically evolving system to be in state α at time t with $P(\alpha; t)$. Given the transition rates $W_{\alpha \to \beta}(t)$ from states α to β, a *master equation* essentially balances the transitions into and out of each state:

$$\frac{\partial P(\alpha; t)}{\partial t} = \sum_{\beta \neq \alpha} \left[W_{\beta \to \alpha}(t) P(\beta; t) - W_{\alpha \to \beta}(t) P(\alpha; t) \right] .$$

$$(13)$$

The dynamics of many complex systems can be cast into the language of 'chemical' reactions, wherein certain particle species (upon encounter, say) transform into different species with fixed (time-independent) reaction rates. The 'particles' considered here could be atoms or molecules in chemistry, but also individuals in population dynamics (as in our example in Sect. "Example: Lotka–Volterra Model"), or appropriate effective degrees of freedom governing the system's kinetics, such as domain walls in mag-

nets, etc. To be specific, we envision our particles to propagate via unbiased random walks (diffusion) on a d-dimensional hypercubic lattice, with the reactions occuring according to prescribed rules when particles meet on a lattice site. This stochastic interacting particle system is then at any time fully characterized by the number of particles $n_A, n_B, \ldots$ of each species $A, B, \ldots$ located on any lattice site. The following describes the construction of an associated field theory action. As important examples, we briefly discuss annihilation reactions and absorbing state phase transitions.

Master Equation and Fock Space Representation

The formal procedures are best explained by means of a simple example; thus consider the irreversible binary annihilation process $A + A \to A$, happening with rate λ. In terms of the occupation numbers n_i of the lattice sites i, we can construct the master equation associated with these on-site reactions as follows. The annihilation process locally changes the occupation numbers by one; the transition rate from a state with n_i particles at site i to $n_i - 1$ particles is $W_{n_i \to n_i - 1} = \lambda n_i(n_i - 1)$, whence

$$\frac{\partial P(n_i; t)}{\partial t} = \lambda(n_i + 1)n_i P(n_i + 1; t) - \lambda n_i(n_i - 1)P(n_i; t)$$
(14)

represents the master equation for this reaction at site i. As an initial condition, we can for example choose a Poisson distribution $P(n_i) = \bar{n}_0^{n_i} e^{-\bar{n}_0}/n_i!$ with mean initial particle density $\bar{n}_0$. In order to capture the complete stochastic dynamics, we just need to add similar contributions describing other processes, and finally sum over all lattice sites i.

Since the reactions all change the site occupation numbers by integer values, a Fock space representation (borrowed from quantum mechanics) turns out particularly useful. To this end, we introduce the harmonic oscillator or bosonic ladder operator algebra $[a_i, a_j] = 0 = [a_i^\dagger, a_j^\dagger]$, $[a_i, a_j^\dagger] = \delta_{ij}$, from which we construct the particle number eigenstates $|n_i\rangle$, namely $a_i |n_i\rangle = n_i |n_i - 1\rangle$, $a_i^\dagger |n_i\rangle = |n_i + 1\rangle$, $a_i^\dagger a_i |n_i\rangle = n_i |n_i\rangle$). (Notice that a different normalization than in ordinary quantum mechanics has been employed here.) A general state with n_i particles on sites i is obtained from the 'vacuum' configuration $|0\rangle$, defined via $a_i |0\rangle = 0$, through the product $|\{n_i\}\rangle = \prod_i a_i^{\dagger n_i} |0\rangle$.

To implement the stochastic kinetics, we introduce a formal state vector as a linear combination of all possible states weighted by the time-dependent configurational probability:

$$|\Phi(t)\rangle = \sum_{\{n_i\}} P(\{n_i\}; t)|\{n_i\}\rangle .$$
(15)

Simple manipulations then transform the linear time evolution according to the master equation into an 'imaginary-time' Schrödinger equation

$$\frac{\partial |\Phi(t)\rangle}{\partial t} = -H|\Phi(t)\rangle , \quad |\Phi(t)\rangle = e^{-Ht}|\Phi(0)\rangle$$
(16)

governed by a stochastic quasi-Hamiltonian (rather, the Liouville time evolution operator). For on-site reaction processes, $H_{\text{reac}} = \sum_i H_i(a_i^\dagger, a_i)$ is a sum of local contributions; e.g., for the binary annihilation reaction, $H_i(a_i^\dagger, a_i) = -\lambda(1 - a_i^\dagger)a_i^\dagger a_i^2$. It is a straightforward exercise to construct the corresponding expressions within this formalism for the generalization $kA \to \ell A$,

$$H_i(a_i^\dagger, a_i) = -\lambda \left(a_i^{\dagger \ell} - a_i^{\dagger k} \right) a_i^k,$$
(17)

and for nearest-neighbor hopping with rate D between adjacent sites $\langle ij \rangle$,

$$H_{\text{diff}} = D \sum_{<ij>} \left(a_i^\dagger - a_j^\dagger \right) \left(a_i - a_j \right) .$$
(18)

The two contributions for each process may be interpreted as follows: The first term in Eq. (17) corresponds to the actual process, and describes how many particles are annihilated and (re-)created in each reaction. The second term encodes the 'order' of each reaction, i. e., the number operator $a_i^\dagger a_i$ appears to the kth power, but in the normal-ordered form $a_i^{\dagger k} a_i^k$, for a kth-order process. These procedures are readily adjusted for reactions involving multiple particle species. We merely need to specify the occupation numbers on each site and correspondingly introduce additional ladder operators $b_i, c_i, \ldots$ for each new species, with $[a_i, b_i^\dagger] = 0 = [a_i, c_i^\dagger]$ etc. For example, consider the reversible reaction $kA + \ell B \rightleftharpoons mC$ with forward rate λ and backward rate σ; the associated reaction Hamiltonian reads

$$H_{\text{reac}} = -\sum_i \left(c_i^{\dagger m} - a_i^{\dagger k} b_i^{\dagger \ell} \right) \left(\lambda a_i^k b_i^\ell - \sigma c_i^m \right) .$$
(19)

Similarly, for the Lotka–Volterra model of Sect. "Example: Lotka–Volterra Model", one finds

$$H_{\text{reac}} = -\sum_i \left[\mu \left(1 - a_i^\dagger \right) a_i + \sigma \left(b_i^\dagger - 1 \right) b_i^\dagger b_i \right.$$
$$\left. + \lambda \left(a_i^\dagger - b_i^\dagger \right) a_i^\dagger a_i b_i \right]. \quad (20)$$

Note that all the above quasi-Hamiltonians are non-Hermitean operators, which naturally reflects the creation and destruction of particles.

Our goal is to compute averages and correlation functions with respect to the configurational probability $P(\{n_i\}; t)$. Returning to a single-species system (again, the generalization to many particle species is obvious), this is accomplished with the aid of the projection state $\langle \mathcal{P} | = \langle 0 | \prod_i e^{a_i}$, for which $\langle \mathcal{P} | 0 \rangle = 1$ and $\langle \mathcal{P} | a_i^\dagger = \langle \mathcal{P} |$, since $[e^{a_i}, a_j^\dagger] = e^{a_i} \delta_{ij}$. For the desired statistical averages of observables (which must all be expressible as functions of the occupation numbers $\{n_i\}$), one obtains

$$\langle \mathcal{O}(t) \rangle = \sum_{\{n_i\}} \mathcal{O}(\{n_i\}) P(\{n_i\}; t) = \langle \mathcal{P} | \mathcal{O}(\{a_i^\dagger a_i\}) | \Phi(t) \rangle . \tag{21}$$

For example, as a consequence of probability conservation, $1 = \langle \mathcal{P} | \Phi(t) \rangle = \langle \mathcal{P} | e^{-Ht} | \Phi(0) \rangle$. Thus necessarily $\langle \mathcal{P} | H = 0$; upon commuting $e^{\sum_i a_i}$ with H, the creation operators are shifted $a_i^\dagger \to 1 + a_i^\dagger$, whence this condition is fulfilled provided $H_i(a_i^\dagger \to 1, a_i) = 0$, which is indeed satisfied by our above explicit expressions (17) and (18). Through this prescription, we may replace $a_i^\dagger a_i \to a_i$ in all averages; e.g., the particle density becomes $a(t) = \langle a_i(t) \rangle$.

In the bosonic operator representation above, we have assumed that no restrictions apply to the particle occupation numbers n_i on each site. If $n_i \leq 2s + 1$, one may instead employ a representation in terms of spin s operators. For example, particle exclusion systems with $n_i = 0$ or 1 can thus be mapped onto non-Hermitean spin 1/2 'quantum' systems (for recent overviews, see [28,29]). Specifically in one dimension, such representations in terms of integrable spin chains have been very fruitful. An alternative approach uses the bosonic theory, but incorporates the site occupation restrictions through exponentials in the number operators $e^{-a_i^\dagger a_i}$ [30].

Continuum Limit and Field Theory

As a next step, we follow an established route in quantum many-particle theory [8] and proceed towards a field theory representation through constructing the path integral equivalent to the 'Schrödinger' dynamics (16) based on coherent states, which are right eigenstates of the annihilation operator, $a_i | \phi_i \rangle = \phi_i | \phi_i \rangle$, with complex eigenvalues ϕ_i. Explicitly, $| \phi_i \rangle = \exp\left(-\frac{1}{2} |\phi_i|^2 + \phi_i a_i^\dagger \right) | 0 \rangle$, and these coherent states satisfy the overlap formula $\langle \phi_j | \phi_i \rangle = \exp\left(-\frac{1}{2} |\phi_i|^2 - \frac{1}{2} |\phi_j|^2 + \phi_j^* \phi_i \right)$, and the

(over-)completeness relation $\int \prod_i d^2 \phi_i | \{\phi_i\} \rangle \langle \{\phi_i\} | = \pi$. Upon splitting the temporal evolution (16) into infinitesimal increments, standard procedures (elaborated in detail in [25]) eventually yield an expression for the configurational average

$$\langle \mathcal{O}(t) \rangle \propto \int \prod_i d\phi_i \, d\phi_i^* \, \mathcal{O}(\{\phi_i\}) e^{-\mathcal{A}[\phi_i^*, \phi_i; t]} , \tag{22}$$

which is of the form (3), with the action

$$\mathcal{A}[\phi_i^*, \phi_i; t_f] = \sum_i \left(-\phi_i(t_f) \right.$$
$$\left. + \int_0^{t_f} dt \left[\phi_i^* \frac{\partial \phi_i}{\partial t} + H_i(\phi_i^*, \phi_i) \right] - \bar{n}_0 \phi_i^*(0) \right), \tag{23}$$

where the first term originates from the projection state, and the last one stems from the initial Poisson distribution. Through this procedure, in the original quasi-Hamiltonian the creation and annihilation operators $a_i^\dagger$ and a_i are simply replaced with the complex numbers ϕ_i^* and ϕ_i.

Finally, we proceed to the continuum limit, $\phi_i(t) \to \psi(\mathbf{x}, t)$, $\phi_i^*(t) \to \hat{\psi}(\mathbf{x}, t)$. The 'bulk' part of the action then becomes

$$\mathcal{A}[\hat{\psi}, \psi] = \int d^d x$$
$$\cdot \int dt \left[\hat{\psi} \left(\frac{\partial}{\partial t} - D\nabla^2 \right) \psi + \mathcal{H}_{\text{reac}}(\hat{\psi}, \psi) \right], \tag{24}$$

where the discrete hopping contribution (18) has naturally turned into a continuum diffusion term. We have thus arrived at a *microscopic* field theory for stochastic reaction–diffusion processes, without invoking any assumptions on the form or correlations of the internal reaction noise. Note that we require two independent fields $\hat{\psi}$ and ψ to capture the stochastic dynamics. Actions of the type (24) may serve as a basis for further systematic coarse-graining, constructing a perturbation expansion as outlined in Sect. "Perturbation Expansion", and perhaps a subsequent renormalization group analysis [25,26,27]. We remark that it is often useful to perform a shift in the field $\hat{\psi}$ about the mean-field solution, $\hat{\psi}(x, t) = 1 + \widetilde{\psi}(x, t)$. For occasionally, the resulting field theory action allows the derivation of an equivalent Langevin dynamics, see Sect. "Stochastic Differential Equations" below.

Annihilation Processes

Let us consider our simple single-species example $kA \to \ell A$. The reaction part of the corresponding field

theory action reads

$$\mathcal{H}_{\text{reac}}(\hat{\psi}, \psi) = -\lambda \left(\hat{\psi}^{\ell} - \hat{\psi}^{k} \right) \psi^{k} , \qquad (25)$$

see Eq. (17). It is instructive to study the *classical field equations*, namely $\delta \mathcal{A} / \delta \psi = 0$, which is always solved by $\hat{\psi} = 1$, reflecting probability conservation, and $\delta \mathcal{A} / \delta \hat{\psi} = 0$, which, upon inserting $\hat{\psi} = 1$ yields

$$\frac{\partial \psi(x,t)}{\partial t} = D\nabla^2 \psi(x,t) - (k-\ell)\lambda \psi(x,t)^k , \qquad (26)$$

i. e., the mean-field equation for the local particle density $\psi(x,t)$, supplemented with a diffusion term. For $k = 1$, the particle density grows ($k < \ell$) or decays ($k > \ell$) exponentially. The solution of the rate equation for $k > 1$, $a(t) = \langle \psi(x,t) \rangle = \left[a(0)^{1-k} + (k-l)(k-1)\lambda t \right]^{-1/(k-1)}$ implies a divergence within a finite time for $k < \ell$, and an algebraic decay $\sim (\lambda t)^{-1/(k-1)}$ for $k > \ell$.

The full field theory action, which was derived from the master equation defining the very stochastic process, provides a means of systematically including fluctuations in the mathematical treatment. Through a dimensional analysis, we can determine the (upper) *critical dimension* below which fluctuations become sufficiently strong to alter these power laws. Introducing an inverse length scale κ, $[x] \sim \kappa^{-1}$, and applying diffusive temporal scaling, $[Dt] \sim \kappa^{-2}$, and $[\hat{\psi}(x,t)] \sim \kappa^0$, $[\psi(x,t)] \sim \kappa^d$ in d spatial dimensions, the reaction rate in terms of the diffusivity scales according to $[\lambda/D] \sim \kappa^{2-(k-1)d}$. In large dimensions, the kinetics is *reaction-limited*, and at least qualitatively correctly described by the mean-field rate equation. In low dimensions, the dynamics becomes *diffusion-limited*, and the annihilation reactions generate depletion zones and spatial particle anti-correlations that slow down the density decay. The nonlinear coupling λ/D becomes dimensionless at the boundary critical dimension $d_c(k) = 2/(k-1)$ that separates these two distinct regimes. Thus in physical dimensions, intrinsic stochastic fluctuations are relevant only for pair and triplet annihilation reactions. By means of a renormalization group analysis (for details, see [25]) one finds for $k = 2$ and $d < d_c(2) = 2$: $a(t) \sim (Dt)^{-d/2}$ [21,22], as confirmed by exact solutions in one dimension. Precisely at the critical dimension, the mean-field decay laws acquire logarithmic corrections, namely $a(t) \sim (Dt)^{-1} \ln(Dt)$ for $k = 2$ at $d_c(2) = 2$, and $a(t) \sim \left[(Dt)^{-1} \ln(Dt) \right]^{1/2}$ for $k = 3$ at $d_c(3) = 1$. Annihilation reaction between different species (e. g., $A + B \to \emptyset$) may introduce additional correlation effects, such as particle segregation and the confinement of active dynamics to narrow reaction zones [23]; a recent overview can be found in [25].

Active to Absorbing State Phase Transitions

Competition between particle production and decay processes leads to even richer scenarios, and can induce genuine nonequilibrium transitions that separate 'active' phases (wherein the particle densities remain nonzero in the long-time limit) from 'inactive' stationary states (where the concentrations ultimately vanish). A special but abundant case are *absorbing states*, where, owing to the absence of any agents, stochastic fluctuations cease entirely, and no particles can be regenerated [31,32]. These occur in a variety of systems in nature ([33,34] contain extensive discussions of various model systems); examples are chemical reactions involving an inert state Ø, wherefrom no reactants A are released anymore, or stochastic population dynamics models, combining diffusive migration of a species A with asexual reproduction $A \to 2A$ (with rate σ), spontaneous death $A \to \emptyset$ (at rate μ), and lethal competition $2A \to A$ (with rate λ). In the inactive state, where no population members A are left, clearly all processes terminate. Similar effective dynamics may be used to model certain nonequilibrium physical systems, such as the domain wall kinetics in Ising chains with competing Glauber and Kawasaki dynamics. Here, spin flips $\uparrow\uparrow\downarrow\downarrow \to \uparrow\uparrow\uparrow\downarrow$ and $\uparrow\uparrow\downarrow\uparrow \to \uparrow\uparrow\uparrow\uparrow$ may be viewed as domain wall (A) hopping and pair annihilation $2A \to \emptyset$, whereas spin exchange $\uparrow\uparrow\downarrow\downarrow \to \uparrow\downarrow\uparrow\downarrow$ represents a branching process $A \to 3A$. Notice that the para- and ferromagnetic phases respectively map onto the active and inactive 'particle' states. The ferromagnetic state becomes absorbing if the spin flip rates are taken at zero temperature.

The reaction quasi-Hamiltonian corresponding to the stochastic dynamics of the aforementioned population dynamics model reads

$$\mathcal{H}_{\text{reac}}(\hat{\psi}, \psi) = \left(1 - \hat{\psi} \right) \left(\sigma \hat{\psi}\psi - \mu\psi - \lambda\hat{\psi}\psi^2 \right) . \quad (27)$$

The associated rate equation is the Fisher–Kolmogorov equation (see Murray 2002 [3])

$$\dot{a}(t) = (\sigma - \mu)a(t) - \lambda a(t)^2 , \qquad (28)$$

which yields both inactive and active phases: For $\sigma < \mu$ we have $a(t \to \infty) \to 0$, whereas for $\sigma > \mu$ the density eventually saturates at $a_s = (\sigma - \mu)/\lambda$. The explicit time-dependent solution $a(t) = a(0)a_s / [a(0) + [a_s - a(0)]e^{(\mu-\sigma)t}]$ shows that both stationary states are approached exponentially in time. They are separated by a continuous nonequilibrium phase transition at $\sigma = \mu$, where the temporal decay becomes algebraic, $a(t) = a(0)/[1+a(0)\lambda t]) \to 1/(\lambda t)$ as $t \to \infty$, independent of the initial density $a(0)$. As in second-order equilibrium phase transitions, however, critical fluctuations are expected to

invalidate the mean-field power laws in low dimensions $d < d_c$.

If we now shift the field $\hat{\psi}$ about its stationary value 1 and rescale according to $\hat{\psi}(\mathbf{x}, t) = 1 + \sqrt{\sigma/\lambda} \, \widetilde{S}(\mathbf{x}, t)$ and $\psi(\mathbf{x}, t) = \sqrt{\lambda/\sigma} \, S(\mathbf{x}, t)$, the (bulk) action becomes

$$\mathcal{A}\left[\widetilde{S}, S\right] = \int d^d x \int dt \left[\widetilde{S}\left(\frac{\partial}{\partial t} + D\left(r - \mathbf{\nabla}^2\right)\right) S \right.$$
$$\left. - u\left(\widetilde{S} - S\right)\widetilde{S}S + \lambda\widetilde{S}^2 S^2\right]. \quad (29)$$

Thus, the three-point vertices have been scaled to identical coupling strengths $u = \sqrt{\sigma\lambda}$, which in fact represents the effective coupling of the perturbation expansion. Its scaling dimension is $[u] = \mu^{2-d/2}$, whence we infer the upper critical dimension $d_c = 4$. The four-point vertex $\propto \lambda$, with $[\lambda] = \mu^{2-d}$, is then found to be irrelevant in the renormalization group sense, and can be dropped for the computation of universal, asymptotic scaling properties. The action (29) with $\lambda = 0$ is known as Reggeon field theory [35]; it satisfies a characteristic symmetry, namely invariance under so-called rapidity inversion $S(\mathbf{x}, t) \leftrightarrow -\widetilde{S}(\mathbf{x}, -t)$. Remarkably, it has moreover been established that the field theory action (29) describes the scaling properties of critical directed percolation clusters [36,37,38]. The fluctuation-corrected universal power laws governing the vicinity of the phase transition can be extracted by renormalization group methods (reviewed for directed percolation in [39]). Table 1 compares the analytic results obtained in an ε expansion about the critical dimension ($\epsilon = 4 - d$) with the critical exponent values measured in Monte Carlo computer simulations [33,34].

According to a conjecture originally formulated by Janssen and Grassberger, any continuous nonequilibrium phase transition from an active to an absorbing state in a system governed by Markovian stochastic dynamics that is decoupled from any other slow variable, and in the absence of special additional symmetries or quenched randomness, should in fact fall in the directed percolation universality class [38,40]. This statement has indeed been confirmed in a large variety of model sytems (many examples are listed in [33,34]). It even pertains to multi-species generalizations [41], and applies for instance to the predator extinction threshold in the stochastic Lotka–Volterra model with restricted site occupation numbers mentioned in Sect. "Example: Lotka–Volterra Model" [4].

Stochastic Differential Equations

This section explains how dynamics governed by *Langevin-type stochastic differential equations* can be represented through a field-theoretic formalism [14,15,16]. Such a description is especially useful to capture the effects of external noise on the temporal evolution of the relevant quantities under consideration, which encompasses the case of thermal noise induced by the coupling to a heat bath in thermal equilibrium at temperature T. The underlying assumption in this approach is that there exists a natural *separation of time scales* between the slow variables S_i, and all other degrees of freedom ζ_i which in comparison fluctuate rapidly, and are therefore summarily gathered in zero-mean noise terms, assumed to be uncorrelated in space and time,

$$\langle \zeta_i(x, t) \rangle = 0 \,,$$
$$\langle \zeta_i(x, t)\zeta_j(x', t') \rangle = 2L_{ij}[S_i]\delta(x - x')\delta(t - t') \,. \quad (30)$$

Here, the noise correlator $2L_{ij}[S_i]$ may be a function of the slow system variables S_i, and also contain operators such as spatial derivatives. A general set of coupled Langevin-type stochastic differential equations then takes the form

$$\frac{\partial S_i(t)}{\partial t} = F_i[S_i] + \zeta_i \,, \quad (31)$$

where we may decompose the 'systematic forces' into reversible terms of microscopic origin and relaxational contributions that are induced by the noise and drive the system towards its stationary state (see below), i. e.: $F_i[S_i] = F_i^{\text{rev}}[S_i] + F_i^{\text{rel}}[S_i]$. Both ingredients may contain nonlinear terms as well as mode couplings between

Field Theoretic Methods, Table 1
Comparison of the values for the critical exponents of the directed percolation universality class measured in Monte Carlo simulations with the analytic renormalization group results within the $\epsilon = 4 - d$ expansion: ξ denotes the correlation length, t_c the characteristic relaxation time, a_s the saturation density in the active state, and $a_c(t)$ the critical density decay law

Scaling exponent	d = 1	d = 2	d = 4 − ε
$\xi \sim \|\tau\|^{-\nu}$	$\nu \approx 1.100$	$\nu \approx 0.735$	$\nu = 1/2 + \epsilon/16 + O(\epsilon^2)$
$t_c \sim \xi^z \sim \|\tau\|^{-z\nu}$	$z \approx 1.576$	$z \approx 1.73$	$z = 2 - \epsilon/12 + O(\epsilon^2)$
$a_s \sim \|\tau\|^\beta$	$\beta \approx 0.2765$	$\beta \approx 0.584$	$\beta = 1 - \epsilon/6 + O(\epsilon^2)$
$a_c(t) \sim t^{-\alpha}$	$\alpha \approx 0.160$	$\alpha \approx 0.46$	$\alpha = 1 - \epsilon/4 + O(\epsilon^2)$

different variables. Again, we first introduce the abstract formalism, and then proceed to discuss relaxation to thermal equilibrium as well as some examples for nonequilibrium Langevin dynamics.

Field Theory Representation of Langevin Equations

The shortest and most general route towards a field theory representation of the Langevin dynamics (31) with noise correlations (30) starts with one of the most elaborate ways to expand unity, namely through a product of functional delta functions (for the sake of compact notations, we immediately employ a functional integration language, but in the end all the path integrals are defined through appropriate discretizations in space and time):

$$
1 = \int \prod_i \mathcal{D}[S_i]
$$
$$
\cdot \prod_{(x,t)} \delta\left(\frac{\partial S_i(x,t)}{\partial t} - F_i[S_i](x,t) - \zeta_i(x,t) \right)
$$
$$
= \int \prod_i \mathcal{D}[i\widetilde{S}_i]\mathcal{D}[S_i] \exp\left[-\int d^d x \right.
$$
$$
\left. \cdot \int dt \sum_i \widetilde{S}_i \left(\frac{\partial S_i}{\partial t} - F_i[S_i] - \zeta_i \right) \right]. \quad (32)
$$

In the second line we have used the Fourier representation of the (functional) delta distribution by means of the purely imaginary auxiliary variables $\widetilde{S}_i$ (also called Martin–Siggia–Rose response fields [42]). Next we require the explicit form of the noise probability distribution that generates the correlations (30); for simplicity, we may employ the Gaussian

$$
\mathcal{W}[\zeta_i] \propto \exp\left[-\frac{1}{4} \int d^d x \right.
$$
$$
\left. \cdot \int_0^{t_f} dt \sum_{ij} \zeta_i(x,t) \left[L_{ij}^{-1} \zeta_j(x,t) \right] \right]. \quad (33)
$$

Inserting the identity (32) and the probability distribution (33) into the desired stochastic noise average of any observable $\mathcal{O}[S_i]$, we arrive at

$$
\langle \mathcal{O}[S_i] \rangle_\zeta \propto \int \prod_i \mathcal{D}[i\widetilde{S}_i]\mathcal{D}[S_i]
$$
$$
\exp\left[-\int d^d x \int dt \sum_i \widetilde{S}_i \left(\frac{\partial S_i}{\partial t} - F_i[S_i] \right) \right] \mathcal{O}[S_i]
$$
$$
\cdot \int \prod_i \mathcal{D}[\zeta_i] \exp\left(-\int d^d x \int dt \right.
$$
$$
\left. \cdot \sum_i \left[\frac{1}{4}\zeta_i \sum_j L_{ij}^{-1}\zeta_j - \widetilde{S}_i\zeta_i \right] \right). \quad (34)
$$

Subsequently evaluating the Gaussian integrals over the noise ζ_i yields at last

$$
\langle \mathcal{O}[S_i] \rangle_\zeta = \int \prod_i \mathcal{D}[S_i]\mathcal{O}[S_i]\mathcal{P}[S_i] ,
$$
$$
\mathcal{P}[S_i] \propto \int \prod_i \mathcal{D}[i\widetilde{S}_i]e^{-\mathcal{A}[\widetilde{S}_i, S_i]} , \quad (35)
$$

with the statistical weight governed by the Janssen–De Dominicis 'response' functional [14,15]

$$
\mathcal{A}[\widetilde{S}_i, S_i] = \int d^d x \int_0^{t_f} dt
$$
$$
\cdot \sum_i \left[\widetilde{S}_i \left(\frac{\partial S_i}{\partial t} - F_i[S] \right) - \widetilde{S}_i \sum_j L_{ij}\widetilde{S}_j \right]. \quad (36)
$$

It should be noted that in the above manipulations, we have omitted the functional determinant from the variable change $\{\zeta_i\} \to \{S_i\}$. This step can be justified through applying a forward (Itô) discretization (for technical details, see [16,27,43]). Normalization implies $\int \prod_i \mathcal{D}[i\widetilde{S}_i]\mathcal{D}[S_i]e^{-\mathcal{A}[\widetilde{S}_i, S]_i} = 1$. The first term in the action (36) encodes the temporal evolution according to the systematic terms in the Langevin Equations (31), whereas the second term specifies the noise correlations (30). Since the auxiliary fields appear only quadratically, they could be eliminated via completing the squares and Gaussian integrations. This results in the equivalent Onsager–Machlup functional which however contains squares of the nonlinear terms and the inverse of the noise correlator operators; the form (36) is therefore usually more convenient for practical purposes. The Janssen–De Dominicis functional (36) takes the form of a $(d + 1)$-dimensional statistical field theory with again *two* independent sets of fields S_i and $\widetilde{S}_i$. It may serve as a starting point for systematic approximation schemes including perturbative expansions, and subsequent renormalization group treatments. Causality is properly incorporated in this formalism which has important technical implications [16,27,43].

Thermal Equilibrium and Relaxational Critical Dynamics

Consider the dynamics of a system that following some external perturbation relaxes towards thermal equilibrium governed by the canonical Boltzmann distribution at fixed temperature T,

$$
\mathcal{P}_{\text{eq}}[S_i] = \frac{1}{\mathcal{Z}(T)} \exp\left(-\mathcal{H}[S_i]/k_B T\right) . \quad (37)
$$

The relaxational term in the Langevin Equation (31) can then be specified as

$$F_i^{\text{rel}}[S_i] = -\lambda_i \frac{\delta \mathcal{H}[S_i]}{\delta S_i} \,, \tag{38}$$

with Onsager coefficients λ_i; for nonconserved fields, λ_i is a positive relaxation rate. On the other hand, if the variable S_i is a conserved quantity (such as the energy density), there is an associated continuity equation $\partial S_i / \partial t + \nabla \cdot J_i = 0$, with a conserved current that is typically given by a gradient of the field S_i: $J_i = -D_i \nabla S_i + \ldots$; as a consequence, the fluctuations of the fields S_i will relax diffusively with diffusivity D_i, and $\lambda_i = -D_i \nabla^2$ becomes a spatial Laplacian.

In order for $\mathcal{P}(t) \to \mathcal{P}_{\text{eq}}$ as $t \to \infty$, the stochastic Langevin dynamics needs to satisfy *two* conditions, which can be inferred from the associated Fokker–Planck equation [27,44]. First, the reversible probability current is required to be divergence-free in the space spanned by the fields S_i:

$$\int d^d x \sum_i \frac{\delta}{\delta S_i(x)} \left(F_i^{\text{rev}}[S_i] e^{-\mathcal{H}[S_i]/k_B T} \right) = 0 \,. \tag{39}$$

This condition severely constrains the reversible force terms. For example, for a system whose microscopic time evolution is determined through the Poisson brackets $Q_{ij}(x, x') = \{S_i(x), S_j(x')\} = -Q_{ji}(x', x)$ (to be replaced by commutators in quantum mechanics), one finds for the reversible mode-coupling terms [44]

$$F_i^{\text{rev}}[S_i](x) = -\int d^d x'$$
$$\cdot \sum_j \left[Q_{ij}(x, x') \frac{\delta \mathcal{H}[S_i]}{\delta S_j(x')} - k_B T \frac{\delta Q_{ij}(x, x')}{\delta S_j(x')} \right] \,. \tag{40}$$

Second, the noise correlator in Eq. (30) must be related to the Onsager relaxation coefficients through the Einstein relation

$$L_{ij} = k_B T \lambda_i \delta_{ij} \,. \tag{41}$$

To provide a specific example, we focus on the case of purely relaxational dynamics (i. e., reversible force terms are absent entirely), with the (mesoscopic) Hamiltonian given by the Ginzburg–Landau–Wilson free energy that describes second-order phase transitions in thermal equilibrium for an n-component order parameter S_i,

$i = 1, \ldots, N$ [6,7,8,9,10,11,12,13]:

$$\mathcal{H}[S_i] = \int d^d x \sum_{i=1}^{N} \left[\frac{r}{2} [S_i(x)]^2 + \frac{1}{2} [\nabla S_i(x)]^2 \right.$$
$$\left. + \frac{u}{4!} [S_i(x)]^2 \sum_{j=1}^{N} [S_j(x)]^2 \right], \tag{42}$$

where the control parameter $r \propto T - T_c$ changes sign at the critical temperature T_c, and the positive constant u governs the strength of the nonlinearity. If we assume that the order parameter itself is not conserved under the dynamics, the associated response functional reads

$$\mathcal{A}\left[\widetilde{S}_i, S_i \right] = \int d^d x \int dt$$
$$\cdot \sum_i \widetilde{S}_i \left(\frac{\partial}{\partial t} + \lambda_i \frac{\delta \mathcal{H}[S_i]}{\delta S_i} - k_B T \lambda_i \widetilde{S}_i \right) \,. \tag{43}$$

This case is frequently referred to as model A critical dynamics [45]. For a diffusively relaxing conserved field, termed model B in the classification of [45], one has instead

$$\mathcal{A}\left[\widetilde{S}_i, S_i \right] = \int d^d x \int dt$$
$$\cdot \sum_i \widetilde{S}_i \left(\frac{\partial}{\partial t} - D_i \nabla^2 \frac{\delta \mathcal{H}[S_i]}{\delta S_i} + k_B T D_i \nabla^2 \widetilde{S}_i \right) \,. \tag{44}$$

Consider now the external fields h_i that are thermodynamically conjugate to the mesoscopic variables S_i, i. e., $\mathcal{H}(h_i) = \mathcal{H}(h_i = 0) - \int d^d x \sum_i h_i(x) S_i(x)$. For the simple relaxational models (43) and (44), we may thus immediately relate the dynamic susceptibility to two-point correlation functions that involve the auxiliary fields $\widetilde{S}_i$ [43], namely

$$\chi_{ij}(x - x', t - t') = \frac{\delta \langle S_i(x, t) \rangle}{\delta h_j(x', t')} \bigg|_{h_i=0}$$
$$= k_B T \lambda_i \left\langle S_i(x, t) \widetilde{S}_j(x', t') \right\rangle \tag{45}$$

for nonconserved fields, while for model B dynamics

$$\chi_{ij}(x - x', t - t') = -k_B T D_i \left\langle S_i(x, t) \nabla^2 \widetilde{S}_j(x', t') \right\rangle \,. \tag{46}$$

Finally, in thermal equilibrium the dynamic response and correlation functions are related through the fluctuation-dissipation theorem [43]

$$\chi_{ij}(x - x', t - t') = \Theta(t - t') \frac{\partial}{\partial t'} \left\langle S_i(x, t) S_j(x', t') \right\rangle \,. \tag{47}$$

Driven Diffusive Systems and Interface Growth

We close this section by listing a few intriguing examples for Langevin sytems that describe genuine out-of-equilibrium dynamics. First, consider a driven diffusive lattice gas (an overview is provided in [46]), namely a particle system with conserved total density with biased diffusion in a specified ('∥') direction. The coarse-grained Langevin equation for the scalar density fluctuations thus becomes spatially anisotropic [47,48],

$$\frac{\partial S(x,t)}{\partial t}$$
$$= D\left(\nabla_\perp^2 + c\nabla_\parallel^2\right) S(x,t) + \frac{Dg}{2}\nabla_\parallel S(x,t)^2 + \zeta(x,t),$$
(48)

and similarly for the conserved noise with $\langle \zeta \rangle = 0$,

$$\langle \zeta(x,t)\zeta(x',t')\rangle = -2D\left(\nabla_\perp^2 + \tilde{c}\nabla_\parallel^2\right)\delta(x-x')\delta(t-t').$$
(49)

Notice that the drive term $\propto g$ breaks both the system's spatial reflection symmetry as well as the Ising symmetry $S \to -S$. In one dimension, Eq. (48) coincides with the noisy Burgers equation [49], and since in this case (only) the condition (39) is satisfied, effectively represents a system with equilibrium dynamics. The corresponding Janssen–De Dominicis response functional reads

$$\mathcal{A}\left[\tilde{S}, S\right] = \int d^d x \int dt \tilde{S}\left[\frac{\partial S}{\partial t} - D\left(\nabla_\perp^2 + c\nabla_\parallel^2\right) S\right.$$
$$\left. + D\left(\nabla_\perp^2 + \tilde{c}\nabla_\parallel^2\right)\tilde{S} - \frac{Dg}{2}\nabla_\parallel S^2\right].$$
(50)

It describes a 'massless' theory, hence we expect the system to generically display scale-invariant features, without the need to tune to a special point in parameter space. The large-scale scaling properties can be analyzed by means of the dynamic renormalization group [47,48].

Another famous example for generic scale invariance emerging in a nonequilibrium system is curvature-driven interface growth, as captured by the Kardar–Parisi–Zhang equation [50]

$$\frac{\partial S(x,t)}{\partial t} = D\nabla^2 S(x,t) + \frac{Dg}{2}[\nabla S(x,t)]^2 + \zeta(x,t), \quad (51)$$

with again $\langle \zeta \rangle = 0$ and the noise correlations

$$\langle \zeta(x,t)\zeta(x',t')\rangle = 2D\delta(x-x')\delta(t-t').$$
(52)

(For more details and intriguing variants, see e.g. [51, 52,53].) The associated field theory action

$$\mathcal{A}\left[\tilde{S}, S\right] = \int d^d x$$
$$\cdot \int dt\left[\tilde{S}\left(\frac{\partial S}{\partial t} - D\nabla^2 S - \frac{Dg}{2}[\nabla S]^2\right) - D\tilde{S}^2\right] \quad (53)$$

encodes surprisingly rich behavior including a kinetic roughening transition separating two distinct scaling regimes in dimensions $d > 2$ [51,52,53].

Future Directions

The rich phenomenology in many complex systems is only inadequately captured within widely used mean-field approximations, wherein both statistical fluctuations and correlations induced by the subunits' interactions or the system's kinetics are neglected. Modern computational techniques, empowered by recent vast improvements in data storage and tact frequencies, as well as the development of clever algorithms, are clearly invaluable in the theoretical study of model systems displaying the hallmark features of complexity. Yet in order to gain a deeper understanding and to maintain control over the typically rather large parameter space, numerical investigations need to be supplemented by analytical approaches. The field-theoretic methods described in this article represent a powerful set of tools to systematically include fluctuations and correlations in the mathematical description of complex stochastic dynamical systems composed of many interacting degrees of freedom. They have already been very fruitful in studying the intriguing physics of highly correlated and strongly fluctuating many-particle systems. Aside from many important quantitative results, they have provided the basis for our fundamental understanding of the emergence of universal macroscopic features.

At the time of writing, the transfer of field-theoretic methods to problems in chemistry, biology, and other fields such as sociology has certainly been initiated, but is still limited to rather few and isolated case studies. This is understandable, since becoming acquainted with the intricate technicalities of the field theory formalism requires considerable effort. Also, whereas it is straightforward to write down the actions corresponding the stochastic processes defined via microscopic classical discrete master or mesoscopic Langevin equations, it is usually not that easy to properly extract the desired information about large-scale structures and long-time asymptotics. Yet if successful, one tends to gain insights that are not accessible by any other means. I therefore anticipate that the now well-

developed methods of quantum and statistical field theory, with their extensions to stochastic dynamics, will find ample successful applications in many different areas of complexity science. Naturally, further approximation schemes and other methods tailored to the questions at hand will have to be developed, and novel concepts be devised. I look forward to learning about and hopefully also participating in these exciting future developments.

Acknowledgments

Bibliography

1. Lindenberg K, Oshanin G, Tachiya M (eds) (2007) J Phys: Condens Matter 19(6): Special issue containing articles on Chemical kinetics beyond the textbook: fluctuations, many-particle effects and anomalous dynamics; see: http://www.iop.org/EJ/toc/0953-8984/19/6
2. Alber M, Frey E, Goldstein R (eds) (2007) J Stat Phys 128(1/2): Special issue on Statistical physics in biology; see: http://springerlink.com/content/j4q1ln243968/
3. Murray JD (2002) Mathematical biology, vols. I, II, 3rd edn. Springer, New York
4. Mobilia M, Georgiev IT, Täuber UC (2007) Phase transitions and spatio-temporal fluctuations in stochastic lattice Lotka–Volterra models. J Stat Phys 128:447–483. several movies with Monte Carlo simulation animations can be accessed at http://www.phys.vt.edu/~tauber/PredatorPrey/movies/
5. Washenberger MJ, Mobilia M, Täuber UC (2007) Influence of local carrying capacity restrictions on stochastic predator-prey models. J Phys: Condens Matter 19:065139, 1–14
6. Ramond P (1981) Field theory – a modern primer. Benjamin/Cummings, Reading
7. Amit DJ (1984) Field theory, the renormalization group, and critical phenomena. World Scientific, Singapore
8. Negele JW, Orland H (1988) Quantum many-particle systems. Addison-Wesley, Redwood City
9. Parisi G (1988) Statistical field theory. Addison-Wesley, Redwood City
10. Itzykson C, Drouffe JM (1989) Statistical field theory. Cambridge University Press, Cambridge
11. Le Bellac M (1991) Quantum and statistical field theory. Oxford University Press, Oxford
12. Zinn-Justin J (1993) Quantum field theory and critical phenomena. Clarendon Press, Oxford
13. Cardy J (1996) Scaling and renormalization in statistical physics. Cambridge University Press, Cambridge
14. Janssen HK (1976) On a Lagrangean for classical field dynamics and renormalization group calculations of dynamical critical properties. Z Phys B 23:377–380
15. De Dominicis C (1976) Techniques de renormalisation de la théorie des champs et dynamique des phénomènes critiques. J Physique (France) Colloq 37:C247–C253
16. Janssen HK (1979) Field-theoretic methods applied to critical dynamics. In: Enz CP (ed) Dynamical critical phenomena and related topics. Lecture Notes in Physics, vol 104. Springer, Heidelberg, pp 26–47
17. Doi M (1976) Second quantization representation for classical many-particle systems. J Phys A: Math Gen 9:1465–1477
18. Doi M (1976) Stochastic theory of diffusion-controlled reactions. J Phys A: Math Gen 9:1479–1495
19. Grassberger P, Scheunert M (1980) Fock-space methods for identical classical objects. Fortschr Phys 28:547–578
20. Peliti L (1985) Path integral approach to birth-death processes on a lattice. J Phys (Paris) 46:1469–1482
21. Peliti L (1986) Renormalisation of fluctuation effects in the $A + A \rightarrow A$ reaction. J Phys A: Math Gen 19:L365–L367
22. Lee BP (1994) Renormalization group calculation for the reaction $kA \rightarrow \emptyset$. J Phys A: Math Gen 27:2633–2652
23. Lee BP, Cardy J (1995) Renormalization group study of the $A + B \rightarrow \emptyset$ diffusion-limited reaction. J Stat Phys 80:971–1007
24. Mattis DC, Glasser ML (1998) The uses of quantum field theory in diffusion-limited reactions. Rev Mod Phys 70:979–1002
25. Täuber UC, Howard MJ, Vollmayr-Lee BP (2005) Applications of field-theoretic renormalization group methods to reaction-diffusion problems. J Phys A: Math Gen 38:R79–R131
26. Täuber UC (2007) Field theory approaches to nonequilibrium dynamics. In: Henkel M, Pleimling M, Sanctuary R (eds) Ageing and the glass transition. Lecture Notes in Physics, vol 716. Springer, Berlin, pp 295–348
27. Täuber UC, Critical dynamics: a field theory approach to equilibrium and nonequilibrium scaling behavior. To be published at Cambridge University Press, Cambridge. for completed chapters, see: http://www.phys.vt.edu/~tauber/utaeuber.html
28. Schütz GM (2000) Exactly solvable models for many-body systems far from equilibrium. In: Domb C, Lebowitz JL (eds) Phase transitions and critical phenomena, vol 19. Academic Press, London
29. Stinchcombe R (2001) Stochastic nonequilibrium systems. Adv Phys 50:431–496
30. Van Wijland F (2001) Field theory for reaction-diffusion processes with hard-core particles. Phys Rev E 63:022101, 1–4
31. Chopard B, Droz M (1998) Cellular automaton modeling of physical systems. Cambridge University Press, Cambridge
32. Marro L, Dickman R (1999) Nonequilibrium phase transitions in lattice models. Cambridge University Press, Cambridge
33. Hinrichsen H (2000) Nonequilibrium critical phenomena and phase transitions into absorbing states. Adv Phys 49:815–958
34. Ódor G (2004) Phase transition universality classes of classical, nonequilibrium systems. Rev Mod Phys 76:663–724
35. Moshe M (1978) Recent developments in Reggeon field theory. Phys Rep 37:255–345
36. Obukhov SP (1980) The problem of directed percolation. Physica A 101:145–155
37. Cardy JL, Sugar RL (1980) Directed percolation and Reggeon field theory. J Phys A: Math Gen 13:L423–L427
38. Janssen HK (1981) On the nonequilibrium phase transition in reaction-diffusion systems with an absorbing stationary state. Z Phys B 42:151–154
39. Janssen HK, Täuber UC (2005) The field theory approach to percolation processes. Ann Phys (NY) 315:147–192
40. Grassberger P (1982) On phase transitions in Schlögl's second model. Z Phys B 47:365–374

41. Janssen HK (2001) Directed percolation with colors and flavors. J Stat Phys 103:801–839
42. Martin PC, Siggia ED, Rose HA (1973) Statistical dynamics of classical systems. Phys Rev A 8:423–437
43. Bausch R, Janssen HK, Wagner H (1976) Renormalized field theory of critical dynamics. Z Phys B 24:113–127
44. Chaikin PM, Lubensky TC (1995) Principles of condensed matter physics. Cambridge University Press, Cambridge
45. Hohenberg PC, Halperin BI (1977) Theory of dynamic critical phenomena. Rev Mod Phys 49:435–479
46. Schmittmann B, Zia RKP (1995) Statistical mechanics of driven diffusive systems. In: Domb C, Lebowitz JL (eds) Phase transitions and critical phenomena, vol 17. Academic Press, London
47. Janssen HK, Schmittmann B (1986) Field theory of long time behaviour in driven diffusive systems. Z Phys B 63:517–520
48. Leung KT, Cardy JL (1986) Field theory of critical behavior in a driven diffusive system. J Stat Phys 44:567–588
49. Forster D, Nelson DR, Stephen MJ (1977) Large-distance and long-time properties of a randomly stirred fluid. Phys Rev A 16:732–749
50. Kardar M, Parisi G, Zhang YC (1986) Dynamic scaling of growing interfaces. Phys Rev Lett 56:889–892
51. Barabási AL, Stanley HE (1995) Fractal concepts in surface growth. Cambridge University Press, Cambridge
52. Halpin-Healy T, Zhang YC (1995) Kinetic roughening phenomena, stochastic growth, directed polymers and all that. Phys Rep 254:215–414
53. Krug J (1997) Origins of scale invariance in growth processes. Adv Phys 46:139–282

Finance, Agent Based Modeling in

SEBASTIANO MANZAN
Department of Economics and Finance, Baruch College
CUNY, New York, USA

Article Outline

Glossary

Rational expectations (RE) An assumption often introduced in economic models. It assumes that agents subjective distribution is equal to the true probability distribution of a random variable. The implication is that expectation errors are purely random.

Bounded rationality The assumption that agents have limited ability to acquire and process information and to solve complex economic problems. These limitations imply that expectations can diverge from RE.

Efficient markets hypothesis (EMH) An application of rational expectations to asset prices. The EMH assumes that asset prices reflect all available information. It implies that asset prices behave like a random walk process and their changes are purely random.

Artificial financial markets A market populated by agents that have bounded rational expectations and learning from available information. Trading in these markets occurs based on traditional price setting mechanisms or more realistic mechanisms inspired by electronic markets.

Definition of the Subject

Finance can be broadly defined as studying the allocation of resources over time in an uncertain environment. Consumers are interested in saving part of their current income and transfer it for consumption in the future (e. g., saving for retirement). On the other hand, firms are looking to raise capital to finance productive investments that will payoff in the future. In both decisions, the future is uncertain and individuals and firms are required to evaluate the risks involved in buying an asset (e. g., stocks and bonds) or investing in a project.

The traditional modeling approach in finance is to introduce strong assumptions on the behavior of agents. They are assumed to have perfect knowledge of the structure of the economy and to correctly process the available information. Based on these two assumptions, agents are able to form Rational Expectations (RE) such that their beliefs are not systematically wrong (in other words, the forecasting errors are random). Common sense suggests that these assumptions impose unreasonable requirements on the cognitive and computational abilities of agents. In practice, investors and firms are trying to learn to behave "*rationally*" in an economic system that is continuously evolving and where information is imperfect. In addition, there is an increasing amount of empirical evidence that is not consistent with RE theories.

These limitations have motivated an interest in finance to relax the strong assumptions on agents' behavior. Agent-based modeling contributes to this literature by assuming that consumers and firms have limited computational abilities (also known as bounded rationality) and learning (rather than knowing) the mechanisms governing the economy. These models have two main targets. First, to determine the conditions that lead a population of

bounded-rational interacting agents to produce an aggregate behavior that resembles the one of a RE representative agent model. Second, they aim at explaining the empirical facts and anomalies that the standard approach fails to explain.

This entry is structured as follows. In Sect. "Introduction" we discuss in more detail the application of agent-based modeling in finance. In particular, most of the early literature has focused on one specific aspect of financial economics, asset pricing. Sects. "The Standard RE Model" to "Computational Agent-Based Models" introduce the standard asset pricing model and describe the agent-based approaches that have been proposed in the literature. Sect. "Other Applications in Finance" presents some (more recent) applications of agent-based models to corporate finance and market microstructure and, finally, Sect. "Future Directions" discusses some possible future directions on the application of agent-based models in finance.

Introduction

The goal of asset pricing models is to provide an explanation for the "fair" valuation of a financial asset paying an uncertain cash flow. A key role in asset pricing models is played by agents expectations regarding the future cash flow of the asset. Standard economic models assume that agents have *Rational Expectations* (RE). The RE hypothesis is the outcome of some more basic assumptions on agents behavior: they know and use all the information available, they have unlimited computational ability, and rationality is common knowledge in the population. Common sense and introspection suggest that these are quite strong assumptions if the target is to build a realistic model of agents behavior. A justification for assuming RE in asset pricing models is provided by [37]:

> ... this hypothesis (like utility maximization) is not "behavioral": it does not describe the way agents think about their environment, how they learn, process information, and so forth. It is rather a property likely to be (approximately) possessed by the outcome of this unspecified process of learning and adapting.

Agent-based models try to address the issues left unspecified by the RE proponents: how do agents learn and process the information available? In other words, how do they form expectations? In fact, in the intent of the RE proponents, rationality is simply a property of the outcome (e. g., asset price) rather than an assumption about the subjective expectation formation process.

The innovative aspect of the agent-based approach is that it explicitly models "*this unspecified process of learning and adaptation*" (in Lucas's words). The common elements of the wide range of agent-based asset pricing models are:

Expectations agents hold subjective expectations that are *bounded rational*, that is, they are based on processing the available (and possibly imperfect and costly) information and that evolve over time. Agent-based models explicitly specify the way individuals form their expectation, instead of leaving it totally unspecified as in the RE approach.

Heterogeneity agents have different subjective expectations about the future due to heterogeneity in the way they process or interpret information. The RE setup suppresses agents heterogeneity: given the same information set, there is only one way to be rational and agents are thus homogeneous.

Evolution agents evolve in the sense that they abandon a belief if it performs poorly. Instead, rational models typically rely on the latent assumption that non-rational agents will not survive a (unspecified) process of evolutionary market competition.

Based on these basic ingredients, the agent based literature has now grown in different directions and we can distinguish two clearly defined approaches to agent-based modeling. The main difference between them is how they combine the different characteristics discussed above:

Analytical models these models assume that there are many expectation types and agents switch between them according to a deterministic or stochastic process. In the deterministic case, evolution is based on the past performance of the beliefs: agents discard belief types that perform badly compared to the other available. Instead, models with stochastic switching assume that a process governs the imitation and mutation of types, with possible additional features of herding. These models are simple and analytically tractable.

Computational models agents beliefs can change (or mutate) over time, due to the evolutionary selection of the best performing beliefs. Contrary to the analytical approach, the computational models renounce to analytical tractability in order to investigate more realistic expectation formation processes. Most of these models adopt fitness criteria (e. g., a Genetic Algorithm) to model the evolution of expectations.

The first aim of the agent-based literature is to understand whether introducing less restrictive assumptions on
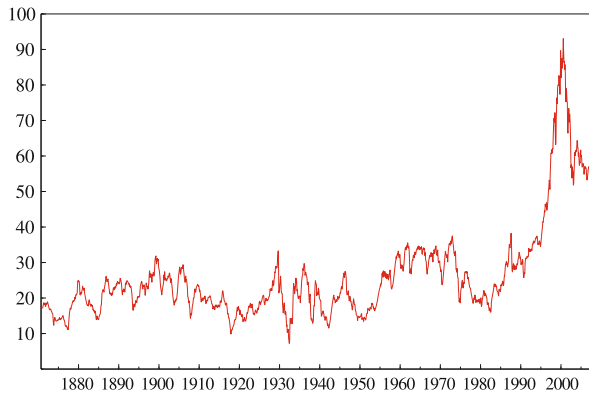
Finance, Agent Based Modeling in, Figure 1
Monthly Price-to-Dividend Ratio for the S&P500 Index from 1871 to 2006

agents behavior (bounded rationality, heterogeneity, and evolutionary selection of expectations) is consistent with the economy converging to the RE equilibrium. If this is the case, it can be argued that relaxing the homogeneity and rationality of agents represents a feasible way to describe the way individuals learn and adapt to achieve an outcome consistent with RE. The second aim of this literature is to provide an explanation for the empirical behavior of asset prices. To illustrate the main stylized facts of financial returns, we consider the Standard & Poors 500 (S&P500) Composite Index (a U.S. equity index). Figure 1 shows the Price-to-Dividend (PD) ratio from 1871 until 2006 at the monthly frequency. It is clear that the PD ratio fluctuates significantly with some periods of extreme valuations, as in the late 1990s. The debate on the reasons for these fluctuations has not reached (yet) a widely accepted conclusion. On the one hand, there are RE models that explain the variation in the PD ratio by changes in the risk premium, i. e., the ex-ante rate of return required by agents to invest in the risky asset. Instead, other models attribute these swings to irrational expectations of investors, that are prone to optimism (and overvaluation) when asset prices are increasing. The two explanations are not mutually excluding since both factors might contribute to explain the observed fluctuations of the PD ratio.

Figure 2 considers the S&P500 Index from 1977 until 2007 at the daily frequency. The figure shows the returns (defined as the percentage change of the price of a financial asset) and the absolute value of the returns. Figure 3 describes the statistical properties of returns, such as the histogram and the autocorrelation function of the returns and absolute returns. The main stylized facts of daily returns are:

Volatility clustering returns alternate periods of high and low volatility (or variability). In calm periods, returns oscillate within a small range, while in turbulent periods they display a much wider range of variation. This is a feature common to different asset classes (e. g., equities, exchange rates, and bonds). The time series of returns and absolute returns on the S&P500 in Fig. 2 clearly show this pattern. In certain periods returns vary in a narrow interval between $\pm 1\%$, while in other periods their variability is higher (e. g., between ± 3 and 5%).

Leptokurtic distribution the distribution of asset returns has a sharp peak around the mean and fat tails (compared to the normal distribution). Large events (positive and negative) are more likely to occur compared to what is expected under the assumption of normality. This property emerges clearly from the top plot of Fig. 3 that shows the histogram of the S&P500 returns and the normal distribution (based on the estimated mean and variance).

No serial correlation returns do not display significant linear serial correlation. The autocorrelation function of the returns (mid-plot of Fig. 3) is close to 0 at all lags considered.

Persistence in volatility on the other hand, volatility (measured by absolute or square returns) has significant linear dependence. The autocorrelation of the absolute returns in Fig. 3 is about 0.1 (and still significant) at lag 100.
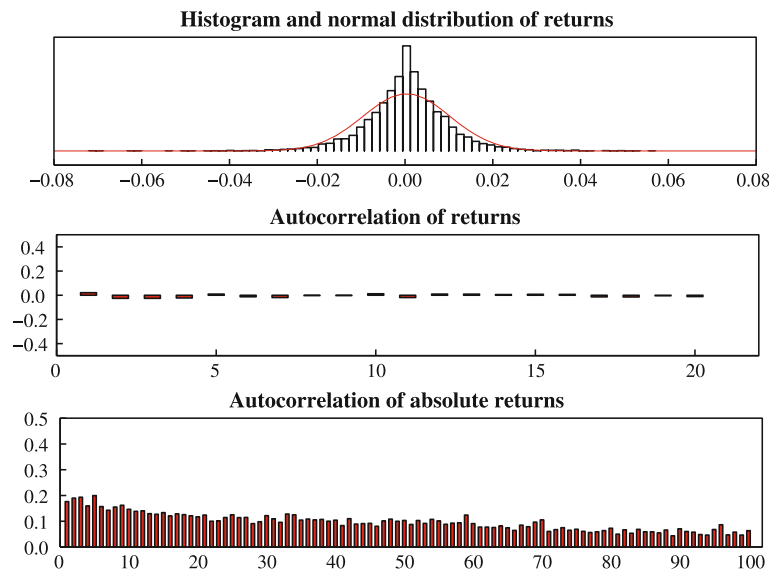
Another relevant fact that is short of explanations is the large trading volume that occurs in financial markets. A model in which everybody is rational and knows that everybody else is rational cannot account for the existence of such relevant volume of trade. Agent-based models aim at explaining this phenomenon based on the assumption that agents hold heterogeneous expectations. Volume can arise, for example, if an optimistic agent is willing to buy an asset from a pessimistic agent (that is willing to sell). An interesting feature of trading volume is its asymmetry during markets cycles: it is typically high when financial markets are booming, and low when the prices are decreasing. There is also empirical evidence that trading volume and volatility are correlated, suggesting that the same economic mechanism might be able to explain both phenomena.

Summarizing, the aim of the agent-based approach to asset pricing is to introduce more realistic assumptions on the way agents form expectations, learn from new information, and adapt to a changing environment. The re-

S&P500 Index

Returns of the S&P500 Index

Absolute returns of the S&P500 Index
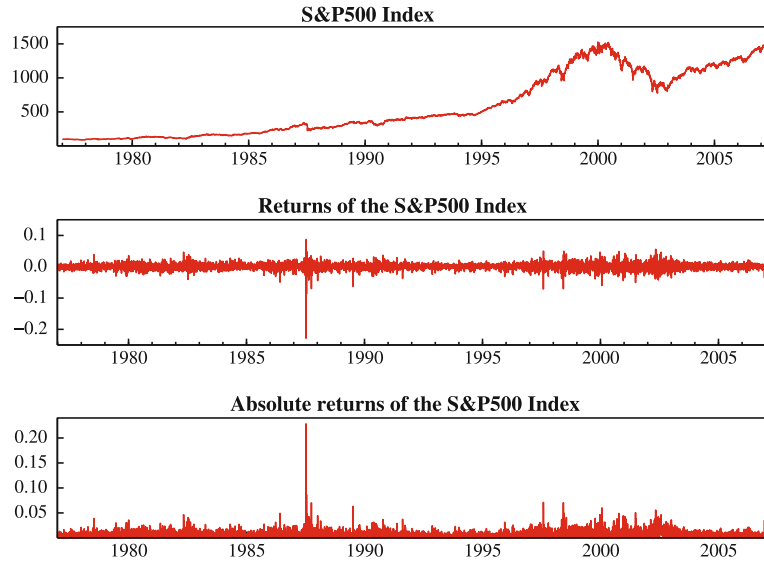
Finance, Agent Based Modeling in, Figure 2
Daily observations of the S&P500 Index from 1977 to 2007. (*top*) Time series of the Index, (*middle*) the returns, (*bottom*) the absolute value of returns

Histogram and normal distribution of returns

Autocorrelation of returns

Autocorrelation of absolute returns

Finance, Agent Based Modeling in, Figure 3
Statistical properties of the S&P500 returns: (*top*) histogram and normal distribution, (*middle*) autocorrelation function (max lag 20) for the returns, (*bottom*) autocorrelation function (max lag 100) for the absolute returns

search questions the agent-based approach is trying to answer are:

1. Under what conditions are these models able to reproduce the RE equilibrium (although starting from a more general setup where agents are not – a priori – assumed to have RE)?

2. Another issue is the empirical validity of these models: are they able to explain the empirical features of financial returns that standard RE models fail to account for?

In the following Sections, we describe some of the most well-known examples of agent-based models in finance, both in the *analytical* and *computational* group. However,

we first introduce a basic RE model that is the starting point for most of the literature. In Sect. "Other Applications in Finance" we discuss other interesting applications of agent-based models in finance.

The Standard RE Model

We briefly consider the standard asset pricing model that is used as a benchmark in the agent-based literature. A more detailed discussion can be found in [25]. The model assumes that agent i faces the choice of investing her wealth among two assets: a riskless asset that pays a constant return r, and a risky asset that pays a stochastic dividend in period t denoted by D_t. A typical assumption is that agents have Constant Absolute Risk Aversion (CARA) preferences defined as $U(W_i) = -e^{-\lambda W_i}$, where $U(\cdot)$ indicates the utility function, W_i denotes the wealth of agent i and λ is the coefficient of absolute risk aversion. These preferences imply the following demand of shares of the risky asset, $X_{i,t}$:

$$X_{i,t} = \frac{E_{i,t}(P_{t+1} + D_{t+1}) - (1+r)P_t}{\lambda \sigma_{i,t}^2 (P_{t+1} + D_{t+1})}, \qquad (1)$$

where P_t is the price of the risky asset in period t, $E_{i,t}(\cdot)$ is the conditional expectation of agent i about next-period payoff of the risky investment, and $\sigma_{i,t}^2(\cdot)$ is the conditional variance of the payoff for agent i. Agents buy shares of the risky asset ($X_{i,t} > 0$) if they expect the return of a share to be higher compared to investing the same amount (P_t) in the riskless asset.

The equilibrium price of the risky asset is such that the aggregate demand and supply are equal. Assuming that there are S number of shares of the risky asset available, the equilibrium condition is

$$S = \sum_i X_{i,t}. \qquad (2)$$

The aggregation across different individuals is simplified by assuming a representative agent with expectation $E_t(P_{t+1} + D_{t+1})$ (and similarly for the conditional variance) for all i's in the economy. This is equivalent to assume that agents are homogeneous in their expectation about the future payoff of the risky asset. In addition, assuming that the representative agent holds RE, it can be shown that the equilibrium price of the risky asset is a linear function of D_t given by

$$P_t = a + bD_t,$$

where a and b are constant (and related to the structural parameters of the model).

There is an extensive literature that aims at relaxing the strong restrictions imposed by the RE hypothesis. Models of *rational learning* assume that agents (typically in a representative agent setup) have to learn (rather than know) the structure of the economy, e. g., the parameters governing the cash flow process. In this case, agents are rational in the sense that they process optimally the information available. However, they do not hold rational expectations since they have imperfect knowledge of the structure of the economy. An alternative route followed in the literature is to consider the effect of *behavioral* biases in the expectation formation process. A comparison of the vast literature on rational learning and behavioral models is provided by [6].

Agent-based models build on these extensions of the basic asset pricing model by considering both rational learning and irrational expectations in a richer economic structure where agents hold heterogeneous expectations. We will now discuss some of the most well-known models in the analytical and computational agent-based literature and deal with their main differences.

Analytical Agent-Based Models

The analytical models assume that the population of agents can choose among a small number of beliefs (or predictors) about next period payoff of the risky asset. Heterogeneity is introduced by allowing agents with different predictors to co-exist, and learning might occur if they are allowed to switch between different beliefs in an evolutionary way.

These models can be described as follows. Assume there are a set of H belief types publicly available to agents. Denote the belief of type h (for $h = 1, \dots, H$) about next period payoff by $E_{h,t}(P_{t+1} + D_{t+1})$ and the conditional variance by $\sigma_{h,t}^2(P_{t+1} + D_{t+1})$. Since these models depart from the assumption of RE, they typically introduce a behavioral assumption that the beliefs are either of the *fundamentalist* or the *trend-following* type. [20] and [21] conducted survey studies of exchange rate traders and found that their expectations could be represented as trend-following in the short-run, but fundamentalist in the long run. Fundamentalist expectations are characterized by the belief that the market price is anchored to the asset fundamental valuation and deviations (of the price from the fundamental) are expected to disappear over time. In this case, the belief uses both information about the asset price and the dividend process (that drives the fundamental value) to form an expectation about the future. On the other hand, trend-following expectations exploit only information contained in the price series to extrapolate the future

dynamics. These types of beliefs are obviously not consistent with the RE equilibrium although they are supported by empirical evidence of their widespread use in financial markets.

Another key assumption of agent-based models concerns the evolution of beliefs: agents switch between expectations based on their past performance or because of interaction with other agents in the population. It is possible to distinguish two families of models with different evolutionary dynamics:

Deterministic evolution agents switch between the different beliefs based on a deterministic function. Typically, the switching is determined by past forecast accuracy of the predictors or their realized profits.
Stochastic evolution a stochastic process governs the switching of agents between beliefs.

Deterministic Evolution

An example of an agent-based model with deterministic evolution is proposed by [7]. A simple version of their model assumes that there are only two types of beliefs: fundamentalists and trend-followers. Some simplifying assumptions are used in deriving the equilibrium price: the dividend process D_t is assumed to be $i.i.d$ (with mean $\bar{D}$) and agents have homogeneous expectations about the dividend process. In this case, the expectation about next period payoff $E_{h,t}(P_{t+1} + D_{t+1})$ in Eq. (1) becomes $E_{h,t}(P_{t+1}) + \bar{D}$.

Lets denote by $P^*(= \bar{D}/r)$ the constant RE fundamental price. The fundamentalist type has the following belief:

$$E_{F,t}(P_{t+1}) = P^* + g_F(P_{t-1} - P^*) . \qquad (3)$$

When $0 < g_F < 1$, fundamentalists believe the asset price will revert toward its fundamental value, and g_F can be interpreted as the speed at which this adjustment is expected to occur. This model assumes that when agents form their belief at time t they actually do not observe the realized asset price for period t. This explains the fact that the expectation is a function of the last observed price, P_{t-1}.

Brock and Hommes assume that agents pay a cost C to acquire the fundamentalist predictor. The underlying idea is to let them choose whether to buy a "sophisticated" predictor (that requires calculating the fundamental value) or, alternatively, to extrapolate from past realized prices. The belief of the trend-followers is given by:

$$E_{TF,t}(P_{t+1}) = g_{TF}P_{t-1} . \qquad (4)$$

The value of the parameter g_{TF} determines the strength of extrapolation of the trend-followers. If $g_{TF} > 1$, they expect an upward trend in prices and, vice versa, for $0 < g_{TF} < 1$.

Assuming the supply of the risky asset, S, in Eq. (2) is equal to 0, the equilibrium asset price, P_t, is given by:

$$P_t = \left(\frac{n_{F,t}(1 - g_F) - r}{1 + r} \right) P^*$$
$$+ \left(\frac{n_{F,t}(g_F - g_{TF}) + g_{TF}}{1 + r} \right) P_{t-1} , \qquad (5)$$

where $n_{F,t}$ indicates the fraction of agents in the population using the fundamentalist belief and the remaining $n_{TF,t}(= 1 - n_{F,t})$ using the trend-following one. [7] assumes the evolution of the fractions $n_{F,t}$ is governed by a discrete choice probability model:

$$n_{F,t} = \frac{1}{1 - \exp\left[\beta(U_{TF,t-1} - U_{F,t-1})\right]} , \qquad (6)$$

where $U_{h,t-1}(h = F, TF)$ is a measure of the fitness of belief h defined as:

$$U_{F,t-1} = \pi_{F,t-1} + \eta U_{F,t-2} - C, \quad \text{and}$$
$$U_{TF,t-1} = \pi_{TF,t-1} + \eta U_{TF,t-2} ,$$

where $\pi_{h,t-1}$ measures the fitness performance (measured by realized profits or forecast accuracy) of the belief h at time $t - 1$ and η is a parameter that determines the memory in the performance measure. C in $U_{F,t-1}$ represents the cost that agents face if they adopt the fundamentalist belief (while the trend-following is available at no cost). The fraction in Eq. (6) depends on the parameter $\beta(> 0)$ that determines the speed at which agents respond to differentials of performance among beliefs. If β is small, agents are very reluctant to switch and require a significantly large difference in fitness to adopt another predictor. On the other hand, when β is large, even small differences of performance cause dramatic changes in the fractions. For a given value of β, if the fundamentalist belief significantly outperforms the trend-following (that is, $U_{F,t-1} \gg U_{TF,t-1}$), then the fraction $n_{F,t-1}$ tends to 1, meaning that most agents in the economy switch to the fundamentalist expectation.

The interesting feature of this model is that it can converge to the RE equilibrium or generate complicated dynamics depending on the value of the parameters. For some combinations of the g_F and g_{TF} parameters, the system converges to the RE equilibrium (i. e., the deviation is equal to 0). However, trend-followers can destabilize the economy when their extrapolation rate, g_{TF} is high enough. For small values of β the dynamical system converges to the RE equilibrium. However, for increasing values of β the system experiences a transition toward a non-

fundamental steady state and complicated dynamics (limit cycles and strange attractors) emerge.

In the presence of information cost (to buy the fundamentalist predictor) and evolutionary switching between strategies, the economy might still converge to the RE equilibrium for a large set of parameter values. However, it is also able to generate large fluctuations of the asset price around the fundamental value. Figure 4 shows a time series of the asset price P_t and the fraction of fundamentalists described in Eqs. (5) and (6). The constant fundamental value in this Figure is equal to 25. As it is clear from the picture, the asset price experiences large swings away from the fundamentals that are explained by the increased importance of agents using the trend-following belief. When the mispricing becomes too large, the economy experiences a sudden change of sentiment with most agents switching to the fundamentalist belief. In this sense, the model is more appropriate to explain the boom-bust dynamics of financial markets.

Although the purely deterministic model captures the relevant features of the dynamics of financial markets, adding a stochastic component provides simulated series that better resemble the observed ones (such as Fig. 1). The model can be extended by considering an approximation error term in Eq. (5) that interacts with the dynamics of the model. Figure 5 shows the asset price and the fraction of fundamentalists for a normally distributed error term with standard deviation equal to 2. The asset price

shows large and persistent deviations from the fundamental value ($P^* = 25$), in some periods as extreme as reaching 100 while, in other periods, more moderate. Since the dividend process is assumed to be $i.i.d.$, the price can also be interpreted as a valuation (PD) ratio. Comparing the properties of this time series with the one for the S&P500 in Fig. 1, it seems that it is able to capture its main qualitative features. [5] provide empirical evidence of the ability of a similar model to explain the long-run behavior of stock prices.

The model proposed by Brock and Hommes is an example of a literature interested in the (possible) emergence of complicated dynamics in asset pricing models. An early contribution to the deterministic literature is [15]. They assume that there are two types of investors: some extrapolating from past deviations while the other group is more sophisticated and able to evaluate whether the asset is over- or under-valued (and sells or buys more aggressively if the mispricing is large). A third actor in the model is the market maker. The role of the market maker is to aggregate demand and supply and to fix the price of the asset. This mechanism is different from the assumption in Eq. (2). In that case, agents submit their demand function (quantity as a function of price) and the price is set at the value that clears the market. Instead, the market maker receives orders from the agents and moves the price to offset excess demand or supply. This represents a disequilibrium mechanism since market makers use their in-
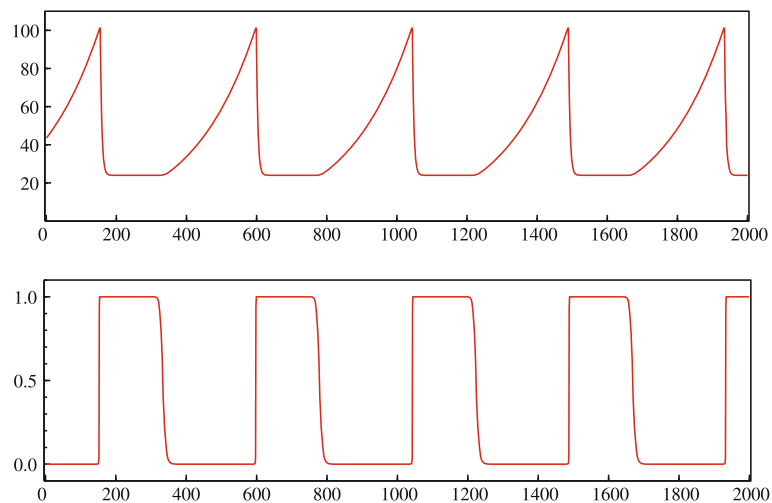
Finance, Agent Based Modeling in, Figure 4
Brock and Hommes model with 2 belief types, fundamentalists and trend-followers. The top plot represents a time series of the asset price and the bottom plot depicts the fraction of fundamentalists, $n_{F,t}$. The parameters of the model: intensity of choice $\beta = 0.5$, the interest rate $r = 0.0083$, the parameter of the fundamentalists $g_F = 0.8$, the parameter of the trend-followers belief $g = 1.014$, the cost of the fundamentalist predictor $C = 0.5$, memory parameter $\eta = 0.99$
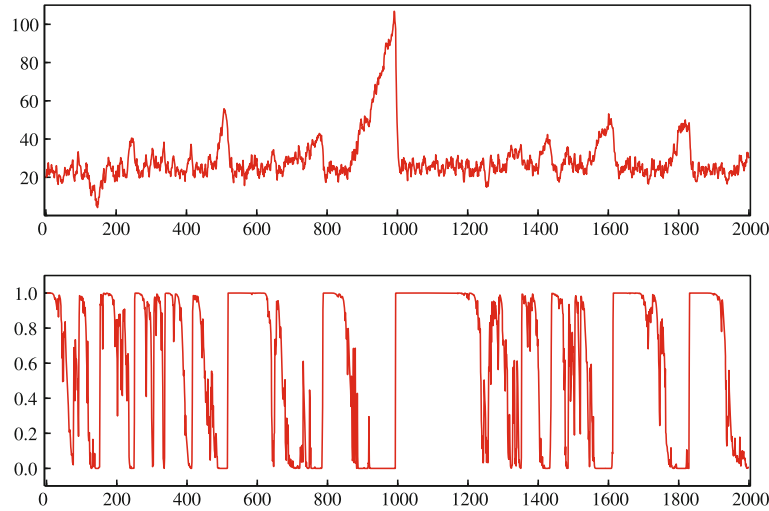
Finance, Agent Based Modeling in, Figure 5
Same model and parameter values used in Fig. 4 with an error term added to Eq. (5) that is normally distributed with mean zero and standard deviation equal to 2

ventory of stocks to provide liquidity in case of excess demand and accumulate stocks in case of excess supply. The results for this model are similar to what was discussed above. The RE equilibrium is obtained when the sophisticated agents dominate the market. However, limit cycles and chaos arise when the trend-following agents are relatively important and the economy fluctuates between periods of increasing asset prices and sudden crashes. Another model that assumes the market maker mechanism is proposed by [9]. In this case, agents form their expectations based on either the fundamental or extrapolative approach. However, the excess demand function of the chartist is assumed to be nonlinear. When the extrapolation rate of the chartist is sufficiently high, the system becomes unstable and limit cycles arise. While these early models assumed that the fractions of agents are fixed, [16] and [17] introduced, in a similar setup, time-variation in those fractions. The driving force for the variation of the fractions is the relative performance of the beliefs (similar to what we discussed above for the model of Brock and Hommes). Some of the more recent models that extend these early contributions are [10,11,12,18,19,24,50], and [51]. A comprehensive survey of the literature is provided in [26].

Stochastic Evolution

An early example of an agent-based model in which individuals switch in a stochastic fashion was proposed by [27]. He uses a slightly different setup compared to the Standard RE Model. In his model the asset is a foreign bond and the agent has to decide whether to invest at home (at the riskless interest rate r) or buy a unit of foreign currency and invest abroad (at the risky interest rate ρ_t, assumed to be normally distributed with mean ρ and variance σ_ρ^2). The price P_{t+1} represents the exchange rate. The only difference with the model described earlier is that in the demand of type h agent in Eq. (1), $E_{h,t}(P_{t+1} + D_{t+1})$ is replaced by $(1 + \rho)E_{h,t}(P_{t+1})$. The fundamental value of the asset in this model is assumed to evolve as a random walk, that is, $P_t^* = P_{t-1}^* + \epsilon_t$ where $\epsilon_t \sim N(0, \sigma_\epsilon^2)$.

Similarly to the previous model, there are two types of beliefs: fundamentalists and chartists. The fundamentalist belief is the same as in Eq. (3), while the chartists have belief given by:

$$E_{\text{TF},t}(P_{t+1}) = (1 - g_{\text{TF}})P_t + g_{\text{TF}}P_{t-1} .$$

The switching between beliefs in Kirman's model is driven by two mechanism: social interactions and herding. Interaction means that agents meet in pairs and communicate about their beliefs. The result of this communication is that, with a probability $(1 - \delta)$, an agent changes her belief to the one of the other agent. In this model, market information (such as prices or dividends) do not play any role in the decision of the agents to adopt the fundamentalist or trend-following beliefs. This is in sharp contrast to the model of [7] where the selection of the belief is endogenous and based on their past performance. In addition to the probability of switching belief due to social interaction, there is a probability ϵ that an agent indepen-

dently changes belief. If we denote by $N_{F,t}$ the number of agents in the population (N is the total number of agents) using the fundamentalist belief at time t, Kirman models the evolution from $N_{F,t-1}$ to $N_{F,t}$ according to a markov chain with the following transition probabilities:

$$P(N_{F,t} - N_{F,t-1} = 1) = \left(1 - \frac{N_{F,t-1}}{N}\right)$$
$$\left(\epsilon + (1-\delta)\frac{N_{F,t-1}}{N-1}\right)$$
$$P(N_{F,t} - N_{F,t-1} = -1) = \frac{N_{F,t-1}}{N}$$
$$\left(\epsilon + (1-\delta)\frac{N - N_{F,t-1}}{N-1}\right)$$
$$P(N_{F,t} - N_{F,t-1} = 0) = 1 - P(N_{F,t} - N_{F,t-1} = 1)$$
$$- P(N_{F,t} - N_{F,t-1} = -1).$$

The second part of the opinion formation can be characterized as herding. Kirman assumes that the agents receive a noisy signal, $q_{i,t}$, about the fraction of the population that is fundamentalist:

$$q_{i,t} = \frac{N_{F,t}}{N} + \xi_{i,t},$$

where $\xi_{i,t}$ is distributed as $N(0, \sigma_\xi^2)$ and $i = 1, \ldots, N$. Based on this signal about the average opinion in the economy, agents herd by coordinating in adopting the belief that is more popular. Agent i uses the fundamentalist belief if her signal, $q_{i,t}$, is larger than 0.5 and the trend-following belief otherwise. The fraction of agents using the fundamentalist belief (denoted by $n_{F,t}$) is then given by:

$$n_{F,t} = \frac{1}{N}\sum_{i=1}^{N} I(q_{i,t} \geq 0.5).$$

Given the beliefs of the two types, the fractions and assuming that the supply of foreign currency is proportional to the time varying fundamental value, the equilibrium price of the model is given by:

$$P_t = \frac{n_{F,t} - \gamma}{A}P_t^* - \frac{n_{F,t}g_F}{A}P_{t-1}^* + \frac{(1 - n_{F,t})g_{TF}}{A}P_{t-1}, \quad (7)$$

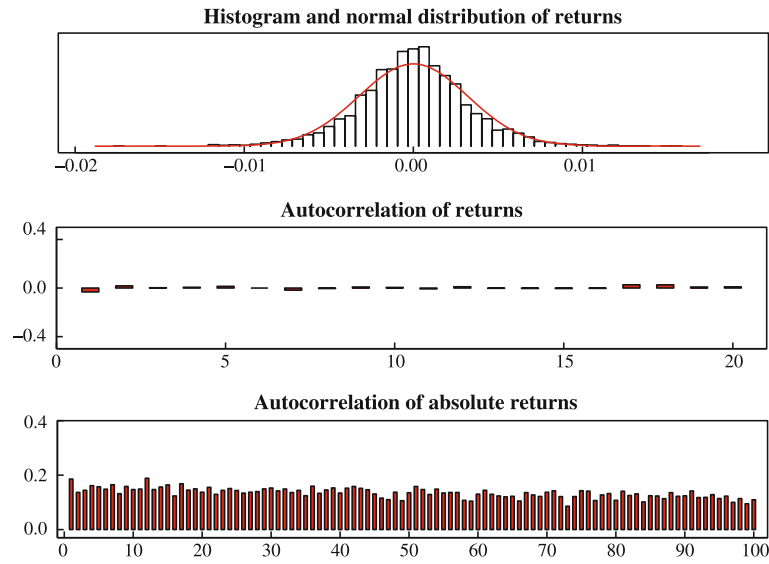where the constants γ and A are functions of the structural parameters of the model.

Figure 6 shows a time series simulated from Kirman's model. The fraction of agents using the fundamentalist belief, $n_{F,t}$, displays significant variation over time, with some periods being close to 1 (most agents are fundamentalists) and other periods close to zero (trend-followers dominate). The resulting price dynamics can be characterized as follows. When fundamentalists dominate the market, the asset price tracks closely the fundamental value and returns volatility is high. On the other hand, when
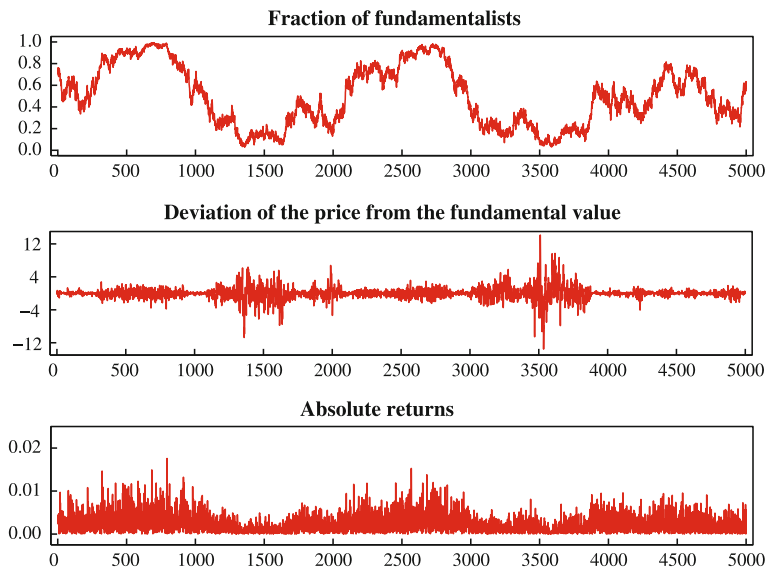
Finance, Agent Based Modeling in, Figure 6
Time series generated from the model proposed by [27] for the following parameter values: $N = 1000$, variance of the $\sigma_\epsilon^2 = 10$, $\rho = 0.00018538$, $r = 0.000133668$, $g_F = 0.6$, $g_{TF} = 0.475$, $\delta = 0.10$, $\epsilon = 0.000325$, and $\sigma_\xi^2 = 0.43/N$. The top plot shows the fraction of fundamentalist agents in the population, the middle plot the deviation of the market price from the fundamental value, and the bottom plot displays the absolute value of the asset returns

Histogram and normal distribution of returns

Autocorrelation of returns

Autocorrelation of absolute returns

Finance, Agent Based Modeling in, Figure 7
Statistical properties of the time series in Fig. 6. The top plot shows histogram of the series and the parametric distribution under the assumption of normality, and the bottom plots show the autocorrelation of the returns and absolute returns, respectively

trend-followers dominate the market the price tends to deviate significantly from the fundamental and volatility is lower. The time series provide a clear intuition about the ability of the model to account for periods of large deviation from the fundamentals and of persistent volatility.

A main objective of this model is to provide an explanation for the stylized facts of financial returns that were discussed earlier. Figure 7 shows some of the statistical properties for the simulated series. The histogram shows the typical leptokurtic property of financial returns. The distribution of the simulated returns shows a higher concentration of probability mass around zero and in the tails (compared to the normal distribution). The returns autocorrelations are very close to zero and statistically insignificant. However, the absolute returns show significantly positive and slowly-decaying autocorrelations. Hence, the simulated series from Kirman's model are able to replicate the relevant empirical features of daily financial returns.

[41] and [42] propose a model inspired by the opinion formation mechanism of Kirman. The model assumes that agents are either fundamentalists or chartists. In addition, the chartist group is composed of agents that are either optimistic or pessimistic. Agents can switch between the two sub-groups due to herding (following the majority opinion) and also to incorporate the recent trend in asset prices. Instead, the switching between fundamentalist and chartist beliefs is based on the excess profits of the rules. In this aspect, the model allows for a feedback ef-

fect from market price to the fractions similarly to [7]. A market maker mechanism aggregates the demand for the risky asset of the agents and determines the prices. Another model based on interacting agents and herding behavior is proposed by [14]. They model the communication among (groups of) agents as a random graph and the returns dynamics closely match the stylized facts of financial returns. [26] and [49] provide extensive overviews of models based on random communication in a population of interacting agents.

Computational Agent-Based Models

Computational agent-based models move one step further compared to analytical models. The setup is very similar: the simple asset pricing model described above, the assumption of heterogeneity of beliefs in the population, and evolutionary pressure to use the best performing predictors. However, computational models do not assume a priori the form of agents' beliefs. Instead, they let agents learn, adapt and explore a large set of strategies and use the best performing ones. In this sense these models allow to investigate the *emergence* of trading strategies and their survival in the market. A relevant question, and an unsettled dispute between academics and practitioners, is the role and importance of technical analysis. Computational agent-based models do not assume (a priori) that trend-following rules are used by agents (as in the analyti-

cal approach), but allow for these rules to emerge from the evolutionary and learning processes. Hence, they can indicate the conditions that lead to the emerge and survival of trend-following rules in the market.

One of the first and most famous example of a computational agent-based model is the Santa Fe Institute (SFI) artificial market proposed by [3]. As mentioned above, the key feature of this and similar models is the way the expectation formation process is represented.

Each agent in the economy is endowed with a set of predictors, in the form of *condition/forecast* rules. These rules are a form of classifier system that identify a state of the world and indicate an action (in this case a forecast of future returns). Each agent in the economy is assumed to rely on a set J of market predictors (classifier rules) that consist of two elements:

Condition a set of bit-strings that characterize different possible states of the market. Each bit represents a state of the world, and the design of the SFI market allows for a set of bits related to *fundamentals* (that relate the asset price to the underlying dividend process) and another set of *technical* bits (that relate the current price to a moving-average of past prices of different length). The role of the bit-strings is to provide the agent with the ability to identify the current state of the market.

Forecast associated with each bit-string j (for $j = 1, \ldots, J$) is a parameter vector (a_j, b_j) that together with the linear forecasting rule $E_{i,t}^{j}(P_{t+1}+D_{t+1}) = a_j(P_t+D_t)+b_j$ provides agent i with the forecast for next period payoff. The agent then combines the forecast of the H most accurate predictors that are active, based on the observed market condition.

The next building block of the SFI artificial market is the learning process. This is implemented using a Genetic Algorithm (GA) that enables learning in both the condition and the forecast part of the classifier. Agents have a probability p to update their predictors through learning in every period. The frequency of learning (measured by p) plays a relevant role in the resulting dynamics of the model since it identifies how quickly agents adapt to changes in the environment and respond to it. In the learning phase, 15% of the worst performing predictors are dropped, and new rules are generated using a GA with uniform crossover and mutation.

The aim of the model is to test the hypotheses discussed above:

1. Does the SFI market converge to the RE equilibrium?
2. Can the SFI market explain the stylized facts of financial returns?

It turns out that the answer is positive to both questions, depending on the speed of learning parameter p. This parameter plays a crucial role in the resulting dynamics of the SFI market and two regimes can be identified:

Slow-learning in this case the agents are engaged in learning (via the GA) every 1000 periods (on average). The resulting price dynamics shows convergence to the RE equilibrium characterized by agents having homogeneous expectations, negligible trading volume (although some occurs when agents change their beliefs due to learning), and returns that are normal and homoskedastic. What is remarkable is that the convergence to the RE equilibrium is not built-in the model, but it is achieved by the learning and evolutionary process taking place in the SFI market. Another interesting result is that the technical trading bits of the classifier play no role and are never active.

Fast-learning in this experiment the agents update their predictors via learning (on average) every 250 periods. The price dynamics shows the typical features of financial time series, such as alternating periods of high and low volatility, fat tailed distribution, high trading volume, and bubbles and crashes. An interesting result of the fast-learning experiment is the *emergence* of the trend-following rules. The technical trading bits of the predictors are activated and their effect on the asset price spurs even more agents to activate them. In this sense, trend-following beliefs emerge endogenously in the economy and they are not eliminated in the evolution of the system, but survive. This is a quite relevant result also from an empirical point of view. As we mentioned earlier, technical analysis is widely used by practitioners and the SFI market provides an explanation for its emergence (and survival).

[29,30], and [31] have recently proposed an artificial market that is largely inspired by the SFI market. However, LeBaron changed some very relevant assumptions compared to the earlier model. An innovation in this new artificial market is the assumption on the preferences of agents. While the SFI market (and many analytical models) rely on CARA preferences, [29] considers Constant Relative Risk Aversion (CRRA) preferences. In this case, wealth plays a role in the demand of agents (while with CARA does not) and, consequently, allows for differential market impact of agents based on their wealth. Another innovation concerns the expectation formation process. [29] departs from the SFI market assumption of different "*speed of learning*" across agents. Instead, LeBaron assumes that agents have different memory length in eval-

uating strategies. In every period agents assess the profitability of the strategy adopted. However, agents evaluate their strategy using different backtesting periods: some agents test their strategies on only the last few months, while other agents consider the last 20 years. In this sense, they are heterogeneous in their *memory* rather than in the speed of learning. Another feature of this model is that the classifier predictor is replaced by a neural network. The learning and evolution is always characterized by a GA. Despite these innovations, the earlier results of the SFI market are confirmed: a market populated by long-memory agents converges to the RE equilibrium. However, in an economy with agents holding different memory lengths, the asset price series shows the typical features of financial returns (no serial correlation, volatility clustering, fat-tailed distribution, high trading volume, and correlation between volume and volatility).

Another recent artificial stock market model is proposed by [8]. The setup is the simple asset pricing model described in Sect. "The Standard RE Model". Chen and Yeh assume that the expectation of agent i about next period payoff is of the form $E_{i,t}(P_{t+1}+D_{t+1}) = (P_t+\mu)(1+\theta_1\tanh(\theta_2 f_{i,t}))$. The quantity $f_{i,t}$ characterizes the expectation of the agent and it evolves according to genetic programming. If $f_{i,t}$ is equal to zero the agent believes in the efficiency and rationality of the market, that is, expects the asset price tomorrow to increase by the expected dividend growth rate.

Compared to the SFI market, they adopt a Genetic Programming (GP) approach to model agents' learning and evolution. The model assumes that agents evolve due to two types of pressures: peer-pressure (the agent performance compared to the rest of the population) and self-pressure (own evaluation of the performance). The probability that an agent searches for better forecasting rules depends on both forms of pressure. If agents rank low in terms of performance compared to their peers, then the probability that they will search for other forecasting rules is higher. The population of forecasting rules evolves due to competition with new rules that are generated by applying genetic operators (reproduction, cross-over, mutation) to the existing rules. The rules space evolves independently of the rules adopted by the agents. When an agent decides to search (due to the pressures mentioned above), forecasting rules are randomly selected from the population until a rule is found that outperforms the one currently adopted by the agent. Chen and Yeh show that the price dynamics of the model is consistent with an efficient market. The investigation of the statistical properties of the returns generated by the model shows that the series does not have any linear and nonlinear dependence, although there is some evidence for volatility clustering. Analyzing the type of rules that agents use, they show that only a small fraction of them are actually using forecasting rules that are consistent with an efficient market (in the sense that they believe that $E_{i,t}(P_{t+1}+D_{t+1}) = P_t + \mu$ in which case $f_{i,t}$ is equal to 0). In other words, although a large majority of agents uses rules that imply some form of predictability in asset returns, the aggregation of their beliefs delivers an asset price that looks "*unpredictable*". In this sense they claim that the efficiency (or unpredictability) of the artificial market is an emerging property that results from the aggregation of heterogeneous beliefs in the economy. Another property that emerges from the analysis of this market is the rationality of a representative agent. Chen and Yeh consider the expectation of a representative agent by averaging the expectations across the agents in the economy. The forecasting errors of this "*representative*" expectation indicate that they satisfy a test for rationality: there is no evidence of systematic forecasting errors in the expectation (in statistical terms, the errors are independent).

Evolution and learning (via GA) have received quite a lot of attention in the literature. Other artificial-market models have been proposed in the literature. Some early examples are [4], and [46]. Some more recent examples are [1,2,47,48]. [32] is an extensive survey of the computational agent-based modeling approach.

Other Applications in Finance

The common theme across the models presented above is to show that departing from a representative rational agent is a viable way to explain the empirical behavior of asset prices. The more recent agent-based literature has shifted interest toward nesting this type of models in more realistic market structures. There are two typical assumptions used in the agent-based literature to determine the asset price: a market clearing or a market maker mechanism. Recent progress in the analysis of the micro-structure of financial markets has indicated the increasing importance of alternative trading mechanisms, such as order-driven markets. In these markets, traders decide whether to buy (sell) using a market or limit order. A market order means that the trader is ready to buy (sell) a certain quantity of stocks at the best available price; instead, with limit orders traders fix both a quantity of shares and a price at which they are willing to buy (sell). Limit orders are stored in the book until a matching order arrives to the market. They are different from quote-driven markets, where a market maker provides quotes and liquidity to investors. This has spurred a series of articles that propose agent-

based models in this more realistic market structure. In particular, [13,36,44], and [45] consider an order-driven market where agents submit limit orders. Typically these models make simple behavioral assumptions on the belief formation process and do not consider learning and evolution of agents' expectations (typical of the computational agent-based models). In this sense, these models are closer to the stochastic agent-based approach reviewed in Sect. "Analytical Agent-Based Models". Recently, [33] has proposed a computational model for an order-driven market in which strategies evolve by imitation of the most successful rules.

[13] propose an order-driven market model in which the demand for the risky asset of the agents is determined by a fundamentalist, a chartist, and a noise component. The agents share the same demand function but the weights on the components are different across agents. Simulating the model suggests that the stylized facts of financial returns can be explained when all behavioral components (fundamentalist, chartist, and noise) participate to determine agents' beliefs. An additional feature of this setup is that it accounts for the persistence in the volatility of returns and trading volume. Such a micro-structural model allows also to investigate the effect of some key market design parameters (such as tick size, liquidity, and average life of an order) on the price formation process.

[44] consider a market structure where agents submit limit orders and the price is determined by market clearing of supply (sell orders) and demand (buy orders) schedules. The behavioral assumptions are closely related to the clustering approach of [14]: a probabilistic mechanism governs the formation of clusters and, within a clusters, all agents coordinate in buying or selling the risky asset. Another behavioral assumption introduced in this model concerns the (positive) relation between market volatility and the limit order price. When the volatility is low, agents set the price of their limit order close to yesterday's asset price. However, when the market is experiencing wide swings in prices, agents' set limit prices that are significantly above or below yesterday's price for their orders. The results suggest that the model is able to explain the main stylized facts of financial returns. [45] consider an economy with a similar market structure but more sophisticated assumption on agents' behavior. They assume the population is composed of four types of agents: random traders (with 50% probability to buy or sell), momentum (buy/sell following an increase/decrease in prices), contrarian (act in the opposite direction of momentum traders), and fundamentalists (buy/sell if the price is below/above the fundamental value). They simulate the model in a way that non-random agents do not

affect the asset price. The idea is to investigate the survival of these types of agents in the economy without affecting the aggregate properties of the model. They show that, on average, the fraction of wealth of momentum agents decreases while it increases for fundamentalist and contrarian traders.

Another recent paper that models an order-driven market is [36]. Agents can submit market and limit orders. They introduce the behavioral assumption that the agents are all fundamentalists, although they are heterogeneous in their belief of the fundamental value of the asset. They show that simulated series from this simple model follow a leptokurtic distribution and attribute this property to the structure of the market (rather than the behavioral assumptions). The same result is also obtained when random traders are considered. However, they are not able to explain other stylized fact such as the autocorrelation structure of volatility. This paper is interesting because it suggest that some of the stylized facts discussed earlier might not be related to agents' bounded rationality, but rather to the details of the market mechanism that is typically neglected in the agent-based literature.

Another area of application of agent-based modeling is corporate finance. [43] propose an agent-based model to investigate the security issuance preferences of firms and investors. The empirical evidence indicates that there is a clear dominance of debt financing, compared to other instruments, such as equities and convertible debt. This is a puzzle for theoretical models where it is customarily assumed that the payoff structure of the financing instruments are common knowledge. Under this assumption, the price should reflect the different characteristics of the securities and investors should be indifferent among them. Noe et al. consider a model that departs from the assumption of perfect knowledge about the security characteristics, and assume that firms and investors are learning (via a GA) about the profitability of the different alternatives. Debt and equity imply different degrees of risk-sharing between investors and firms: in particular, debt provides lower risk and return, contrary to equities that have a more volatile payoff structure. Investors' learning about the risk-return profile of the different instruments leads to the prevalence of debt on equity or convertible debt. Another conclusion of this model is that learning is imperfect: agents learn to price specific contracts and have difficulties in dealing with contracts that rarely occur in the market.

Future Directions

Agent-based modeling in finance has had a significant impact in shaping the way we understand the working of

financial markets. By introducing realistic behavioral assumptions, agent-based models have demonstrated that they provide a coherent explanation for many empirical findings in finance. In addition, they are also able to provide a framework to explain how aggregate rationality can emerge in a population of bounded rational learning agents.

The strength of the agent-based approach is the ability to specify in greater detail the agents' behavior and the structure of market interactions. Simple agent-based models use realistic assumptions and can be solved analytically. However, they sacrifice the important aspect of the emergence of aggregate pattern based on agents' learning. This can be achieved by computational agent-based models. Since the approach is not bounded by the analytical tractability of the model, very detailed (and potentially more realistic) assumption can be introduced. However, this can represent a weakness of the approach since it might lead to over-parametrized models where it is hard to disentangle the role played by each of the assumptions on the aggregate behavior. In this sense, agent-based modeling should aim at balancing parsimony and realism of agents' description.

As already suggested in Sect. "Other Applications in Finance", the application of agent-based models is not limited to asset pricing issues. Recently, they have been used in corporate finance and market microstructure. This is certainly a trend that will increase in the future since these models are particularly suitable to investigate the interaction of market structure and agents' behavior.

Bibliography

Primary Literature

1. Arifovic J (1996) The Behavior of the Exchange Rate in the Genetic Algorithm and Experimental Economies. J Political Econ 104:510–541
2. Arifovic J, Gencay R (2000) Statistical properties of genetic learning in a model of exchange rate. J Econ Dyn Control 24:981–1006
3. Arthur WB, Holland JH, LeBaron B, Tayler P (1997) Asset pricing under endogeneous expectations in an artificial stock market. In: Lane D, Arthur WB, Durlauf S (ed) The Economy as an Evolving Complex System II. Addison–Wesley, Reading
4. Beltratti A, Margarita S (1992) Evolution of trading strategies among heterogeneous artificial economic agents. In: Wilson SW, Meyer JA, Roitblat HL (ed) From Animals to Animats II. MIT Press, Cambridge, pp 494–501
5. Boswijk HP, Hommes CH, Manzan S (2007) Behavioral heterogeneity in stock prices. J Econ Dyn Control 31:1938–1970
6. Brav A, Heaton JB (2002) Competing Theories of Financial Anomalies. Rev Financ Stud 15:575–606
7. Brock WA, Hommes CH (1998) Heterogeneous beliefs and and routes to chaos in a simple asset pricing model. J Econ Dyn Control 22:1235–1274
8. Chen S-H, Yeh C-H (2002) On the emergence properties of artificial stock markets: the efficient market hypothesis and the rational expectations hypothesis. J Econ Behav Organ 49:217–239
9. Chiarella C (1992) The dynamics of speculative behavior. Ann Oper Res 37:101–123
10. Chiarella C, He XZ (2001) Asset price and wealth dynamics under heterogeneous expectations. Quant Financ 1:509–526
11. Chiarella C, He XZ (2002) Heterogeneous Beliefs, Risk and Learning in a Simple Asset Pricing Model. Comput Econ 19:95–132
12. Chiarella C, He XZ (2003) Heterogeneous beliefs, risk, and learning in a simple asset-pricing model with a market maker. Macroecon Dyn 7:503–536
13. Chiarella C, Iori G (2002) A simulation analysis of the microstructure of double auction markets. Quant Financ 2:346–353
14. Cont R, Bouchaud JP (2000) Herd behavior and aggregate fluctuations in financial markets. Macroecon Dyn 4:170–196
15. Day R, Huang W (1990) Bulls, bears and market sheep. J Econ Behav Organ 14:299–329
16. de Grauwe P, Dewachter H (1993) A chaotic model of the exchange rate: The role of fundamentalists and chartists. Open Econ Rev 4:351–379
18. de Grauwe P, Grimaldi M (2005) The exchange rate and its fundamentals in a complex world. Rev Int Econ 13:549–575
17. de Grauwe P, Dewachter H, Embrechts M (1993) Exchange Rate Theory – Chaotic Models of Foreign Exchange Markets. Blackwell, Oxford
19. Farmer JD, Joshi S (2002) The price dynamics of common trading strategies. J Econ Behav Organ 49:149–171
20. Frankel JA, Froot KA (1987) Using survey data to test standard propositions regarding exchange rate expectations. Am Econ Rev 77:133–153
21. Frankel JA, Froot KA (1990) Chartists, fundamentalists, and trading in the foreign exchange market. Am Econ Rev 80:181–185
22. Gaunersdorfer A (2000) Endogenous fluctuations in a simple asset pricing model with heterogeneous agents. J Econ Dyn Control 24:799–831
23. Gaunersdorfer A, Hommes CH (2006) A nonlinear structural model for volatility clustering. In: Kirman A, Teyssiere G (ed) Microeconomic models for long memory in economics. Springer, Berlin, pp 265–288
24. Goldbaum D (2005) Market efficiency and learning in an endogenously unstable environment. J Econ Dyn Control 29:953–978
25. Grossman SJ, Stiglitz JE (1980) On the Impossibility of Informationally Efficient Markets. Am Econ Rev 70:393–408
26. Hommes CH (2006) Heterogeneous agent models in economics and finance. In: Judd KL, Tesfatsion L (ed) Handbook of Computational Economics, vol 2: Agent-Based Computational Economics. North-Holland, Amsterdam, pp 1109–1185
27. Kirman A (1991) Epidemics of opinion and speculative bubbles in financial markets. In: Taylor MP (ed) Money and Financial Markets. Blackwell, Oxford, pp 354–368
28. Kirman A, Teyssiere G (2001) Microeconomics models for long-

memory in the volatility of financial time series. Stud Nonlinear Dyn Econ 5:281–302

29. LeBaron B (2001) Empirical regularities from interacting long and short memory investors in an agent based in stock market. IEEE Trans Evol Comput 5:442–455

30. LeBaron B (2001) Evolution and time horizons in an agent based stock market. Macroecon Dyn 5:225–254

31. LeBaron B (2002) Short-memory traders and their impact on group learning in financial markets. In: Proceedings of the National Academy of Science: Colloquium, Washington DC, 99:7201–7206

32. LeBaron B (2006) Agent-based computational finance. In: Judd KL, Tesfatsion L (ed) Handbook of Computational Economics, vol 2: Agent-Based Computational Economics. North-Holland, Amsterdam, pp 1187–1233

33. LeBaron B, Yamamoto R (2007) Long-memory in an order-driven market. Physica A 383:85–89

34. Levy M, Levy H, Solomon S (1994) A microscopic model of the stock market. Econ Lett 45:103–111

35. Levy M, Solomon S, Levy H (2000) Microscopic Simulation of Financial Markets. Academic Press, New York

36. Licalzi M, Pellizzari P (2003) Fundamentalists clashing over the book: a study of order-driven stock markets. Quant Financ 3:470–480

37. Lucas RE (1978) Asset prices in an exchange economy. Econometrica 46:1429–1445

38. Lux T (1995) Herd Behaviour, Bubbles and Crashes. Econ J 105:881–896

39. Lux T (1997) Time variation of second moments from a noise trader/infection model. J Econ Dyn Control 22:1–38

40. Lux T (1998) The socio-economic dynamics of speculative markets: interacting agents, chaos, and the fat tails of return distributions. J Econ Behav Organ 33:143–165

41. Lux T, Marchesi M (1999) Scaling and criticality in a stochastic multi-agent model of interacting agents. Nature 397:498–500

42. Lux T, Marchesi M (2000) Volatility clustering in financial markets: A micro-simulation of interacting agents. Int J Theoret Appl Financ 3:675–702

43. Noe TH, Rebello MJ, Wang J (2003) Corporate financing: an artificial agent-based analysis. J Financ 58:943–973

44. Raberto M, Cincotti S, Focardi SM, Marchesi M (2001) Agent-based simulation of a financial market. Physica A: Stat Mech Appl 299:319–327

45. Raberto M, Cincotti S, Focardi SM, Marchesi M (2003) Traders'Long-Run Wealth in an Artificial Financial Market. Comput Econ 22:255–272

46. Rieck C (1994) Evolutionary simulations of asset trading strategies. In: Hillebrand E, Stender J (ed) Many-Agent Simulation and Artificial Life. IOS Press, Amsterdam

47. Routledge BR (1999) Adaptive Learning in Financial Markets. Rev Financ Stud 12:1165–1202

48. Routledge BR (2001) Genetic algorithm learning to choose and use information. Macroecon Dyn 5:303–325

49. Samanidou E, Zschischang E, Stauffer D, Lux T (2007) Microscopic Models of Financial Markets. In: Schweitzer F (ed) Microscopic Models of Economic Dynamics. Springer, Berlin

50. Westerhoff FH (2003) Expectations driven distortions in the foreign exchange market. J Econ Behav Organ 51:389–412

51. Westerhoff FH (2004) Greed, fear and stock market dynamics. Physica A: Stat Mech Appl 343:635–642

Books and Reviews

Anderson PW, Arrow KJ, Pines D (1989) The Economy as an Evolving Complex System. Addison-Wesley, Reading

Arthur WB, Durlauf SN, Lane DA (1999) The Economy as an Evolving Complex System: vol 2. Addison-Wesley, Reading

Blume LE, Durlauf SN (2005) The Economy as an Evolving Complex System: vol 3. Oxford University Press, Oxford

Judd KL, Tesfatsion L (2006) Handbook of Computational Economics, vol 2: Agent-based Computational Economics. North-Holland, Amsterdam

Finance and Econometrics, Introduction to

BRUCE MIZRACH
Department of Economics, Rutgers University,
New Jersey, USA

Article Outline

Introduction

Economics and finance have slowly emerged from the Walrasian, representative agent paradigm exemplified by the research agenda in general equilibrium theory. This program may have reached its pinnacle in the 1970s, with a highly abstract treatment of the existence of a market clearing mechanism. The normative foundation of this research was provided by powerful welfare theorems that demonstrated the optimality of the market allocations. Unfortunately, this abstract world had little economics in it. The models rarely provided empirical implications. Lifetime consumption and portfolio allocation plans were formed in infancy, unemployment was Pareto optimal, and the role for government was largely limited to public goods provision.

The demonstration by Benhabib, Brock, Day, Gale, Grandmont, [1,4,8,9] and others, that even simple math-

ematical models could display highly complex dynamics was the beginning of a new research program in economics. This section on finance and econometrics surveys some of the developments of the last 20 years that were inspired by this research.

Econometrics

Time series econometrics was originally built on the representation theorems for Euclidean spaces. The existence of a Wold decomposition in linear time series led to the widespread use of Box–Jenkins [3] style modeling as an alternative to structural or reduced form models.

A number of stylized facts about the economy emerged that simply could not be explained in this linear world. Rob Engle [2] and Tim Bollerslev [5] showed that volatility was quite persistent, even in markets that appeared to be nearly random walks. In ▶ GARCH Modeling, Christian Hafner surveys the extensive development in this area.

James Hamilton [10] and Salih Neftci [11] demonstrated that the business cycle was asymmetric and could be well described by a Markov switching model. James Morley ▶ Macroeconomics, Non-linear Time Series in and Jeremy Piger ▶ Econometrics: Models of Regime Changes describe the developments in this area. Virtually all the moments, not just the conditional mean, are now thought to be varying over the business cycle. These models help us to understand why recessions are shorter than expansions and why certain variables lead and lag the cycle.

Nearly all the business cycle models involve the use of latent or unobservable state variables. This reflects a reality that policy makers themselves face. We rarely know whether we are in a recession until it is nearly over. These latent variable models are often better described in a Bayesian rather than a classical paradigm. Oleg Korenok ▶ Bayesian Methods in Non-linear Time Series provides an introduction to the frontier research in this area.

Markets are often drawn towards equilibrium states in the absence of exogenous shocks, and, since the 1940s, this simple idea was reflected in the building of macroeconometric models. In linear models, Engle and Granger [6] formalized this notion in an error correction framework. When the adjustment process is taking place between two variables that are not stationary, we say that they are cointegrated. Escanciano and Escribano extend the error correction framework and cointegration analysis to nonlinear models in ▶ Econometrics: Non-linear Cointegration.

Because we often know very little about the data generating mechanism for an economy, nonparametric methods have become increasingly popular in the analysis of time series. Cees Diks discusses in ▶ Nonparametric Tests for Independence methods to analyze both data and the residuals from an econometric model.

Our last two entries look at the data generated by individual consumers and households. Pravan Trivedi ▶ Microeconometrics surveys the microeconometric literature, and Jeff Wooldridge ▶ Econometrics: Panel Data Methods examines the tools and techniques useful for analyzing cross-sectional data.

Agent Based Modeling

The neo-classical synthesis in economics was built upon the abstraction of a single optimizing agent. This assumption simplified the model building and allowed for analytical solutions of the standard models. As computational power became cheaper, it became easier to relax these assumptions. Many economists underestimated the complexity of a world in which multiple agents interact in a dynamic setting. Econophysicists, as Bertrand Roehner describes in ▶ Econophysics, Observational, were not surprised. Roehner is just one of scores of physicists who have brought their tools and perspectives to economics.

Agent based modeling has had a large impact on finance. Financial economics had been led by a Chicago influenced school that saw markets as both rational and efficient. Behavioral finance has eroded the view that people always make optimizing decisions even when large sums of money are at stake. The boundedly rational agents in Sebastiano Manzan's ▶ Finance, Agent Based Modeling in are prone to speculative bubbles. Markets crash suddenly in agent based computational models and in large scale experimental stock markets.

Finance

The foundation of financial economics is the theory of optimal consumption and saving. The goal of the empirical literature was to identify a set of risk factors that would explain why certain assets have a higher return than others. Ralitsa Petkova ▶ Financial Economics, The Cross–Section of Stock Returns and the Fama-French Three Factor Model surveys the canonical model of Fama and French [7] and the extensions to this model in the last decade.

With risk averse agents, asset returns are often predictable. Stijn van Nieuwerburgh and Ralph S.J. Koijen ▶ Financial Economics, Return Predictability and Market Efficiency demonstrate the robustness of this result in a structural model and show that the dividend price ratio does predict future stock returns.

Mototsugu Shintani addresses in ▶ Financial Forecasting, Sensitive Dependence the concept of predictability from an information theoretic perspective through the use of Lyapunov exponents. The exponents not only tell us which systems display sensitive dependence on initial conditions ("chaos") but also provide a predictive horizon for data generated by the model. Shintani finds that financial data appear to not be chaotic, even though they display local dependence on initial conditions.

Mark Kamstra and Lisa Kramer's entry on ▶ Financial Economics, Time Variation in the Market Return primarily focus on the equity premium, the substantially higher return in the US and other countries on equities, over default free securities like Treasury bonds. They document its statistical significance and discuss some behavioral explanations. They demonstrate that behavioral moods can influence asset prices.

Terence Mills' ▶ Financial Economics, Non-linear Time Series in surveys the use of nonlinear time series techniques in finance. Gloria Gonzalez-Rivera and Tae-Hwy Lee look at the ability of nonlinear models to forecast in ▶ Financial Forecasting, Non-linear Time Series in. They also cover the methodology for assessing forecast improvement. The best forecast may not be the one that predicts the mean most accurately; it may instead be the one that keeps you from large losses.

Our last two papers in this area focus on volatility. Markus Haas and Christian Pigorsch discuss the ubiquitous phenomenon of fat-tailed distributions in asset markets in ▶ Financial Economics, Fat-Tailed Distributions. They provide evidence on the frequency of extreme events in many different markets, and develop the implications for risk management when the world is not normally distributed. Torben Andersen and Luca Benzoni ▶ Stochastic Volatility introduce the standard volatility model from the continuous time finance literature. They contrast it with the GARCH model discussed earlier and develop econometric methods for estimating volatility from discretely sampled data.

Market Microstructure

Market microstructure examines the institutional mechanisms by which prices adjust to their fundamental values. The literature has grown with the availability of transactions frequency databases. Clara Vega and Christian Miller ▶ Market Microstructure survey the topic largely from a theoretical perspective. Because disparate markets are likely to have different mechanisms and regulators, the literature has evolved by instrument. Carol Osler ▶ Market Microstructure, Foreign Exchange examines the microstructure of the foreign currency market, the largest and most liquid asset market. Bruce Mizrach and Chris Neely ▶ Treasury Market, Microstructure of the U.S. look at the government bond market in the US as it has evolved into an electronic market. Michael Piwowar ▶ Corporate and Municipal Bond Market Microstructure in the U.S. looks at two bond markets with a large number of issues that trade only very infrequently. Both the markets which he examines have become substantially more transparent through recent government initiatives.

Conclusion

This section covers a wide range of material from theoretical time series analysis to descriptive modeling of financial markets. The theme of complexity is a unifying one in the sense that the models are generally nonlinear and can produce a wide range of possible outcomes. There is complexity in the data which now evolves at a millisecond frequency. Readers should find a variety of perspectives and directions for future research in a heterogenous but interconnected range of fields.

Acknowledgments

Bibliography

1. Benhabib J, Day RH (1982) A characterization of erratic dynamics in the overlapping generations models. J Econ Dyn Control 4:37–55
2. Bollerslev TP (1986) Generalized autoregressive conditional heteroscedasticity. J Econ 31:307–327
3. Box G, Jenkins G (1994) Time Series Analysis Forecasting and Control, 3rd ed. Prentice Hall
4. Brock WA (1986) Distinguishing random and deterministic systems. J Econ Theory 40:168–195
5. Engle RF (1982) Autoregressive conditional heteroscedasticity with estimates of the variance of UK inflation. Econ 50:987–1008
6. Engle RF, Granger CWJ (1987) Co-integration and error correction: representation, estimation, and testing. Econometrica 55:251–276
7. Fama E, French K (1992) The cross-section of expected stock returns. J Finance 47:427–465
8. Gale D (1973) Pure exchange equilibrium of dynamic economic models. J Econ Theory 6:12–36
9. Grandmont JM (1985) On Endogenous Competitive Business Cycles. Econometrica 53:995–1045
10. Hamilton JD (1989) A new approach to the economic analysis of nonstationary time series and the business cycle. Econometrica 57:357–384
11. Neftçi SH (1984) Are economic time series asymmetric over the business cycle? J Political Econ 92:307–328

Financial Economics, The Cross-Section of Stock Returns and the Fama-French Three Factor Model

RALITSA PETKOVA
Mays Business School, Texas A&M University,
College Station, USA

Article Outline

Glossary

Market capitalization Market capitalization is a measure of the size of a public company. It is equal to the share price times the number of shares outstanding. Small stocks have small market capitalizations, while large stocks have large market capitalizations.

Book-to-market ratio A ratio used to compare a company's book value to its market capitalization value. It is calculated by dividing the latest book value by the latest market value of the company.

Value stocks Value stocks tend to trade at lower prices relative to fundamentals like dividends, earnings, sales and others. These stocks are considered undervalued by value investors. Value stocks usually have high dividend yields, and high book-to-market ratios.

Growth stocks Growth stocks tend to trade at higher prices relative to fundamentals like dividends, earnings, sales and others. Growth stocks usually do not pay dividends and have low book-to-market ratios.

Market beta The market beta is a measure of the systematic risk of a security in comparison to the market as a whole. It measures the tendency of the security return to respond to market movements.

Capital asset pricing model (CAPM) The CAPM describes the relationship between risk and expected return and it is used in the pricing of risky securities. According to the CAPM, the expected return of a security equals the rate on a risk-free security plus a risk premium that increases in the security's market beta.

Definition of the Subject

Different stocks have different expected rates of return and many asset pricing models have been developed to understand why this is the case. According to such models, different assets earn different average returns because they differ in their exposures to systematic risk factors in the economy. Fama and French [12] derive a model in which the systematic risk factors are the market index, and two portfolios related to the size of a company, and its ratio of book value to market value (book-to-market). The size and book-to-market factors are empirically motivated by the observation that small stocks and stocks with high book-to-market ratios (value stocks) earn higher average returns than justified by their exposures to market risk (beta) alone. These observations suggest that size and book-to-market may be proxies for exposures to sources of systematic risk different from the market return.

Introduction

An important class of asset pricing models in finance are linear beta models. They assume that the expected return of an asset in excess of the risk-free rate is a linear function of exposures to systematic sources of risk. Usually, the asset's exposures to common sources of risk in the economy are referred to as betas. In general, linear beta models assume the following form for the unconditional expected excess return on assets:

$$E(R_i) = \gamma_M \beta_{i,M} + \sum \gamma_K \beta_{i,K}, \quad \text{for all } i \qquad (1)$$

where $E(R_i)$ is the expected excess return of asset i, γ_M is the market risk premium or the price for bearing market risk, and γ_K is the price of risk for factor K. The model stated above implies that exposures to systematic sources of risk are the only determinants of expected returns. Thus, assets with high betas earn higher expected returns. The betas are the slope coefficients from the following return-generating process:

$$R_{i,t} = \alpha_i + \beta_{i,M} R_{M,t} + \sum \beta_{i,K} K_t + \varepsilon_{i,t}, \quad \text{for all } i \quad (2)$$

where $R_{i,t}$ is the return on asset i in excess of the risk-free rate at the end of period t, $R_{M,t}$ is the excess return on the market portfolio at the end of period t, and K_t is the realization for factor K at the end of period t.

One approach of selecting the pervasive risk factors is based on empirical evidence. For example, many empirical studies document that small stocks have higher average returns than large stocks, and value stock have higher average returns than growth stocks (see [12] for a review). The differences in average returns of these classes of stocks

are statistically and economically significant. If the market sensitivities of small and value stocks were high then their high average returns would be consistent with the Capital Asset Pricing Model (CAPM), which predicts that the market beta is the only determinant of average returns. However, the patterns in returns for these stocks cannot be explained by the CAPM.

In a series of papers, Fama and French [12,13,14] show that a three-factor model performs very well at capturing the size and value effects in average stock returns. The three factors are the excess return on the market portfolio, the return on a portfolio long in value stocks and short in growth stocks, and the return on a portfolio long in small stocks and short in large stocks.

The impressive performance of the Fama–French three-factor model has spurred an enthusiastic debate in the finance literature over what underlying economic interpretation to give to the size and book- to-market factors. One side of the debate favors a risk-based explanation and contends that these factors reflect systematic risks that the static CAPM has failed to capture. For example, if the return distributions of different assets change over time (i. e., expected returns, variances, correlation), then the investment opportunity set available to investors varies over time as well. If individual assets covary with variables that track this variation then the expected returns of these assets will reflect that. Fama and French argue that the factors in their model proxy for such variables.

Another side of the debate favors a non-risk explanation. For example, Lakonishok, Shleifer, and Vishny [22] argue that the book-to-market effect arises since investors over-extrapolate past earnings growth into the future and overvalue companies that have performed well in the past. Namely, investors tend to over-extrapolate recent performance: they overvalue the firms with good recent performance (growth) and undervalue the firms with bad recent performance (value). When the market realizes its mistake, the prices of the former fall, while the prices of the latter rise. Therefore on average, growth firms tend to underperform value firms. Daniel and Titman [9] suggest that stocks characteristics, rather than risks, are priced in the cross-section of average returns. Other authors attribute the success of the size and book-to-market factors to data-snooping and other biases in the data [21,27]. Berk, Green, and Naik [1] and Gomes, Kogan, and Zhang [17] derive models in which problems in the measurement of market beta may explain the Fama–French results.

This article focuses on the risk-based explanation behind the success of the Fama–French three-factor model. If the Fama–French factors are to be explained in the context of a rational asset pricing model, then they should be correlated with variables that characterize time variation in the investment opportunity set. The rest of the article is organized as follows. Section "The Fama–French Model as a Linear Beta Pricing Model" discusses the set-up of the Fama–French model and presents some empirical tests of the model. Section "Explaining the Performance of the Fama–French Model: A Risk-Based Interpretation" argues that the Fama–French factors proxy for fundamental variables that describe variation in the investment opportunity set over time, and presents empirical results. Section "Other Risk-Based Interpretations" presents additional arguments for the relation between the Fama–French factors and more fundamental sources of risk. Section "Future Directions" summarizes and concludes.

The Fama–French Model as a Linear Beta Pricing Model

Model Set-up

Fama and French [12] propose a three-factor linear beta model to explain the empirical performance of small and high book-to-market stocks. The intuition behind the factors they propose is the following.

If small firms earn higher average returns than large firms as a compensation for risk, then the return differential between a portfolios of small firms and a portfolio of large firms would mimic the factor related to size provided the two portfolios have similar exposures to other sources of risk. Similarly, if value firms earn higher average returns than growth firms as a compensation for risk, then the return differential between a portfolio of value firms and a portfolio of growth firms, would mimic the factor related to book-to-market provided the two portfolios have similar exposure to other sources of risk. Fama and French [12] construct two pervasive risk factors in this way that are now commonly used in empirical studies. The composition of these factors is explained below.

In June of each year independent sorts are used to allocate the NYSE, AMEX, and NASDAQ stocks to two size groups and three book-to-market groups. Big stocks are above the median market equity of NYSE firms and small stocks are below. Similarly, low book-to-market stocks are below the 30th percentile of book-to-market for NYSE firms, medium book-to-market stocks are in the middle 40 percent, and high book-to-market stocks are in the top 30 percent. Size is market capitalization at the end of June. Book-to-market is book equity at the last fiscal year end of the prior calendar year divided by market cap as of 6 months before formation. Firms with negative book equity are not considered. At the end of June of each year, six value-weight portfolios are formed, SL, SM, SH, BL, BM,

and BH, as the intersections of the size and book-to-market groups. For example, SL is the value-weight return on the portfolio of stocks that are below the NYSE median in size and in the bottom 30 percent of book-to-market. The portfolios are rebalanced annually. SMB in each period is the difference between the equal-weight averages of the returns on the three small stock portfolios and the three big stock portfolios, constructed to be neutral with respect to book-to-market:

$$SMB = (SL + SM + SH)/3 - (BL + BM + BH)/3 . \quad (3)$$

Similarly, HML in each period is the difference between the return on a portfolio of high book-to-market stocks and the return on a portfolio of low book-to-market stocks, constructed to be neutral with respect to size:

$$HML = (SH + BH)/2 - (SL + BL)/2 . \quad (4)$$

Therefore, the Fama–French three-factor linear model implies that:

$$E(R_i) = \gamma_M \beta_{i,M} + \gamma_{SMB} \beta_{i,SMB} + \gamma_{HML} \beta_{i,HML}, \quad \text{for all } i \quad (5)$$

where $E(R_i)$ is the excess return of asset i, γ_M is the market risk premium, γ_{SMB} is the price of risk for the size factor, and γ_{HML} is the price of risk for the book-to-market factor. The betas are the slope coefficients from the following return-generating process:

$$R_{i,t} = \alpha_i + \beta_{i,M} R_{M,t} + \beta_{i,SMB} R_{SMB,t}$$
$$+ \beta_{i,HML} R_{HML,t} + \varepsilon_{i,t}, \quad \text{for all } i \quad (6)$$

where $R_{i,t}$ is the return on asset i in excess of the risk-free rate at the end of period t, $R_{M,t}$ is the excess return on the market portfolio at the end of period t, $R_{SMB,t}$ is the return on the SMB portfolio at the end of period t, and $R_{HML,t}$ is the return on the HML portfolio at the end of period t.

Testing the Fama–French Model and Results

The return-generating process is Eq. (6) applies to the excess return of any asset. The Fama–French model is usually tested on a set of portfolios sorted by book-to-market and size. Similarly to the construction of HML and SMB, 25 value-weighted portfolios are formed as the intersections of five size and five book-to-market groups. These 25 portfolios are the test assets used most often in testing competing asset-pricing models. These assets represent one of the most challenging set of portfolios in the asset pricing literature.

In this article, monthly data for the period from July of 1963 to December of 2001 is used. The returns on the market portfolio, the risk-free rate, HML, and SMB are taken from Ken French's web site, as well as the returns on 25 portfolios sorted by size and book-to-market.

To test the Fama–French specification in Eq. (5), the Fama–MacBeth [15] cross-sectional method can be used. In the first pass of this method, a multiple time-series regression as in (6) is estimated for each one of the 25 portfolios mentioned above which provides estimates of the assets' betas with respect to the market return, and the size and book-to-market factors.

Table 1 reports the estimates of the factor loadings computed in the first-pass time-series regression (6) for each portfolio. The table also present joint tests of the significance of the corresponding loadings, computed from a seemingly unrelated regressions (SUR) system. This is done in order to show that the Fama–French factors are relevant in the sense that the 25 portfolios load significantly on them.

The results from Table 1 reveal that within each size quintile, the loadings of the portfolios with respect to HML increase monotonically with book-to-market. Within each size group, portfolios in the lowest book-to-market quintile (growth) have negative betas with respect to HML, while portfolios in the highest book-to-market quintile (value) have positive betas with respect to HML. Further, within each book-to-market quintile, the loadings of the portfolios with respect to SMB decrease monotonically with size. Within each book-to-market group, portfolios in the lowest size quintile (small) have positive betas with respect to SMB, while portfolios in the highest size quintile (large) have negative betas with respect to SMB. The table shows that small and large portfolios, and value and growth portfolios have similar market betas.

Note that only six of the 25 intercepts in Table 1 are significant (although the intercepts are jointly significant). The large R-square statistics show that the excess returns of the 25 portfolios are explained well by the three-factor model. Furthermore the large t-statistics on the size and book-to-market betas show that these factors contribute significantly to the explanatory power of the model.

The second step of the Fama–MacBeth procedure involves relating the average excess returns of the 25 portfolios to their exposures to the risk factors in the model. More specifically, the following cross-sectional relation is estimated

$$\overline{R}_{i,t} = \gamma_0 + \gamma_M \widehat{\beta}_{i,M} + (\gamma_{HML}) \widehat{\beta}_{i,HML} + (\gamma_{SMB}) \widehat{\beta}_{i,SMB}$$
$$+ e_{i,t} . \quad (7)$$

Financial Economics, The Cross-Section of Stock Returns and the Fama-French Three Factor Model, Table 1

Loadings on the Fama–French Factors from Time-Series Regressions

This table reports loadings on the excess market return, R_M, and the Fama–French factors R_{HML} and R_{SMB} computed in time-series regressions for 25 portfolios sorted by size and book-to-market. The corresponding t-statistics are also reported and are corrected for autocorrelation and heteroscedasticity using the Newey–West estimator with five lags. The sample period is from July 1963 to December 2001. The intercepts are in percentage form. The last column reports F-statistics and their corresponding p-values from an SUR system, testing the joint significance of the corresponding loadings. The p-values are in percentage form. R^2s from each time-series regression are reported in percentage form

Regression: $R_{i,t} = \alpha_i + \beta_{i,M} R_{M,t} + \beta_{i,HML} R_{HML,t} + \beta_{i,SMB} R_{SMB,t} + \varepsilon_{i,t}$

	Low	2	3	4	High		Low	2	3	4	High	F
			α						t_α			F
Small	−0.38	0.01	0.04	0.18	0.12		−3.40	0.18	0.56	2.84	1.91	2.96
2	−0.17	−0.10	0.08	0.08	−0.00		−2.25	−1.45	1.15	1.28	−0.01	0.01
3	−0.07	−0.00	−0.09	0.01	0.00		−1.03	−0.03	−1.26	0.17	0.06	
4	0.16	0.21	−0.08	0.04	−0.05		1.67	−2.27	−0.99	0.61	−0.54	
Large	0.21	−0.04	−0.02	−0.09	−0.21		3.25	−0.53	−0.27	−1.29	−2.36	
			β_M						t_{β_M}			F
Small	1.04	0.96	0.93	0.92	0.98		44.38	39.40	50.88	46.60	43.39	> 100
2	1.11	1.03	1.00	0.99	1.08		48.84	45.42	46.47	60.69	52.11	< 0.01
3	1.09	1.07	1.03	1.01	1.10		52.59	38.53	32.93	52.70	38.97	
4	1.05	1.11	1.08	1.03	1.17		46.03	36.33	36.86	41.15	36.74	
Large	0.96	1.04	0.99	1.01	1.04		45.08	49.22	36.71	46.18	31.59	
			β_{HML}						$t_{\beta_{HML}}$			F
Small	−0.31	0.09	0.31	0.47	0.69		−5.86	1.79	9.62	14.97	17.10	> 100
2	−0.38	0.18	0.43	0.59	0.76		−8.52	2.96	7.36	13.97	23.28	< 0.01
3	−0.43	0.22	0.52	0.67	0.82		−14.90	3.10	7.39	10.58	15.94	
4	−0.45	0.26	0.51	0.61	0.83		−10.55	3.42	7.43	11.92	16.07	
Large	−0.38	0.14	0.27	0.64	0.85		−10.47	2.58	5.65	11.82	20.56	
			β_{SMB}						$t_{\beta_{SMB}}$			F
Small	1.41	1.33	1.12	1.04	1.09		36.39	24.68	36.50	24.34	25.40	> 100
2	1.00	0.89	0.75	0.70	0.82		27.61	18.51	15.90	25.31	25.68	< 0.01
3	0.72	0.51	0.44	0.38	0.53		24.97	7.68	6.81	8.28	8.87	
4	0.37	0.20	0.16	0.20	0.26		9.26	3.42	2.64	6.70	4.22	
Large	−0.26	−0.24	−0.24	−0.22	−0.08		−9.25	−6.92	−6.12	−6.81	−2.11	
						R^2						
		92.61	94.32	94.89	94.51	94.58						
		95.16	93.99	93.56	93.85	94.62						
		94.88	90.22	89.49	89.69	90.31						
		93.52	88.31	87.65	88.41	85.77						
		93.35	89.79	84.32	87.39	80.60						

The $\widehat{\beta}$ terms are the independent variables in the regression, while the average excess returns of the assets are the dependent variables. If loadings with respect to the Fama–French factors are important determinants of average returns, then there should be a significant price of risk associated with the factors.

Since the betas are estimated from the time-series regression in (6), they represent generated regressors in (7). This is the classical errors-in-variables problem, arising from the two-pass nature of this approach. Following Shanken [33], a correction procedure can be used that accounts for the errors-in-variables problem. Shanken's correction is designed to adjust for the overstated precision of the Fama–MacBeth standard errors. It assumes that the error terms from the time-series regression are independently and identically distributed over time, conditional on the time series of observations for the risk factors. The adjustment also assumes that the risk factors are generated by a stationary process. Jagannathan and Wang [19] argue that if the error terms are heteroscedastic, then the Fama–

Financial Economics, The Cross-Section of Stock Returns and the Fama-French Three Factor Model, Table 2

Cross-Sectional Regressions with the Fama–French Factor Loadings

This table presents Fama–MacBeth cross-sectional regressions using the average excess returns on 25 portfolios sorted by book-to-market and size. The full-sample factor loadings, which are the independent variables in the regressions, are computed in one multiple time-series regression. The coefficients are expressed as percentage per month. The Adjusted R^2 follows Jagannathan and Wang [18] and is reported in percentage form. The first set of t-statistics, indicated by FM t-stat, stands for the Fama–MacBeth estimate. The second set, indicated by SH t-stat, adjusts for errors-in-variables and follows Shanken [33]. The sample period is from July 1963 to December 2001

The Fama–French Three-Factor Model					
	γ_0	γ_M	γ_{HML}	γ_{SMB}	Adj. R^2
Estimate	1.15	−0.65	0.44	0.16	71.00
FM t-stat	3.30	−1.60	3.09	1.04	
SH t-stat	3.19	−1.55	3.07	1.00	

Financial Economics, The Cross-Section of Stock Returns and the Fama-French Three Factor Model, Figure 1

Fitted Expected Returns vs. Average Realized Returns for 1963:07-2001:12.

This figure shows realized average returns (%) on the *horizontal axis* and fitted expected returns (%) on the *vertical axis* for 25 size and book-to-market sorted portfolios. Each two-digit number represents a separate portfolio. The first digit refers to the size quintile (1 being the smallest and 5 the largest), while the second digit refers to the book-to-market quintile (1 being the lowest and 5 the highest). For each portfolio, the realized average return is the time-series average of the portfolio return and the fitted expected return is the fitted value for the expected return from the corresponding model. The straight line is the 45-degree line from the origin

MacBeth procedure does not necessarily result in smaller standard errors of the cross-sectional coefficients. In light of these two issues, researchers often report both unadjusted and adjusted cross-sectional statistics.

Table 2 reports the estimates of the factor prices of risk computed in the second-pass cross-sectional regression (7). The table also presents the t-statistics for the coefficients, adjusted for errors-in-variables following Shanken [33]. The table shows that the market beta is not an important factor in the cross-section of returns sorted by size and book-to-market.[1] Further, the table reveals that loadings on *HML* represent a significant factor in the cross-section of the 25 portfolios, even after correcting for the sampling error in the loadings. Loadings on *SMB* do not appear to be significant in the cross-section of portfolio returns for this time period. The large R-square of 0.71 shows that the loadings from the Fama–French model explain a significant portion of the cross-sectional variation in the average returns of these portfolios.

It is also helpful to examine the performance of the model visually. This is done by plotting the fitted expected return of each portfolio against its realized average return in Fig. 1. The fitted expected return is computed using the estimated parameter values from the Fama–French model specification. The realized average return is the time-series average of the portfolio return. If the fitted expected return

and the realized average return for each portfolio are the same, then they should lie on a 45-degree line through the origin.

Figure 1 shows the fitted versus realized returns for the 25 portfolios in two different models for the period from July of 1963 to December of 2001. Each two-digit number represents a separate portfolio. The first digit refers to the size quintile of the portfolio (1 being the smallest and 5 the biggest), while the second digit refers to the book-to-market quintile (1 being the lowest and 5 the highest). For example, portfolio 15 has the highest book-to-market value among the portfolios in the smallest size quintile. In other words, it is the smallest value portfolio.

It can be seen form the graph that the model goes a long way toward explaining the value effect: in general, the fitted expected returns on value portfolios (bigger second digit) are higher than the fitted expected returns on growth portfolios (lower second digit). This is consistent with the data on realized average returns for these portfolios. By inspection of Fig. 1, a few portfolios stand out as problematic for the FF model, in terms of distance from

---

[1]The estimate of the market risk premium tends to be negative. This result is consistent with previous results reported in the literature. Fama and French [11], Jagannathan and Wang [18], and Lettau and Ludvigson [24] report negative estimates for the market risk premium, using monthly or quarterly data.

the 45-degree line, namely the growth portfolios within the smallest and largest size quintiles (11, 41, and 51) and the value portfolios within the largest size quintiles (45, 54, and 55).

In summary, the Fama–French model performs remarkably well at explaining the average return difference between small and large, and value and growth portfolios.

The natural question that arises is what drives the superior performance of the Fama–French model in explaining average stock returns. One possible explanation is that the Fama–French factors *HML* and *SMB* proxy for sources of risk not captured by the return on the market portfolio. This explanation is consistent with a multifactor asset pricing model like the Intertemporal Capital Asset Pricing Model (ICAPM), which states that if investment opportunities change over time, then variables other than the market return will be important factors driving stock returns. Therefore, one possible interpretation of the *HML* and *SMB* portfolios is that they proxy for variables that describe how investment opportunities change over time. The following sections examine the ICAPM explanation behind the performance of the Fama–French model.

Explaining the Performance of the Fama–French Model: A Risk-Based Interpretation

The ICAPM Framework

The analysis in this paper assumes that asset returns are governed by the discrete-time version of the ICAPM of Merton [29]. According to the ICAPM, if investment opportunities change over time, then assets' exposures to these changes are important determinants of average returns in addition to the market beta. Campbell [3] develops a framework to model changes in the investment opportunity set as innovations in state variables that capture uncertainty about investment opportunities in the future. Therefore, the model for the unconditional expected excess returns on assets becomes

$$E(R_i) = \gamma_M \beta_{i,M} + \sum (\gamma_{u^K}) \beta_{i,u^K}, \quad \text{for all } i \qquad (8)$$

where $E(R_i)$ is the excess return of asset i, γ_M is the market risk premium, and γ_{u^K} is the price of risk for innovations in state variable K. The betas are the slope coefficients from the following return-generating process:

$$R_{i,t} = \alpha_i + \beta_{i,M} R_{M,t} + \sum (\beta_{i,u^K}) u_t^K + \varepsilon_{i,t}, \quad \text{for all } i \quad (9)$$

where $R_{i,t}$ is the return on asset i in excess of the risk-free rate at the end of period t, $R_{M,t}$ is the excess return on the market portfolio at the end of period t, and u_t^K is the innovation to state variable K at the end of period t. The

innovation is the unexpected component of the variable. According to the asset-pricing model, only the unexpected component of the state variable should command a risk premium. Note that the innovations to the state variables are contemporaneous to the excess market returns. This equation captures the idea that the market portfolio and the innovations to the state variables are the relevant risk factors.

It is important to specify a process for the time-series dynamics of the state variables in the model. A vector autoregressive (VAR) approach, for example, specifies the excess market return as the first element of a state vector z_t. The other elements of z_t are state variables that proxy for changes in the investment opportunity set. The assumption is that the demeaned vector z_t follows a first-order VAR:

$$z_t = \mathbf{A} z_{t-1} + \mathbf{u}_t . \qquad (10)$$

The residuals in the vector $\mathbf{u}_t$ are the innovations terms which are the risk factors in Eq. (2). These innovations are risk factors since they represent the surprise components of the state variables that proxy for changes in the investment opportunity set.

The State Variables of Interest

For the empirical implementation of the model described above, it is necessary to specify the identity of the state variables. Petkova [31] chooses a set of state variables to model the following aspects of the investment opportunity set: the yield curve and the conditional distribution of asset returns. In particular, she chooses the short-term Treasury bill, the term spread, the aggregate dividend yield, and the default spread.

The choice of these state variables is motivated as follows. The ICAPM dictates that the yield curve is an important part of the investment opportunity set. Furthermore, Long [28] points out that the yield curve is important in an economy with a bond market. Therefore, the short-term Treasury bill yield (*RF*) and the term spread (*TERM*) are good candidates that capture variations in the level and the slope of the yield curve. Litterman and Scheinkman [26] show that the two most important factors driving the term structure of interest rates are its level and its slope.

In addition to the yield curve, the conditional distribution of asset returns is a relevant part of the investment opportunity set facing investors in the ICAPM world. There is growing evidence that the conditional distribution of asset returns, as characterized by its mean and variance, changes over time. The time-series literature has identified variables that proxy for variation in the mean and variance

of returns. The aggregate dividend yield (*DIV*), the default spread (*DEF*), and interest rates are among the most common.[2]

The variables described above are good candidates for state variable within the ICAPM. Merton [29] states that stochastic interest rates are important for changing investment opportunities. In addition, the default spread, the dividend yield, and interest rate variables have been used as proxies for time-varying risk premia under changing investment opportunities. Therefore, all these variables are likely to capture the hedging concerns of investors related the changes in interest rates and to variations in risk premia.

As argued in the previous sections of this article, two other variables proposed as candidates for state variables within the ICAPM are the returns on the *HML* and *SMB* portfolios. Fama and French [12] show that these factors capture common variation in portfolio returns that is independent of the market and that carries a different risk premium. The goal of the following section is to show that the FF factors proxy for the state variables described above that have been shown to track time-variation in the market risk premium and the yield curve.

Econometric Approach

First, a vector autoregressive (VAR) process for the vector of state variables is specified. The first element of the vector is the excess return on the market, while the other elements are *DIV*, *TERM*, *DEF*, *RF*, R_{HML}, and R_{SMB}, respectively. For convenience, all variables in the state vector have been demeaned. The first-order VAR is as follows:

$$\left\{ \begin{array}{c} R_{M,t} \\ DIV_t \\ TERM_t \\ DEF_t \\ RF_t \\ R_{HML,t} \\ R_{SMB,t} \end{array} \right\} = \mathbf{A} \left\{ \begin{array}{c} R_{M,t-1} \\ DIV_{t-1} \\ TERM_{t-1} \\ DEF_{t-1} \\ RF_{t-1} \\ R_{HML,t-1} \\ R_{SMB,t-1} \end{array} \right\} + \mathbf{u}_t \qquad (11)$$

where $\mathbf{u}_t$ represents a vector of innovations for each element in the state vector. From $\mathbf{u}_t$ six surprise series can be extracted, corresponding to the dividend yield, the term spread, the default spread, the one-month T-bill yield, and the FF factors. They are denoted as follows: u^{DIV}, u^{TERM}, u^{DEF}, u^{RF}, u^{HML}, and u^{SMB}, respectively. This VAR rep-

resents a joint specification of the dynamics of all candidate state variables within the ICAPM. This specification treats the FF factors as potential candidates for state variables that command separate risk premia from the other variables.

The innovations derived from the VAR model are risk factors in addition to the excess return of the market portfolio. Asset's exposures to these risk factors are important determinants of average returns according to the ICAPM. To test the ICAPM specification, the Fama–MacBeth [15] cross-sectional method can be used as previously discussed. In the first pass of this method, a multiple time-series regression is specified which provides estimates of the assets' loadings with respect to the market return and the innovations in the state variables. More precisely, the following time-series regression is examined for each asset:

$$\begin{aligned} R_{i,t} = \alpha_i &+ \beta_{i,M} R_{M,t} + (\beta_{i,\hat{u}^{DIV}}) \hat{u}_t^{DIV} + (\beta_{i,\hat{u}^{TERM}}) \hat{u}_t^{TERM} \\ &+ (\beta_{i,\hat{u}^{DEF}}) \hat{u}_t^{DEF} + (\beta_{i,\hat{u}^{RF}}) \hat{u}_t^{RF} + (\beta_{i,\hat{u}^{HML}}) \hat{u}_t^{HML} \\ &+ (\beta_{i,\hat{u}^{SMB}}) \hat{u}_t^{SMB} + \varepsilon_{i,t}, \ \ \text{for all} \ \ i \ . \end{aligned}$$
$$(12)$$

The $\hat{u}$-terms represent the estimated surprises in the state variables. Note that the innovations terms are generated regressors and they appear on the right-hand side of the equation. However, as pointed out by Pagan [30], the OLS estimates of the parameters' standard errors will still be correct if the generated regressor represents the unanticipated part of a certain variable. On the other hand, if the $\hat{u}$-terms are only noisy proxies for the true surprises in the state variables, then the estimates of the factor loadings in the above regression will be biased downwards. This will likely bias the results against finding a relation between the innovations and asset returns.

The second step of the Fama–MacBeth procedure involves relating the average excess returns of all assets to their exposures to the risk factors in the model. Therefore, the following cross-sectional relation applies

$$\begin{aligned} \overline{R}_{i,t} = \gamma_0 &+ \gamma_M \widehat{\beta}_{i,M} + (\gamma_{\hat{u}^{DIV}}) \widehat{\beta}_{i,\hat{u}^{DIV}} + (\gamma_{\hat{u}^{TERM}}) \widehat{\beta}_{i,\hat{u}^{TERM}} \\ &+ (\gamma_{\hat{u}^{DEF}}) \widehat{\beta}_{i,\hat{u}^{DEF}} + (\gamma_{\hat{u}^{RF}}) \widehat{\beta}_{i,\hat{u}^{RF}} + (\gamma_{\hat{u}^{HML}}) \widehat{\beta}_{i,\hat{u}^{HML}} \\ &+ (\gamma_{\hat{u}^{SMB}}) \widehat{\beta}_{i,\hat{u}^{SMB}} + e_{i,t}, \ \ \text{for all} \ \ t \ . \end{aligned}$$
$$(13)$$

Data, Time-Series Analysis, and Results

In this section, monthly data for the period from July of 1963 to December of 2001 is used. The state variables in the context of the ICAPM are the dividend yield of the

---

[2]The following is only a partial list of papers that document time-variation in the excess market return and the variables they use: Campbell [2], term spread; Campbell and Shiller [4], dividend yield; Fama and Schwert [16], T-bill rate; Fama and French [10], default spread.

value-weighted market index (computed as the sum of dividends over the last 12 months, divided by the level of the index), the difference between the yield of a 10-year and a 1-year government bond (term spread), the difference between the yield of a long-term corporate Baa bond and a long-term government bond (default spread), and the one-month Treasury-bill yield. Data on bond yields is taken from the FRED® database of the Federal Reserve Bank of St. Louis. The T-bill yield and the term spread are used to measure the level and the slope of the yield curve, respectively.

VAR Estimation   The state variables are the FF factors and the four predictive variables described above. All of them are included in a first-order VAR system. Campbell [3] emphasizes that it is hard to interpret estimation results for a VAR factor model unless the factors are orthogonalized and scaled in some way. In his paper the innovations to the state variables are orthogonal to the excess market return and to labor income. Following Campbell, the VAR system in Eq. (4) is triangularized in a similar way: the innovation in the excess market return is unaffected, the orthogonalized innovation in DIV is the component of the original DIV innovation orthogonal to the excess market return, and so on. The orthogonalized innovation to DIV is a change in the dividend/price ratio with no change in the market return, therefore it can be interpreted as a shock to the dividend. Similarly, shocks to the term spread, the default spread, the short-term rate, and the FF factors are orthogonal to the contemporaneous stock market return. As in Campbell [3], the innovations are scaled to have the same variance as the innovation in the excess market return.

It is interesting to note that the returns on the FF factors are very highly correlated with their respective innovation series. For example, the correlation between $R_{HML,t}$ and $\hat{u}_t^{HML}$ is 0.90, while the correlation between $R_{SMB,t}$ and $\hat{u}_t^{SMB}$ is 0.92. Therefore, the returns on the HML and SMB portfolios are good proxies for the innovations associated with those variables.

Relation Between R_{HML} and R_{SMB} and the VAR Innovations   As a first step towards testing whether the FF factors proxy for innovations in state variables that track investment opportunities, the joint distribution of R_{HML} and R_{SMB} and innovations to DIV, $TERM$, DEF, and RF is examined. The following time-series regression is analyzed

$$\hat{u}_t = c_0 + c_1 R_{M,t} + c_2 R_{HML,t} + c_3 R_{SMB,t} + \varepsilon_t \quad (14)$$

for each series of innovations in the state variables. The results for these regressions are presented in Table 3, with

Financial Economics, The Cross-Section of Stock Returns and the Fama-French Three Factor Model, Table 3
Time-Series Regressions Showing the Contemporaneous Relations Between Innovations in State Variables and the Fama–French Factors
This table presents time-series regressions of innovations in the dividend yield ($\hat{u}_t^{DIV}$), term spread ($\hat{u}_t^{TERM}$), default spread ($\hat{u}_t^{DEF}$), and one-month T-bill yield ($\hat{u}_t^{RF}$) on the excess market return, R_M, and the Fama–French factors R_{HML} and R_{SMB}. The innovations to the state variables are computed in a VAR system. The t-statistics are below the coefficients and are corrected for heteroscedasticity and autocorrelation using the Newey–West estimator with five lags. The Adjusted R^2 is reported in percentage form. The sample period is from July 1963 to December 2001

Regression: $\hat{u}_t = c_0 + c_1 R_{M,t} + c_2 R_{HML,t} + c_3 R_{SMB,t} + \varepsilon_t$					
Dep. Variable	c_0	c_1	c_2	c_3	Adj. R^2
$\hat{u}_t^{DIV}$	0.00	−0.08	−0.30	−0.01	3.00
	0.85	−0.70	−2.43	−0.09	
$\hat{u}_t^{TERM}$	−0.00	0.06	0.24	0.03	2.00
	−0.56	0.75	2.30	0.59	
$\hat{u}_t^{DEF}$	−0.00	0.07	0.17	−0.12	2.00
	−0.38	1.11	2.10	−1.92	
$\hat{u}_t^{RF}$	0.00	−0.04	−0.13	0.01	0.00
	0.36	−0.51	−1.36	0.14	

the corresponding t-statistics, below the coefficients, corrected for heteroscedasticity and autocorrelation. Innovations in the dividend yield, $\hat{u}_t^{DIV}$, covary negatively and significantly with the return on HML. In addition, $\hat{u}_t^{TERM}$ covaries positively and significantly with the HML return. These results are robust to the presence of the market factor in the regression. The return on the HML portfolio covaries positively and significantly with $\hat{u}_t^{DEF}$, while the return on the SMB factor covaries negatively with $\hat{u}_t^{DEF}$ (the corresponding t-statistic is marginally significant). The last regression in Table 3 indicates that the FF factors are not significant determinants of innovations in the T-bill yield. The results in the table remain unchanged if the independent variables in the equation above are the innovations to R_{HML} and R_{SMB} derived from the VAR system. The R-squares in the regressions reported in Table 3 are rather low. This does not imply, however, that the innovations in the state variables cannot potentially price assets as well as the FF factors. It could be the case that only the information in the FF factors correlated with the state variables is relevant for the pricing of risky assets. A similar point is made by Vassalou [34].

As pointed out by FF [10], the values of the term spread signal that expected market returns are low during expansions and high during recessions. In addition, FF document that the term spread very closely tracks the short-

term fluctuations in the business cycle. Therefore, positive shocks to the term premium are associated with bad times in terms of business conditions, while negative shocks are associated with good times. In light of the results documented by Petkova and Zhang [32], that value stocks are riskier than growth stocks in bad times and less risky during good times, the relation between *HML* and shocks to the term spread seems natural.

Another interpretation of the relation between shocks to the term spread and the *HML* portfolio is in the context of cash flow maturities of assets. This point is discussed by Cornell [8] and Campbell and Vuolteenaho [5]. The argument is that growth stocks are high-duration assets, which makes them similar to long-term bonds and more sensitive to innovations in the long end of the term structure. Similarly, value stocks have lower duration than growth stocks, which makes them similar to short-term bonds and more sensitive to shocks to the short end of the yield curve.

Chan and Chen [6] have argued that small firms examined in the literature tend to be marginal firms, that is, they generally have lost market value due to poor performance, they are likely to have high financial leverage and cash flow problems, and they are less likely to survive poor economic conditions. In light of this argument, it is reasonable to assume that small firms will be more sensitive to news about the state of the business cycle. Therefore, it is puzzling that I find no significant relation between *SMB* and surprises to the term spread. Innovations in the term spread seem to be mostly related to *HML*. This observation suggests that the *HML* portfolio might represent risk related to cash flow maturity, captured by unexpected movements in the slope of the term structure.

Innovations in default spread, u_t^{DEF}, stand for changes in forecasts about expected market returns and changes in forecasts about default spread. FF [10] show that the default premium tracks time variation in expected returns that tends to persist beyond the short-term fluctuations in the business cycle. A possible explanation for the negative relation between *SMB* and shocks to the default spread could be that bigger stocks are able to track long-run trends in the business cycle better than the smaller stocks. The result that *HML* is also related to shocks in the default spread is consistent with the interpretation of *HML* as a measure of distress risk. The distress risk interpretation of the book-to-market effect is advocated by FF [11,12,13,14] and Chen and Zhang [7], among others.

In summary, the empirical literature has documented that both value and small stocks tend to be under distress, with high leverage and cash flow uncertainty. The results in this study suggest that the book-to-market factor might be related to asset duration risk, measured by the slope of

the term structure, while the size factor might be related to asset distress risk, measured by the default premium.

It is reasonable to test whether the significant relation between the state variables surprises and the FF factors gives rise to the significant explanatory power of *HML* and *SMB* in the cross-section of returns. The next section examines whether *HML* and *SMB* remain significant risk factors in the presence of innovations to the other state variables. The results from the cross-sectional regressions suggest that *HML* and *SMB* lose their explanatory power for the cross-section of returns once accounting for the other variables. This supports an ICAPM explanation behind the empirical success of the FF three-factor model.

Cross-Sectional Regressions

Incremental Explanatory Power of the Fama–French Factors   This section examines the pricing performance of the full set of state variables considered before over the period from July 1963 to December 2001. The full set of state variables consists of the dividend yield, the term spread, the default spread, the short-term T-bill yield, and the FF factors. The innovations to these state variables derived from a VAR system are risk factors in the ICAPM model. The objective is to test whether an asset's loadings with respect to these risk factors are important determinants of its average return.

The first specification is

$$
\begin{aligned}
\overline{R}_{i,t} = {} & \gamma_0 + \gamma_{MKT}\widehat{\beta}_{i,MKT} + (\gamma_{\hat{u}^{DIV}})\widehat{\beta}_{i,\hat{u}^{DIV}} \\
& + (\gamma_{\hat{u}^{TERM}})\widehat{\beta}_{i,\hat{u}^{TERM}} + (\gamma_{\hat{u}^{DEF}})\widehat{\beta}_{i,\hat{u}^{DEF}} \\
& + (\gamma_{\hat{u}^{RF}})\widehat{\beta}_{i,\hat{u}^{RF}} + (\gamma_{\hat{u}^{HML}})\widehat{\beta}_{i,\hat{u}^{HML}} \\
& + (\gamma_{\hat{u}^{SMB}})\widehat{\beta}_{i,\hat{u}^{SMB}} + e_{i,t} ,
\end{aligned}
\tag{15}
$$

where the $\widehat{\beta}$ terms stand for exposures to the corresponding factor, while the γ terms stand for the reward for bearing the risk of that factor. The $\widehat{\beta}$ terms are the independent variables in the regression, while the average excess returns of the assets are the dependent variables. If loadings with respect to innovations in a state variable are important determinants of average returns, then there should be a significant price of risk associated with that state variable.

The results are reported in Table 4. The table shows that assets' exposures to innovations in R_{HML} and R_{SMB} are not significant variables in the cross-section in the presence of betas with respect to surprises in the other state variables. The corresponding *t*-statistics are 1.40 and 1.56, respectively, under the errors-in-variables correction. Therefore, based on the results presented in Table 4, the hypothesis that innovations in the dividend yield, the

Financial Economics, The Cross-Section of Stock Returns and the Fama-French Three Factor Model, Table 4
Cross-Sectional Regressions Showing the Incremental Explanatory Power of the Fama–French Factor Loadings
This table presents Fama–MacBeth cross-sectional regressions using the average excess returns on 25 portfolios sorted by book-to-market and size. The full-sample factor loadings, which are the independent variables in the regressions, are computed in one multiple time-series regression. The coefficients are expressed as percentage per month. The table presents results for the model including the excess market return, R_M, and innovations in the dividend yield, term spread, default spread, one-month T-bill yield, and the Fama–French factors HML and SMB. The Adjusted R^2 follows Jagannathan and Wang [18] and is reported in percentage form. The first set of t-statistics, indicated by FM t-stat, stands for the Fama–MacBeth estimate. The second set, indicated by SH t-stat, adjusts for errors-in-variables and follows Shanken [33]. The table examines the sample period from July 1963 to December 2001

The Model with Innovations in All State Variables								
	γ_0	γ_M	$\gamma_{\hat{u}^{DIV}}$	$\gamma_{\hat{u}^{TERM}}$	$\gamma_{\hat{u}^{DEF}}$	$\gamma_{\hat{u}^{RF}}$	$\gamma_{\hat{u}^{HML}}$	$\gamma_{\hat{u}^{SMB}}$	Adj. R^2
Estimate	1.11	−0.57	−0.83	3.87	0.37	−2.90	0.42	0.41	77.26
FM t-stat	3.29	−1.45	−0.94	3.53	0.42	−3.33	1.62	1.75	
SH t-stat	2.36	−1.10	−0.69	2.56	0.31	−2.44	1.40	1.56	

term spread, the default spread, and the short-term T-bill span the information contained in the FF factors cannot be rejected.

A Model Based on R_M, and Innovations in DIV, $TERM$, DEF, and RF   This part examines separately the set of innovations in the variables associated with time-series predictability: the dividend yield, the term spread, the default spread, and the short-term T-bill. The model specification is as follows

$$R_{i,t} = \alpha_i + \beta_{i,M} R_{M,t} + (\beta_{i,\hat{u}^{DIV}})\hat{u}_t^{DIV}$$
$$+ (\beta_{i,\hat{u}^{TERM}})\hat{u}_t^{TERM} + (\beta_{i,\hat{u}^{DEF}})\hat{u}_t^{DEF}$$
$$+ (\beta_{i,\hat{u}^{RF}})\hat{u}_t^{RF} + \varepsilon_{i,t}, \quad \text{for all } i \qquad (16)$$

$$\overline{R}_{i,t} = \gamma_0 + \gamma_M \widehat{\beta}_{i,M} + (\gamma_{\hat{u}^{DIV}})\widehat{\beta}_{i,\hat{u}^{DIV}} + (\gamma_{\hat{u}^{TERM}})\widehat{\beta}_{i,\hat{u}^{TERM}}$$
$$+ (\gamma_{\hat{u}^{DEF}})\widehat{\beta}_{i,\hat{u}^{DEF}} + (\gamma_{\hat{u}^{RF}})\widehat{\beta}_{i,\hat{u}^{RF}} + e, \quad \text{for all } t$$
$$\qquad (17)$$

which corresponds to a model in which the relevant risk factors are innovations to predictive variables. The objective is to compare the pricing performance of this model with that of the Fama–French model for the cross-section of returns sorted by book-to-market and size. The specification is motivated by the previous observation that HML and SMB do not add explanatory power to the set of state variables that are associated with time-series predictability.

Table 5 report the estimates of the factor loadings computed in the first-pass time-series regressions defined in Eq. (16). It also presents joint tests of the significance of the corresponding loadings, computed from a SUR system. This is done in order to show that the innovations

factors are relevant in the sense that the 25 portfolios load significantly on them. A similar analysis was performed on the Fama–French model in Sect. "The Fama–French Model as a Linear Beta Pricing Model".

An F-test implies that the 25 loadings on innovations to the term spread are jointly significant, with the corresponding p-value being 0.47%. Furthermore, portfolios' loadings on $\hat{u}_t^{TERM}$ are related to book-to-market: within each size quintile, the loadings increase monotonically from lower to higher book-to-market quintiles. In fact, the portfolios within the lowest book-to-market quintile have negative sensitivities with respect to $\hat{u}_t^{TERM}$, while the portfolios within the highest book-to-market quintile have positive loadings on $\hat{u}_t^{TERM}$. This pattern resembles very much the one observed in Table 1 for the loadings on R_{HML}.

Similarly, loadings on shocks to default spread are jointly significant in Table 5, with the corresponding p-value being 0.24%. Moreover, the slopes on $\hat{u}_t^{DEF}$ are systematically related to size. Within each book-to-market quintile, the loadings increase almost monotonically from negative values for the smaller size quintiles to positive values for the larger size quintiles. This pattern closely resembles the mirror image of the one observed in Table 1 for the loadings on R_{SMB}. The slopes on dividend yield and T-bill innovations do not exhibit any systematic patterns related to size or book-to-market. However, both of these are jointly significant.

Note that the R^2s in the time-series regressions with the innovations factors in Table 5 are smaller than the ones in the regressions with the FF factors in Table 1. This indicates that potential errors-in-variables problems that arise in measuring the factor loadings will be more serious in the case of the innovations terms. Therefore, the results will be potentially biased against finding significant factor

Financial Economics, The Cross-Section of Stock Returns and the Fama-French Three Factor Model, Table 5

Loadings on R_M, $\hat{u}_t^{DIV}$, $\hat{u}_t^{TERM}$, $\hat{u}_t^{DEF}$, and $\hat{u}_t^{RF}$ from Time-Series Regressions

This table reports loadings on the excess market return, R_M, and innovations in the dividend yield ($\hat{u}_t^{DIV}$), term spread ($\hat{u}_t^{TERM}$), default spread ($\hat{u}_t^{DEF}$), and short-term T-bill ($\hat{u}_t^{RF}$) computed in time-series regressions for 25 portfolios sorted by size and book-to-market. The corresponding t-statistics are also reported and are corrected for autocorrelation and heteroscedasticity using the Newey–West estimator with five lags. The sample period is from July 1963 to December 2001. The last column reports F-statistics and their corresponding p-values from an SUR system, testing the joint significance of the corresponding loadings. The p-values are in percentage form. R^2s from each time-series regression are reported in percentage form

Regression: $R_{i,t} = \alpha_i + \beta_{i,M}R_{M,t} + \beta_{i,\hat{u}^{DIV}}\hat{u}_t^{DIV} + \beta_{i,\hat{u}^{TERM}}\hat{u}_t^{TERM} + \beta_{i,\hat{u}^{DEF}}\hat{u}_t^{DEF} + \beta_{i,\hat{u}^{RF}}\hat{u}_t^{RF} + \varepsilon_{i,t}$											
	Low	2	3	4	High		Low	2	3	4	High
β_{MKT}						$t_{\beta_{MKT}}$						F
Small	1.44	1.23	1.09	1.01	1.02		24.20	22.74	20.76	19.57	18.87	> 100
2	1.44	1.18	1.04	0.98	1.05		31.33	25.11	22.63	21.90	18.76	< 0.01
3	1.38	1.12	0.98	0.90	0.98		39.96	32.34	22.52	21.58	17.66	
4	1.27	1.08	0.97	0.90	0.99		45.46	29.07	24.02	23.95	19.60	
Large	1.01	0.95	0.85	0.78	0.78		42.69	36.55	26.89	20.47	15.34	
$\beta_{\hat{u}^{DIV}}$						$t_{\beta_{\hat{u}^{DIV}}}$						F
Small	4.75	0.43	−5.02	−5.61	−7.88		0.76	0.08	−0.89	−1.10	−1.44	2.33
2	3.38	−4.01	−7.66	−6.76	−6.51		0.76	−0.79	−1.55	−1.35	−1.09	0.02
3	7.45	−1.30	−5.91	−8.27	−9.18		2.34	−0.35	−1.16	−1.53	−1.36	
4	8.65	−5.83	−6.17	−8.18	−11.81		2.90	−1.29	−1.21	−1.72	−2.04	
Large	−0.78	−3.49	−1.73	−9.69	−9.50		−0.29	−1.18	−0.47	−1.83	−1.49	
$\beta_{\hat{u}^{TERM}}$						$t_{\beta_{\hat{u}^{TERM}}}$						F
Small	1.51	1.04	1.69	2.82	8.68		0.26	0.26	0.47	0.79	2.24	1.89
2	−8.21	−2.73	−0.19	1.36	5.16		−1.87	−0.75	0.06	0.46	1.44	0.47
3	−6.34	−3.52	−1.72	2.08	4.39		−1.77	−1.17	−0.55	0.55	1.18	
4	−0.73	−1.51	0.21	0.02	2.13		−0.26	−0.59	0.06	0.01	0.54	
Large	−5.98	−3.26	0.78	−0.90	2.90		−2.22	−1.37	0.31	−0.26	0.74	
$\beta_{\hat{u}^{DEF}}$						$t_{\beta_{\hat{u}^{DEF}}}$						F
Small	−15.45	−14.54	−6.86	−4.79	−8.58		−2.27	−2.17	−1.39	−1.09	−1.68	1.99
2	−10.03	−5.90	−4.78	0.82	−2.20		−2.04	−1.62	−1.37	0.22	−0.49	0.24
3	−11.17	0.22	1.73	4.03	0.81		−2.75	0.08	0.49	1.15	0.18	
4	−5.80	4.81	4.80	8.03	1.08		−2.10	1.92	1.44	2.50	0.25	
Large	−2.45	3.99	9.12	7.25	2.56		−0.96	1.91	3.85	1.91	0.63	
$\beta_{\hat{u}^{RF}}$						$t_{\beta_{\hat{u}^{RF}}}$						F
Small	4.07	−2.58	0.07	1.03	2.77		0.77	−0.50	0.01	0.22	0.56	1.76
2	−4.37	−5.20	−6.25	−4.57	0.97		−1.00	−1.19	−1.60	−1.16	0.20	1.08
3	−7.63	−4.40	−6.53	−4.08	0.71		−2.29	−1.38	−2.07	−1.09	0.15	
4	−3.43	0.47	−2.04	−5.74	−3.71		−1.12	0.16	−0.69	−1.61	−0.90	
Large	−3.55	−0.59	4.81	−0.89	0.30		−1.14	−0.22	1.41	−0.25	0.06	
R^2												
						61.51	60.92	63.41	62.41	59.93		
						73.93	73.95	74.47	71.88	67.96		
						79.81	81.80	77.54	73.58	68.96		
						84.99	86.05	80.32	77.51	69.42		
						87.65	86.11	77.89	67.67	55.88		

loadings on the shocks to the predictive variables. Kan and Zhang [20] emphasize that checking the joint significance of the assets' factor loadings is an important step in detecting useless factors in the cross-section of returns.

Table 6 contains the results for Eq. (17) which correspond to the second pass of the Fama–MacBeth method. The results reveal that the explanatory power of this model is very close to the one for the Fama–French model re-

Financial Economics, The Cross-Section of Stock Returns and the Fama-French Three Factor Model, Table 6

Cross-Sectional Regressions with Loadings on Innovations in State Variables

This table presents Fama–MacBeth cross-sectional regressions using the average excess returns on 25 portfolios sorted by book-to-market and size. The full-sample factor loadings, which are the independent variables in the regressions, are computed in one multiple time-series regression. The coefficients are expressed as percentage per month. The Adjusted R^2 follows Jagannathan and Wang [18] and is reported in percentage form. The first set of t-statistics, indicated by FM t-stat, stands for the Fama–MacBeth estimate. The second set, indicated by SH t-stat, adjusts for errors-in-variables and follows Shanken [33]. The sample period is from July 1963 to December 2001

The Model with R_M and Innovations in DIV, $TERM$, DEF, and RF						
	γ_0	γ_M	$\gamma_{\hat{u}DIV}$	$\gamma_{\hat{u}TERM}$	$\gamma_{\hat{u}DEF}$	$\gamma_{\hat{u}RF}$	Adj. R^2
Estimate	0.64	−0.07	−1.39	4.89	−0.54	−3.22	77.00
FM t-stat	1.74	−0.16	−1.56	4.44	−0.58	−3.79	
SH t-stat	1.08	−0.11	−0.99	2.79	−0.37	−2.40	

Financial Economics, The Cross-Section of Stock Returns and the Fama-French Three Factor Model, Figure 2

Fitted Expected Returns vs. Average Realized Returns for 1963:07-2001:12.

This figure shows realized average returns (%) on the horizontal axis and fitted expected returns (%) on the vertical axis for 25 size and book-to-market sorted portfolios. Each two-digit number represents a separate portfolio. The first digit refers to the size quintile (1 being the smallest and 5 the largest), while the second digit refers to the book-to-market quintile (1 being the lowest and 5 the highest). For each portfolio, the realized average return is the time-series average of the portfolio return and the fitted expected return is the fitted value for the expected return from the corresponding model. The straight line is the 45-degree line from the origin. The Model with the Excess Market Return and Innovations in the Dividend Yield, Term Spread, Default Spread, and Short-Term T-bill

ported previously in Table 2. Figure 2 plots the fitted versus the realized average returns from the model. It can be seen form the graph that the model based on innovation in predictive variables goes a long way toward explaining the value effect: in general, the fitted expected returns on value portfolios (bigger second digit) are higher than the fitted expected returns on growth portfolios (lower second digit). This is consistent with the data on realized average returns for these portfolios. Further, the model with R_M, $\hat{u}^{DIV}$, $\hat{u}^{TERM}$, $\hat{u}^{DEF}$, and $\hat{u}^{RF}$ is more successful at pricing the portfolios that are challenging for the Fama–French model. The realized returns on growth portfolios within the smallest and largest size groups and the value portfolios within the largest size groups are brought closer to the 45-degree line under the model with the four innovations factors.

In summary, this section has shown that the performance of the model based on innovation in predictive variables is very close to the performance of the Fama–French model in the cross-section of average returns sorted by size and book-to-market. This suggest that the Fama–French factors HML and SMB might proxy for fundamental state variables that describe variation in investment opportunities over time.

Other Risk-Based Interpretations

Liew and Vassalou [25] show that there is a relation between the Fama–French portfolios HML and SMB and macroeconomic events. They find that not only in the US but also in several other countries, the corresponding

HML and SMB portfolios contain information about future GDP growth. Therefore, the authors conclude that the size and book-to-market factors are related to future macroeconomic growth. This evidence is consistent with interpreting the HML and SMB factors as proxies for business cycle risk.

Other studies try to relate the difference in average returns between value and growth portfolios to the time-varying nature of the riskiness of those portfolios. Namely, if value stocks are riskier than growth stocks during bad economic times and if the price of bearing risk is higher during those times, then it follows that value stocks should earn higher average returns than growth stocks. Lettau and Ludvigson [24] document that HML is indeed sensitive to bad news in bad macroeconomic times.

Petkova and Zhang [32] is another study that looks are the time-varying risk of value and growth portfolios. They find that the market risk of value stocks is high in bad times when the expected premium for risk is high and it is low

in good times when the expected premium for risk is low. What might lead to this time-varying of value and growth stocks? Zhang [35] suggest that the reason might be irreversible investment. He notes that firms with high book-to-market ratios on average will have larger amounts of tangible capital. In addition, it is more costly for firms to reduce than to expand capital. In bad times, firms want to scale down, especially value firms that are less productive than growth firms (Fama and French [13]). Because scaling down is more difficult, value firms are more adversely affected by economic downturns. In good times, growth firms face less flexibility because they tend to invest more. Expanding is less urgent for value firms because their previously unproductive assets have become more productive. In sum, costly reversibility causes value firms to have higher (lower) betas than growth firms in bad (good) times and this contributes to the return differential between these two classes of stocks.

Future Directions

The Fama–French model states that asset returns are driven by three market-wide factors: the excess return on the market portfolio, and the returns on two portfolios related to size (*SMB*) and book-to-market (*HML*). The *HML* and *SMB* portfolios capture the empirical observation that value firms earn higher average returns than growth firms, and small firms earn higher average returns than large firms. The Fama–French model has been very successful at explaining average stock returns, but the exact economic interpretation of the *HML* and *SMB* portfolios has been an issue of debate.

This article examines the risk-based explanation behind the empirical success of the Fama–French model and suggests that the value and size premia arise due to differences in exposure to systematic sources of risk. As mentioned in the introduction, several authors (e. g., Lakonishok, Shleifer, Vishny [22], La Porta, Lakonishok, Shleifer, Vishny [23]), however, claim that the value premium results from irrationality on the side of investors. Namely, investors tend to over-extrapolate recent stock performance: they overvalue the stocks of growth firms and undervalue the stocks of value firms. When the market realizes its mistake, the prices of the former fall, while the prices of the latter rise, resulting in the value premium.

The Fama–French model provides a useful performance benchmark relative to a set of market-wide factors. The results in this article suggest that the Fama–French factors proxy for systematic sources of risk that capture time variation in investment opportunities. However, the debate about the economic interpretation behind the size

and value premia is still not settled. Whether they arise as a result of rational compensation for risk or irrational investor behavior is still a matter of controversy.

Bibliography

1. Berk J, Green R, Naik V (1999) Optimal investment, growth options, and security returns. J Finance 54:1553–1608
2. Campbell J (1987) Stock returns and the term structure. J Financial Econ 18:373–399
3. Campbell J (1996) Understanding risk and return. J Political Econ 104:298–345
4. Campbell J, Shiller R (1988) Stock prices, earnings, and expected dividends, J Finance 43:661–676
5. Campbell J, Vuolteenaho T (2004) Bad beta, good beta. Am Econ Rev 5:1249–1275
6. Chan KC, Chen N (1991) Structural and return characteristics of small and large firms. J Finance 46:1467–1484
7. Chen N, Zhang F (1998) Risk and return of value stocks. J Bus 71:501–535
8. Cornell B (1999) Risk, duration, and capital budgeting: New evidence on some old questions. J Bus 72:183–200
9. Daniel K, Titman S (1997) Evidence on the characteristics of cross-sectional variation in stock returns. J Finance 52:1–33
10. Fama E, French K (1989) Business conditions and expected returns on stocks and bonds. J Financial Econ 25:23–49
11. Fama E, French K (1992) The cross-section of expected stock returns. J Finance 47:427–465
12. Fama E, French K (1993) Common risk factors in the returns on bonds and stocks. J Financial Econ 33:3–56
13. Fama E, French K (1995) Size and book-to-market factors in earnings and returns. J Finance 50:131–155
14. Fama E, French K (1996) Multifactor explanations of asset pricing anomalies. J Finance 51:55–84
15. Fama E, MacBeth J (1973) Risk, return, and equilibrium: Empirical tests. J Political Econ 81:607–636
16. Fama E, Schwert W (1977) Asset returns and inflation. J Financial Econ 5:115–146
17. Gomes J, Kogan L, Zhang L (2003) Equilibrium cross-section of returns. J Political Econ 111:693–732
18. Jagannathan R, Wang Z (1996) The conditional CAPM and the cross-section of expected returns. J Finance 51:3–53
19. Jagannathan R, Wang Z (1998) Asymptotic theory for estimating beta pricing models using cross-sectional regressions. J Finance 53:1285–1309
20. Kan R, Zhang C (1999) Two-pass tests of asset pricing models with useless factors. J Finance 54:203–235
21. Kothari SP, Shanken J, Sloan R (1995) Another look at the cross-section of expected returns. J Finance 50:185–224
22. Lakonishok J, Shleifer A, Vishny R (1994) Contrarian investment, extrapolation, and risk. J Finance 49:1541–1578
23. La Porta R, Lakonishok J, Shleifer A, Vishny R (1997) Good news for value stocks: Further evidence on market efficiency. J Finance 52:859–874
24. Lettau M, Ludvigson S (2001) Resurrecting the (C)CAPM: A cross-sectional test when risk premia are time-varying. J Political Econ 109:1238–1287
25. Liew J, Vassalou M (2000) Can book-to-market, size, and momentum be risk factors that predict economic growth? J Financial Econ 57:221–245

26. Litterman R, Sheinkman J (1991) Common factors affecting bond returns. J Fixed Income 1:54–61
27. Lo A, MacKinlay C (1990) Data-snooping biases in tests of financial asset pricing models. Rev Financial Stud 3:431–467
28. Long J (1974) Stock prices, inflation, and the term structure of interest rates. J Financial Econ 1:131–170
29. Merton R (1973) An intertemporal capital asset-pricing model. Econometrica 41:867–887
30. Pagan A (1984) Econometric issues in the analysis of regressions with generated regressors. Int Econ Rev 25:221–247
31. Petkova R (2006) Do the Fama–French factors proxy for innovations in predictive variables? J Finance 61:581–612
32. Petkova R, Zhang L (2005) Is value riskier than growth? J Financial Econ 78:187–202
33. Shanken J (1992) On the estimation of beta-pricing models. Rev Financial Stud 5:1–34
34. Vassalou M (2003) News related to future GDP growth as a risk factor in equity returns. J Financial Econ 68:47–73
35. Zhang L (2005) The value premium. J Finance 60:67–103

Financial Economics, Fat-Tailed Distributions

Markus Haas[1], Christian Pigorsch[2]
[1] Department of Statistics, University of Munich, Munich, Germany
[2] Department of Economics, University of Bonn, Bonn, Germany

Article Outline

Glossary

Leptokurtosis A distribution is leptokurtic if it is more peaked in the center and thicker tailed than the normal distribution with the same mean and variance. Occasionally, leptokurtosis is also identified with a moment-based kurtosis measure larger than three, see Sect. "Introduction".

Return Let S_t be the price of a financial asset at time t. Then the *continuous* return, r_t, is $r_t = \log(S_t/S_{t-1})$. The *discrete* return, R_t, is $R_t = S_t/S_{t-1} - 1$. Both are rather similar if $-0.15 < R_t < 0.15$, because $r_t = \log(1 + R_t)$. See Sect. "Introduction".

Tail The (upper) tail, denoted by $\bar{F}(x) = P(X > x)$, characterizes the probability that a random variable X exceeds a certain "large" threshold x. For analytical purposes, "large" is often translated with "as $x \to \infty$". For financial returns, a daily change of 5% is already infinitely large. A Gaussian model essentially excludes such an event.

Tail index The tail index, or *tail exponent*, α, characterizes the rate of tail decay if the tail goes to zero, in essence, like a power function, i. e., $\bar{F}(x) = x^{-\alpha} L(x)$, where L is slowly varying. Moments of order lower (higher) than α are (in)finite.

Definition of the Subject

Have a look at Fig. 1. The top plot shows the daily percentage changes, or *returns*, of the S&P500 index ranging from January 2, 1985 to December 29, 2006, a total of 5,550 daily observations. We will use this data set throughout the article to illustrate some of the concepts and models to be discussed. Two observations are immediate. The first is that both small and large changes come clustered, i. e., there are periods of low and high volatility. The second is that, from time to time, we observe rather large changes which may be hard to reconcile with *the* standard distributional assumption in statistics and econometrics, that is, normality. The most outstanding return certainly occurred on October 19, 1987, the "Black Monday", where the index lost more than 20% of its value, but the phenomenon is chronic. For example, if we fit a normal distribution to the data, the resulting model predicts that we observe an absolute daily change larger than 5% once in approximately 1,860 years, whereas we actually encountered that 13 times during our 22-year sample period. This suggests that, compared to the normal distribution, the distribution of the returns is *fat-tailed*, i. e., the probability of large losses and gains is much higher than would be implied by a time-invariant unconditional Gaussian distribution. The latter is obviously not suitable for describing the booms, busts, bubbles, and bursts of activity which characterize financial markets, and which are apparent in Fig. 1.

The two aforementioned phenomena, i. e., volatility clustering and fat tails, have been detected in almost every financial return series that was subject to statistical analysis since the publication of Mandelbrot's [155] seminal study of cotton price changes, and they are of paramount importance for any individual or institution engaging in the financial markets, as well as for financial economists trying to understand their mode of operation. For exam-
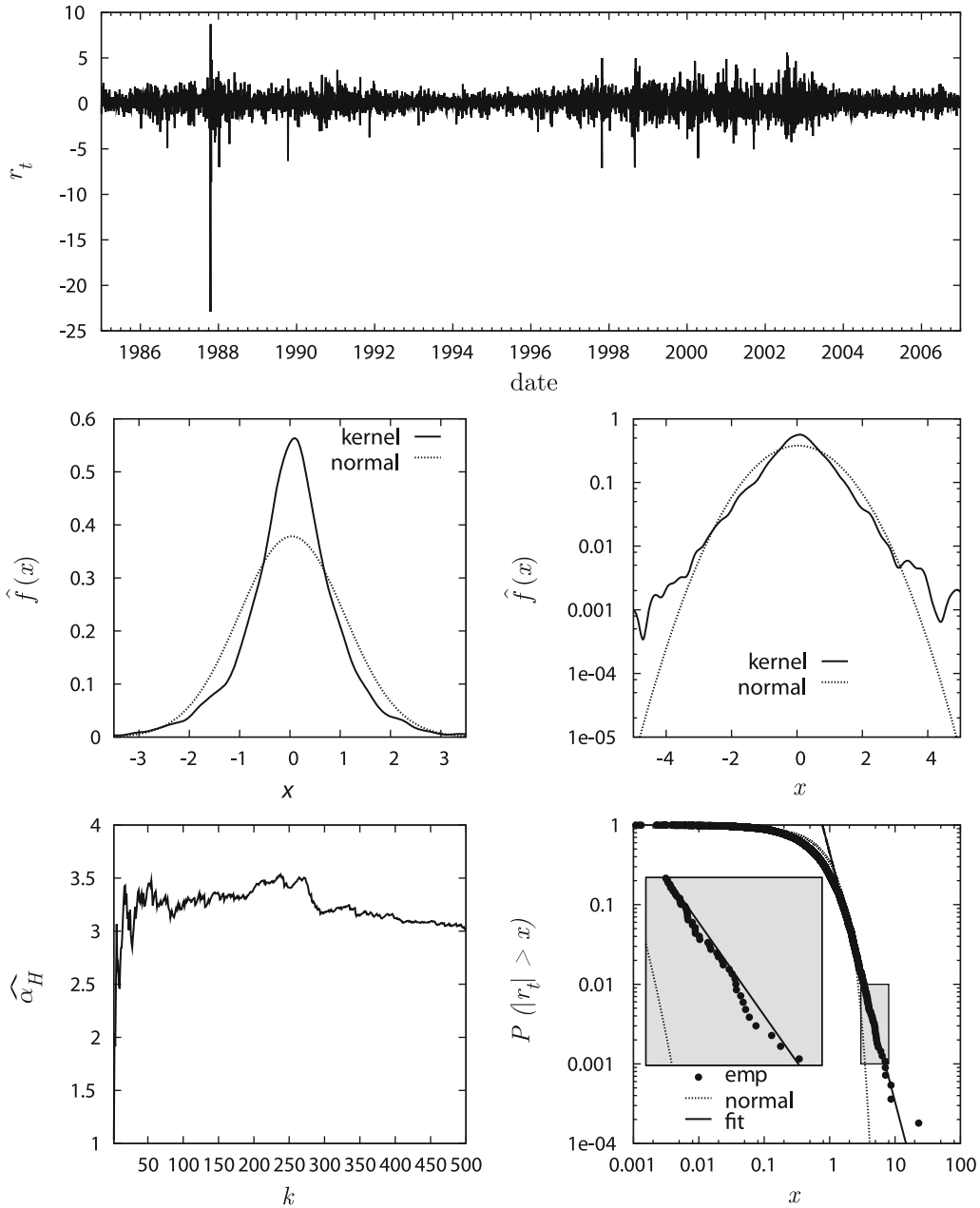
Financial Economics, Fat-Tailed Distributions, Figure 1
The *top plot* shows the S&P500 *percentage* returns, r_t, from January 1985 to December 2006, i. e., $r_t = 100 \times \log(S_t/S_{t-1})$, where S_t is the index level at time t. The left plot of the *middle panel* shows a nonparametric density estimate (*solid*), along with the fitted normal density (*dotted*); the right graph is similar but shows the respective log-densities in order to better visualize the tail regions. The *bottom left plot* represents a *Hill plot* for the S&P500 returns, i. e., it displays $\hat{\alpha}_{k,n}$ defined in (11) for $k \leq 500$. The *bottom right plot* shows the complementary cdf, $\bar{F}(x)$, on a log-log scale, see Sect. "Empirical Evidence About the Tails" for discussion

ple, investors holding significant portions of their wealth in risky assets need a realistic assessment of the likelihood of severe losses. Similarly, economists trying to learn about the relation between risk and return, the pricing of finan-

cial derivatives, such as options, and the inherent dynamics of financial markets, can only benefit from building their models on adequate assumptions about the stochastic properties of the variables under study, and they have

to reconcile the predictions of their models with the actual facts.

This article reviews some of the most important concepts and distributional models that are used in empirical finance to capture the (almost) ubiquitous stochastic properties of returns as indicated above. Section "Introduction" defines in a somewhat more precise manner than above the central variable of interest, the return of a financial asset, and gives a brief account of the early history of the problem. Section "Defining Fat-Tailedness" discusses various operationalizations of the term "fat-tailedness", and Sect. "Empirical Evidence About the Tails" summarizes what is or is at least widely believed to be known about the tail characteristics of typical return distributions. Popular parametric distributional models are discussed in Sect. "Some Specific Distributions". The alpha stable model as the archetype of a fat-tailed distribution in finance is considered in detail, as is the generalized hyperbolic distribution, which provides a convenient framework for discussing, as special or limiting cases, many of the important distributions employed in the literature. An empirical comparison using the S&P500 returns is also included. In Sect. "Volatility Clustering and Fat Tails", the relation between the two "stylized facts" mentioned above, i. e., clusters of volatility and fatness of the tails, is highlighted, where we concentrate on the GARCH approach, which has gained outstanding popularity among financial econometricians. This model has the intriguing property of producing fat-tailed marginal distributions even with light-tailed innovation processes, thus emphasizing the role of the market dynamics. In Sect. "Application to Value-at-Risk", we compare both the unconditional parametric distributional models introduced in Sect. "Some Specific Distributions" as well as the GARCH model of Sect. "Volatility Clustering and Fat Tails" on an economic basis by evaluating their ability to accurately measure the Value-at-Risk, which is an important tool in risk management. Finally, Sect. "Future Directions" identifies some open issues.

Introduction

To fix notation, let S_t be the price of an asset at time t, e. g., a stock, a market index, or an exchange rate. The *continuously compounded* or *log* return from time t to time $t + \Delta t$, $r_{t,t+\Delta t}$, is then defined as

$$r_{t,t+\Delta t} = \log S_{t+\Delta t} - \log S_t . \tag{1}$$

Often the quantity defined in (1) is also multiplied by 100, so that it can be interpreted in terms of *percentage returns*, see Fig. 1. Moreover, in applications, Δt is usually set equal

to one and represents the horizon over which the returns are calculated, e. g., a day, week, or month. In this case, we drop the first subscript and define $r_t := \log S_t - \log S_{t-1}$. The log returns (1) can be additively aggregated over time, i. e.,

$$r_{t,t+\tau} = \sum_{i=1}^{\tau} r_{t+i} . \tag{2}$$

Empirical work on the distribution of financial returns is usually based on log returns. In some applications a useful fact is that, over short intervals of time, when returns tend to be small, (1) can also serve as a reasonable approximation to the *discrete* return, $R_{t,t+\Delta t} := S_{t+\Delta t}/S_t - 1 = \exp(r_{t,t+\Delta t}) - 1$. For further discussion of the relationship between continuous and discrete returns and their respective advantages and disadvantages, see, e. g., [46,76].

The seminal work of Mandelbrot [155], to be discussed in Subsect. "Alpha Stable and Related Distributions", is often viewed as the beginning of modern empirical finance. As reported in [74], "[p]rior to the work of Mandelbrot the usual assumption ... was that the distribution of price changes in a speculative series is approximately Gaussian or normal". The rationale behind this prevalent view, which was explicitly put forward as early as 1900 by Bachelier [14], was clearly set out in [178]: If the logprice changes (1) from transaction to transaction are independently and identically distributed with finite variance, and if the number of transactions is fairly uniformly distributed in time, then (2) along with the central limit theorem (CLT) implies that the return distribution over longer intervals, such as a day, a week, or a month, approaches a Gaussian shape.

However, it is now generally acknowledged that the distribution of financial returns over horizons shorter than a month is not well described by a normal distribution. In particular, the empirical return distributions, while unimodal and approximately symmetric, are typically found to exhibit considerable *leptokurtosis*, i. e., they are more peaked in the center and have fatter tails than the Gaussian with the same variance. Although this has been occasionally observed in the pre-Mandelbrot literature (e. g., [6]), the first systematic account of this phenomenon appeared in [155] and the follow-up papers by Fama [74,75] and Mandelbrot [156], and it was consistently confirmed since then. The typical shape of the return distribution, as compared to a fitted Gaussian, is illustrated in the middle panel of Fig. 1 for the S&P500 index returns, where a nonparametric kernel density estimator (e. g., [198]) is superimposed on the fitted Gaussian curve (dashed line). Interestingly, this pattern has been detected not only for modern

financial markets but also for those of the eighteenth century [103].

The (location and scale-free) standardized fourth moment, or coefficient of *kurtosis*,

$$\mathbb{K}[X] = \frac{\mathbb{E}\left[(X - \mu)^4\right]}{\sigma^4}, \tag{3}$$

where μ and σ are the mean and the standard deviation of the random variable (rv) X, respectively, is sometimes used to assess the degree of leptokurtosis of a given distribution. For the normal distribution, $\mathbb{K} = 3$, and $\mathbb{K} > 3$, referred to as *excess kurtosis*, is taken as an indicator of a leptokurtic shape (e. g., [164], p. 429). For example, the sample analogue of (3) for the S&P500 returns shown in Fig. 1 is 47.9, indicating very strong excess kurtosis. A formal test could be conducted using the fact that, under normality, the sample kurtosis is asymptotically normal with mean 3 and standard deviation $\sqrt{24/T}$ (T being the sample size), but the result can be anticipated.

As is well-known, however, such moment-based summary measures have to be interpreted with care, because a particular moment need not be very informative about a density's shape. We know from Finucan [82] that if two symmetric densities, f and g, have common mean and variance and finite fourth moment, and if g is more peaked and has thicker tails than f, *then* the fourth moment (and hence $\mathbb{K}$) is greater for g than for f, provided the densities cross exactly twice on both sides of the mean. However, the converse of this statement is, of course, not true, and a couple of (mostly somewhat artificial) counterexamples can be found in [16,68,121]. [158] provides some intuition by relating density crossings to moment crossings. For example, (only) if the densities cross more than four times, it may happen that the fourth moment is greater for f, but the sixth and all higher moments are greater for g, reflecting the thicker tails of the latter. Nevertheless, Finucan's result, along with his (in some respects justified) hope that we can view "this pattern as the common explanation of a high observed kurtosis", may serve to argue for a certain degree of usefulness of the kurtosis measure (3), provided the fourth moment is assumed to be finite. However, a nonparametric density estimate will in any case be more informative. Note that the density crossing condition in Finucan's theorem is satisfied for the S&P500 returns in Fig. 1.

Defining Fat-Tailedness

The notion of leptokurtosis as discussed so far is rather vague, and both financial market researchers as well as practitioners, such as risk managers, are interested in a more precise description of the tail behavior of financial variables, i. e., the laws governing the probability of large gains and losses. To this end, we define the *upper tail* of the distribution of a rv X as

$$\bar{F}(x) = P(X > x) = 1 - F(x), \tag{4}$$

where F is the cumulative distribution function (cdf) of X. Consideration of the upper tail is the standard convention in the literature, but it is clear that everything could be phrased just as well in terms of the lower tail.

We are interested in the behavior of (4) as x becomes large. For our benchmark, i. e., the normal distribution with (standardized) density (pdf) $\phi(x) = (2\pi)^{-1/2}$ $\exp(-x^2/2)$, we have (cf. p. 131 in [79])

$$\bar{F}(x) \cong \frac{1}{\sqrt{2\pi} x} \exp\left(-\frac{x^2}{2}\right) = \frac{\phi(x)}{x} \quad \text{as } x \to \infty, \tag{5}$$

where the notation $f(x) \cong g(x)$ as $x \to \infty$ means that $\lim_{x \to \infty} f(x)/g(x) = 1$. Thus, the tails of the normal tend to zero faster than exponentially, establishing its very light tails.

To appreciate the difference between the general concept of leptokurtosis and the approach that focuses on the tails, consider the class of finite normal mixtures as discussed in Subsect. "Finite Mixtures of Normal Distributions". These are leptokurtic in the sense of peakedness and tailedness (compared to the normal), but are light-tailed according to the tail-based perspective.

While it is universally accepted in the literature that the Gaussian is too light-tailed to be an appropriate model for the distribution of financial returns, there is no complete agreement with respect to the actual shape of the tails. This is not surprising because we cannot reasonably expect to find a model that accurately fits all markets at any time and place. However, the current mainstream opinion is that the probability for the occurrence of large (positive and negative) returns can often appropriately be described by Pareto-type tails. Such tail behavior is also frequently adopted as the definition of fat-tailedness per se, but the terminology in the literature is by no means unique.

A distribution has Pareto-type tails if they decay essentially like a power function as x becomes large, i. e., $\bar{F}$ is regularly varying (at infinity) with index $-\alpha$ (written $\bar{F} \in \text{RV}_{-\alpha}$), meaning that

$$\bar{F}(x) = x^{-\alpha} L(x), \quad \alpha > 0, \tag{6}$$

where $L > 0$ is a slowly varying function, which can be interpreted as "slower than any power function" (see [34, 188,195] for a technical treatment of regular variation). The defining property of a slowly varying function is $\lim_{x \to \infty} L(tx)/L(x) = 1$ for any $t > 0$, and the aforemen-

tioned interpretation follows from the fact that, for any $\gamma > 0$, we have (cf. [195], p. 18)

$$\lim_{x\to\infty} x^{\gamma} L(x) = \infty, \quad \lim_{x\to\infty} x^{-\gamma} L(x) = 0. \quad (7)$$

Thus, for large x, the parameter α in (6), called the *tail index* or *tail exponent*, controls the rate of tail decay and provides a measure for the fatness of the tails.

Typical examples of slowly varying functions include $L(x) = c$, a constant, $L(x) = c + o(1)$, or $L(x) = (\log x)^k$, $x > 1$, $k \in \mathbb{R}$. The first case corresponds to strict Pareto tails, while in the second the tails are asymptotically Paretian in the sense that $\bar{F}(x) \cong cx^{-\alpha}$, which includes as important examples in finance the (non-normal) stable Paretian (see (13) in Subsect. "Alpha Stable and Related Distributions") and the Student's t distribution considered in Sect. "The Student t Distribution", where the tail index coincides with the characteristic exponent and the number of degrees of freedom, respectively. As an example for both, the Cauchy distribution with density $f(x) = [\pi(1 + x^2)]^{-1}$ has cdf $F(x) = 0.5 + \pi^{-1}\arctan(x)$. As $\arctan(x) = \sum_0^\infty (-1)^i x^{2i+1}/(2i + 1)$ for $|x| < 1$, and $\arctan(x) = \pi/2 - \arctan(1/x)$ for $x > 0$, we have $\bar{F}(x) \cong (\pi x)^{-1}$.

For the distributions mentioned in the previous paragraph, it is known that their moments exist only up to (and excluding) their tail indices, α. This is generally true for rvs with regularly varying tails and follows from (7) along with the well-known connection between moments and tail probabilities, i. e., for a non-negative rv X, and $r > 0$, $\mathbb{E}[X^r] = r\int_0^\infty x^{r-1}\bar{F}(x)\mathrm{d}x$ (cf. [95], p. 75). The only possible minor variation is that, depending on L, $\mathbb{E}[X^\alpha]$ may be finite. For example, a rv X with tail $\bar{F}(x) = cx^{-1}(\log x)^{-2}$ has finite mean. The property that moments greater than α do not exist provides further intuition for α as a measure of tail-fatness.

As indicated above, there is no universal agreement in the literature with respect to the definition of fat-tailedness. For example, some authors (e. g., [72,196]) emphasize the class of *subexponential* distributions, which are (although not exclusively) characterized by the property that their tails tend to zero slower than any exponential, i. e., for any $\gamma > 0$, $\lim_{x\to\infty} e^{\gamma x}\bar{F}(x) = \infty$, implying that the moment generating function does not exist. Clearly a regularly varying distribution is also subexponential, but further members of this class are, for instance, the lognormal as well as the *stretched exponential*, or heavy-tailed Weibull, which has a tail of the form

$$\bar{F}(x) = \exp\left(-x^b\right), \quad 0 < b < 1. \quad (8)$$

The stretched exponential has recently been considered

by [134,152,153] as an alternative to the Pareto-type distribution (6) for modeling the tails of asset returns. Note that, as opposed to (6), both the lognormal as well as the stretched exponential possess power moments of all orders, although no exponential moment.

In addition, [22] coined the term *semi-heavy tails* for the generalized hyperbolic (GH) distribution, but the label may be employed more generally to refer to distributions with slower tails than the normal but existing moment generating function. The GH, which is now very popular in finance and nests many interesting special cases, will be examined in detail in Subsect. "The Generalized Hyperbolic Distribution".

As will be discussed in Sect. "Empirical Evidence About the Tails", results of extreme value theory (EVT) are often employed to identify the tail shape of return distributions. This has the advantage that it allows one to concentrate fully on the tail behavior, without the need to model the central part of the distribution. To sketch the idea behind this approach, suppose we attempt to classify distributions according to the limiting behavior of their normalized maxima. To this end, let $\{X_i, i \geq 1\}$ be an iid sequence of rvs with common cdf F, $M_n = \max\{X_1, \ldots, X_n\}$, and assume there exist sequences $a_n > 0$, $b_n \in \mathbb{R}$, $n \geq 1$, such that

$$P\left(\frac{M_n - b_n}{a_n} \leq x\right) = F^n(a_n x + b_n) \xrightarrow{n\to\infty} G(x), \quad (9)$$

where G is assumed nondegenerate. To see that normalization is necessary, note that $\lim_{n\to\infty} P(M_n \leq x) = \lim_{n\to\infty} F^n(x) = 0$ for all $x < x_M := \sup\{x : F(x) < 1\} \leq \infty$, so that the limiting distribution is degenerate and of little help. If the above assumptions are satisfied, then, according to the classical Fisher–Tippett theorem of extreme value theory (cf. [188]), the limiting distribution G in (9) must be of the following form:

$$G_\xi(x) = \exp\left(-(1 + \xi x)^{-1/\xi}\right), \quad 1 + \xi x > 0, \quad (10)$$

which is known as the *generalized extreme value distribution* (GEV), or *von Mises representation of the extreme value distributions* (EV). The latter term can be explained by the fact that (10) actually nests three different types of EV distributions, namely

(i)   the Fréchet distribution, denoted by G_ξ^+, where $\xi > 0$ and $x > -1/\xi$,

(i)   the so-called Weibull distribution of EVT, denoted by G_ξ^-, where $\xi < 0$ and $x < -1/\xi$, and

(iii) the Gumbel distribution, denoted by G_0, which corresponds to the limiting case as $\xi \to 0$, i. e., $G_0(x) = \exp(-\exp(-x))$, where $x \in \mathbb{R}$.

A cdf F belongs to the *maximum domain of attraction* (MDA) of one of the extreme value distributions nested in (10), written $F \in \text{MDA}\left(G_\xi\right)$, if (9) holds, i. e., classifying distributions according to the limiting behavior of their extrema amounts to figuring out the MDAs of the extreme value distributions. It turns out that it is the tail behavior of a distribution F that accounts for the MDA it belongs to. In particular, $F \in \text{MDA}(G_\xi^+)$ if and only if its tail $\bar{F} \in \text{RV}_{-\alpha}$, where $\alpha = 1/\xi$. As an example, for a strict Pareto distribution, i. e., $F(x) = 1 - (u/x)^\alpha$, $x \geq u > 0$, with $a_n = n^{1/\alpha} u/\alpha$ and $b_n = n^{1/\alpha} u$, we have $\lim_{n \to \infty} F^n(a_n x + b_n) = \lim_{n \to \infty}(1 - n^{-1}(1 + x/\alpha)^{-\alpha})^n = G_{1/\alpha}^+(x)$. Distributions in $\text{MDA}(G_\xi^-)$ have a finite right endpoint, while, roughly, most of the remaining distributions, such as the normal, the lognormal and (stretched) exponentials, belong to $\text{MDA}(G_0)$. The latter also accommodates a few distributions with finite right endpoint. See [188] for precise conditions. The important case of non-iid rvs is discussed in [136]. A central result is that, rather generally, vis-à-vis an iid sequence with the same marginal cdf, the maxima of stationary sequences converge to the same type of limiting distribution. See [63,167] for an application of this theory to ARCH(1) and GARCH(1,1) processes (see Sect. "Volatility Clustering and Fat Tails"), respectively.

One approach to exploit the above results, referred to as the *method of block maxima*, is to divide a given sample of return data into subsamples of equal length, pick the maximum of each subsample, assume that these have been generated by (10) (enriched with location and scale parameters to account for the unknown a_n and b_n), and find the maximum-likelihood estimate for ξ, location, and scale. Standard tests can then be conducted to assess, e. g., whether $\xi > 0$, i. e., the return distribution has Pareto-type tails. An alternative but related approach, which is based on further theoretical developments and often makes more efficient use of the data, is the *peaks over thresholds* (POT) method. See [72] for a critical discussion of these and alternative techniques.

We finally note that $1 - G_{1/\alpha}^+ (\alpha (x - 1)) \cong x^{-\alpha}$, while $1 - G_0(x) \cong \exp(-x)$, i. e., for the extremes, we have asymptotically a Pareto and an exponential tail, respectively. This may provide, on a meta-level, a certain rationale for reserving the notion of genuine fat-tailedness for the distributions with regularly varying tails.

Empirical Evidence About the Tails

The first application of power tails in finance appeared in Mandelbrot's [155] study of the log-price changes of cotton. Mandelbrot proposed to model returns with non-normal *alpha stable*, or *stable Paretian*, distributions, the properties of which will be discussed in some detail in Subsect. "Alpha Stable and Related Distributions". For the present discussion, it suffices to note that for this model the tail index α in (6), also referred to as *characteristic exponent* in the context of stable distributions, is restricted to the range $0 < \alpha < 2$, and that much of its theoretical appeal derives from the fact that, due to the generalized CLT, "Mandelbrot's hypothesis can actually be viewed as a generalization of the central-limit theorem arguments of Bachelier and Osborne to the case where the underlying distributions of price changes from transaction to transaction … have infinite variances" [75]. For the cotton price changes, Mandelbrot came up with a tail index of about 1.7, and his work was subsequently complemented by Fama [75] with an analysis of daily returns of the stocks belonging to the Dow Jones Industrial Average. [75] came to the conclusion that Mandelbrot's theory was supported by these data, with an average estimated α close to 1.9.

The findings of Mandelbrot and Fama initiated an extensive discussion about the appropriate distributional model for stock returns, partly because the stable model's implication that the tails are so fat that even the variance is infinite appeared to be too radical to many economists used to working with models built on the assumption of finite second moments. The evidence concerning the stable hypothesis gathered in the course of the debate until the end of the 1980s was not ultimately conclusive, but there were many papers reporting mainly negative results [4,28, 36,40,54,67,98,99,109,135,176,180,184].

From the beginning of the 1990s, a number of researchers have attempted to estimate the tail behavior of asset returns directly, i. e., without making specific assumptions about the entire distributional shape. [86,115,142,143] use the method of block maxima (see Sect. "Defining Fat-Tailedness") to identify the maximum domain of attraction of the distribution of stock returns. They conclude that the Fréchet distribution with a tail index $\alpha > 2$ is most likely, implying Pareto-type tails which are thinner than those of the *stable* Paretian.

A second strand of literature assumes a priori the presence of a Pareto-type tail and focuses on the estimation of the tail index α. If, as is often the case, a power tail is deemed adequate, an explicit estimate of α is of great interest both from a practical and an academic viewpoint. For example, investors want to assess the likelihood of large losses of financial assets. This is often done using methods of extreme value theory, which require an accurate estimate of the tail exponent. Such estimates are also important because the properties of statistical tests and other quantities of interest, such as empirical autocorrelation

functions, frequently depend on certain moment conditions (e. g., [144,167]). Clearly the desire to figure out the actual tail shape has also an intrinsic component, as is reflected in the long-standing debate on the stable Paretian hypothesis. People simply wanted to know whether this distribution, with its appealing theoretical properties, is consistent with actual data. Moreover, empirical findings may guide economic theorizing, as they can help both in assessing the validity of certain existing models as well as in suggesting new explanations. Two examples will briefly be discussed at the end of the present section.

Within this second strand of literature, the Hill estimator [106] has become the most popular tool. It is given by

$$\hat{\alpha}_{k,n} = \left(\frac{1}{k-1} \sum_{j=1}^{k-1} \log X_{j,n} - \log X_{k,n} \right)^{-1}, \qquad (11)$$

where $X_{i,n}$ denotes the ith upper order statistic of a sample of length n, i. e., $X_{1,n} \geq X_{2,n} \geq \cdots \geq X_{n,n}$. See [64,72] for various approaches to deriving (11). If the tail is not regularly varying, the Hill estimator does not estimate anything.

A crucial choice to be made when using (11) is the selection of the threshold value k, i. e., the number of order statistics to be included in the estimation. Ideally, only observations from the tail region should be used, but choosing k to small will reduce the precision of the estimator. There exist various methods for picking k optimally in a mean-squared error sense [61,62], but much can be learned by looking at the *Hill plot*, which is obtained by plotting $\hat{\alpha}_{k,n}$ against k. If we can find a range of k-values where the estimate is approximately constant, this can be taken as a hint for where the "true" tail index may be located. As illustrated in [189], however, the Hill plot may not always be so well-behaved, and in this case the semi-automatic methods mentioned above will presumably also be of little help.

The theoretical properties of (11), along with technical conditions, are summarized in [72,189]. Briefly, for iid data generated from a distribution with tail $\bar{F} \in \text{RV}_{-\alpha}$, the Hill estimator has been shown to be consistent [159] and asymptotically normal with standard deviation $\alpha/\sqrt{k}$ [100]. Financial data, however, are usually not iid but exhibit considerable dependencies in higher-order moments (see Sect. "Volatility Clustering and Fat Tails"). In this situation, i. e., with ARCH-type dynamics, (11) will still be consistent [190], but little is known about its asymptotic variance. However, simulations conducted in [123] with an IGARCH model, as defined in Sect. "Volatility Clustering and Fat Tails", indicate that,

under such dependencies, the actual standard errors may be seven to eight times larger than those implied by the asymptotic theory for iid variables.

The Hill estimator was introduced into the econometrics literature in the series of articles [107,113,125,126]. [125,126], using weekly observations, compare the tails of exchange rate returns in floating and fixed exchange rate systems, such as the Bretton Woods period and the EMS. They find that for the fixed systems, most tail index estimates are below 2, i. e., consistent with the alpha stable hypothesis, while the estimates are significantly larger than 2 (ranging approximately from 2.5 to 4) for the float. [126] interpret these results in the sense that "a float lets exchange rates adjust more smoothly than any other regime that involves some amount of fixity". Subsequent studies of floating exchange rates using data ranging from weekly [107,110,111] over daily [58,89,144] to very high-frequency [59,61,170] have confirmed the finding of these early papers that the tails are not fat enough to be compatible with the stable Paretian hypothesis, with estimated tail indices usually somewhere in the region 2.5–5. [58] is the first to investigate the tail behavior of the euro against the US dollar, and finds that it is similar both to the German mark in the pre-euro era as well as to the yen and the British pound, with estimated exponents hovering around 3–3.5.

Concerning estimation with data at different time scales, a comparison of the results reported in the literature reveals that the impact on the estimated tail indices appears to be moderate. [59] observe an increase in the estimates when moving from 30-minute to daily returns, but they argue that these changes, due to the greater bias at the lower frequencies, are small enough to be consistent with α being invariant under time aggregation.

Note that if returns were independently distributed, their tail behavior would in fact be unaffected by time aggregation. This is a consequence of (2) along with Feller's (p. 278 in [80]) theorem on the convolution of regularly varying distributions, stating that any finite convolution of a regularly varying cdf $F(x)$ has a regularly varying tail with the same index. Thus, in principle, the tail survives forever, but, as long as the variance is finite, the CLT ensures that in the course of aggregation an increasing probability weight is allocated to the center of the distribution, which becomes closer to a Gaussian shape. The probability of observing a tail event will thus decrease. However, for fat-tailed distributions, the convergence to normality can be rather slow, as reflected in the observation that pronounced non-normalities in financial returns are often observed even at a weekly and (occasionally) monthly frequency. See [41] for an informative discussion of these is-

sues. The fact that returns are, in general, not iid makes the interpretation of the approximate stability of the tail index estimates observed across papers employing different frequencies not so clear-cut, but Feller's theorem may nevertheless provide some insight.

There is also an extensive literature reporting tail index estimates of stock returns, mostly based on daily [2,89,92, 112,113,144,145,146,177] and higher frequencies [2,91,92, 147,181]. The results are comparable to those for floating exchange rates in that the tenor of this literature, which as a whole covers all major stock markets, is that most stock return series are characterized by a tail index somewhere in the region 2.5–5, and most often below 4. That is, the tails are thinner than expected under the stable Paretian hypothesis, but the finiteness of the third and in particular the fourth moments (and hence kurtosis) may already be questionable. Again, the results appear to be relatively insensitive with respect to the frequency of the data, indicating a rather slow convergence to the normal distribution. Moreover, most authors do not find significant differences between the left and the right tail, although, for stock returns, the point estimates tend to be somewhat lower for the left tail (e. g., [115,145]).

Applications to the bond market appear to be rare, but see [201], who report tail index estimates between 2.5 and 4.5 for 5-minute and 1-hour Bund future returns and somewhat higher values for daily data. [160] compare the tail behaviors of spot and future prices of various commodities (including cotton) and find that, while future prices resemble stock prices with tail indices in the region 2.5–4, spot prices are somewhat fatter tailed with α hovering around 2.5 and, occasionally, smaller than 2.

Summarizing, it is now a widely held view that the distribution of asset returns can typically be described as fat-tailed in the power law sense but with finite variance. Thus, currently there seems to exist a far reaching consensus that the stable Paretian model is not adequate for financial data, but see [162,202] for a different viewpoint. A consequence of the prevalent view is that asset return distributions belong to the Gaussian domain of attraction, but that the convergence appears to be very slow.

To illustrate typical findings as reported above, let us consider the S&P500 returns described in Sect. "Definition of the Subject". A first informal check of the appropriateness of a power law can be obtained by means of a log-log plot of the empirical tail, i. e., if $1 - F(x) = \bar{F}(x) \approx cx^{-\alpha}$ for large x, then a plot of the log of the empirical complementary cdf, $\bar{F}(x)$, against $\log x$ should display a linear behavior in its outer part. For the data at hand, such a plot is shown in the bottom right panel of Fig. 1. Assuming homogeneity across the tails, we pool negative and positive

returns by first removing the sample mean and then taking absolute values. We have also multiplied (1) by 100, so that the returns are interpretable in terms of percentage changes. The plot suggests that a power law regime may be starting from approximately the 90% quantile. Included in Fig. 1 is also a regression line ("fit") fitted to the log-tail using the 500 upper (absolute) return observations. This yields, as a rough estimate for the tail index, a slope of $\hat{\alpha} = 3.13$, with a coefficient of determination $R^2 = 0.99$. A Hill plot for $k \le 500$ in (11) is shown in the bottom left panel of Fig. 1. The estimates are rather stable over the entire region and suggest an α somewhere in the interval $(3, 3.5)$, which is reconcilable with the results in the literature summarized above. A somewhat broader picture can be obtained by considering individual stocks. Here we consider the 176 stocks that were listed in the S&P500 from January 1985 to December 2006. Figure 2 presents, for each $k \le 500$, the 5%, 50%, and 95% quantiles of the distribution of (11) over the different stocks. The median is close to 3 throughout, and it appears that an estimate in $(2.5, 4.5)$ would be reasonable for most stocks.

At this point, it may be useful to note that the issue is not whether a power law is true in the strict sense but only if it provides a reasonable approximation in the relevant tail region. For example, it might be argued that asset returns actually have finite support, implying finiteness of all moments and hence inappropriateness of a Pareto-type tail. However, as concisely pointed out in [144], "saying that the support of an empirical distribution is bounded says very little about the nature of outlier activity that may occur in the data".

We clearly cannot expect to identify the "true" distribution of financial variables. For example, [153] have demonstrated that by standard techniques of EVT it is virtually impossible, even in rather large samples, to discriminate between a power law and a stretched exponential (8) with a small value of b, thus questioning, for example, the conclusiveness of studies relying on the block maxima method, as referred to above. A similar point was made in [137], who showed by simulation that a three-factor stochastic volatility model, with a marginal distribution known to have all its moments finite, can generate apparent power laws in practically relevant sample sizes. As put forward in [152], "for most practical applications, the relevant question is not to determine what is the true asymptotic tail, but what is the best effective description of the tails in the domain of useful applications".

As is evident in Fig. 1, a power law may (and often does) provide a useful approximation to the tail behavior of actual data, but there is no reason to expect that it
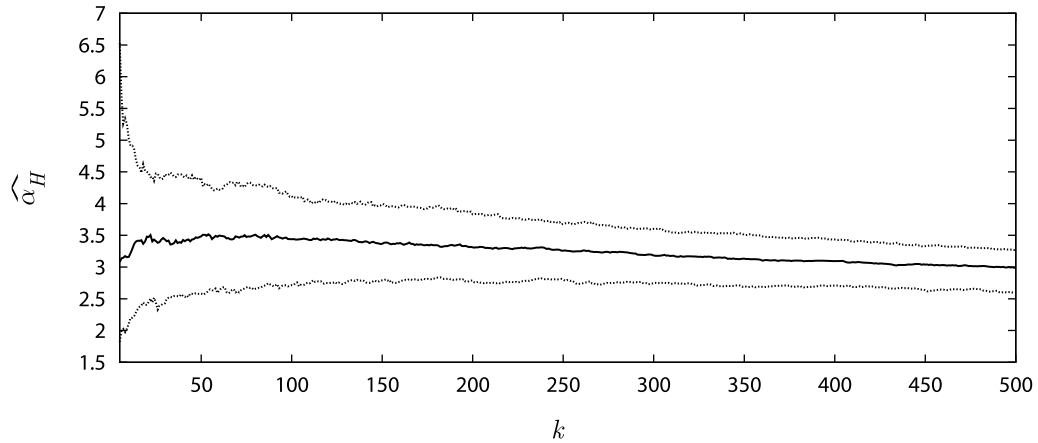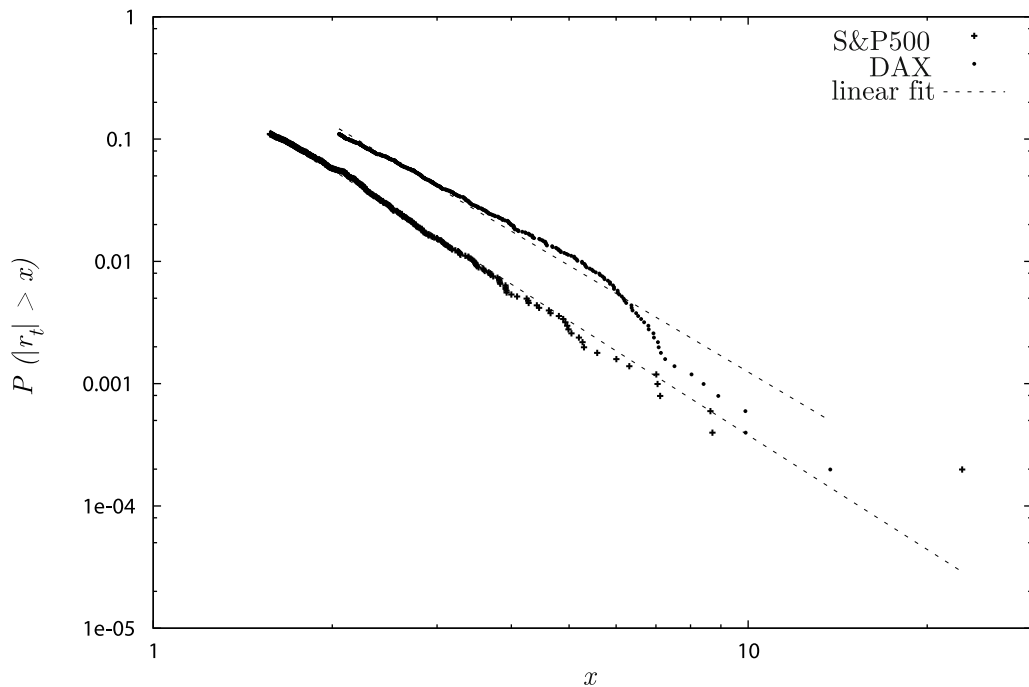
Financial Economics, Fat-Tailed Distributions, Figure 2
Shown are, for $k \leq 500$, the 95%, 50%, and 5% quantiles of the distribution of the Hill estimator $\hat{\alpha}_{k,n}$, as defined in (11), over 176 stocks included in the S&P500 stock index

will appear in every market, and a broad range of heavy and semi-heavy tailed distributions (such as the GH in Subsect. "The Generalized Hyperbolic Distribution") may provide an adequate fit. For instance, [93] investigate the tail behavior of high-frequency returns of one of the most frequently traded stocks on the Paris Stock Exchange (Al-catel) and conclude that the tails decay at an exponential rate, and [119,197] obtain similar results for daily returns of the Nikkei 225 index and various individual US stocks, respectively. As a further illustration, without rigorous statistical testing, Fig. 3 shows the log-log tail plot for daily returns of the German stock market index DAX

Financial Economics, Fat-Tailed Distributions, Figure 3
The figure shows, on a log-log scale, the complementary cdf, $\bar{F}(x)$, for the largest 500 absolute return observations both for the daily S&P500 returns from January 1985 to December 2006 and the daily DAX returns from July 1987 to July 2007

from July 3, 1987 to July 4, 2007 for the largest 500 out of 5,218 (absolute) return observations, along with a regression-based linear fit. For purposes of comparison, the corresponding figure for the S&P500 has also been reproduced from Fig. 1. While the slopes of the fitted power laws exhibit an astonishing similarity (in fact, the estimated tail index of the DAX is 2.93), it is clear from Fig. 3 that an asymptotic power law, although not necessarily inconsistent with the data, is much less self-evident for the DAX than for the S&P500, due to the apparent curvature particularly in the more extreme tail.

It is finally worthwhile to mention that financial theory in general, although some of its models are built on the *assumption* of a specific distribution, has little to say about the distribution of financial variables. For example, according to the efficient markets paradigm, asset prices change in response to the arrival of relevant new information, and, consequently, the distributional properties of returns will essentially reflect those of the news process. As noted by [148], an exception to this rule is the model of rational bubbles of [35]. [148] show that this class of processes gives rise to an (approximate) power law for the return distribution. However, the structure of the model, i. e., the assumption of rational expectations, restricts the tail exponent to be below unity, which is incompatible with observed tail behaviors.

More recently, prompted by the observation that estimated tail indices are often located in a relatively narrow interval around 3, [83,84,85] have developed a model to explain a hypothesized "inverse cubic law for for the distribution of stock price variations" [91], valid for highly developed economies, i. e., a power law tail with index $\alpha = 3$. This model is based on Zipf's law for the size of large institutional investors and the hypothesis that large price movements are generated by the trades of large market participants via a square-root price impact of volume, V, i. e., $r \cong h\sqrt{V}$, where r is the log return and h is a constant. Putting these together with a model for profit maximizing large funds, which have to balance between trading on a perceived mispricing and the price impact of their actions, leads to a power law distribution of volume with tail index 1.5, which by the square-root price impact function and simple power law accounting then produces the "cubic law". See [78,182] for a discussion of this model and the validity of its assumptions. In a somewhat similar spirit, [161] find strong evidence for *exponentially* decaying tails of daily Indian stock returns and speculate about a general inverse relationship between the stage of development of an economy and the closeness to Gaussianity of its stock markets, but it is clear that this is really just speculation.

Some Specific Distributions

Alpha Stable and Related Distributions

As noted in Sect. "Empirical Evidence About the Tails", the history of heavy tailed distributions in finance has its origin in the *alpha stable* model proposed by Mandelbrot [154,155]. Being the first alternative to the Gaussian law, the alpha stable distribution has a long history in financial economics and econometrics, resulting in a large number of books and review articles.

Apart from its good empirical fit the stable distribution has also some attractive theoretical properties such as the stability property and domains of attraction. The stability property states that the index of stability (or shape parameter) remains the same under scaling and addition of different stable rv with the same shape parameter. The concept of domains of attraction is related to a generalized CLT. More specifically, dropping the assumption of a finite variance in the classical CLT, the domains of attraction states, loosely speaking, that the alpha stable distribution is the only possible limit distribution. For a more detailed discussion of this concept we refer to [169], who also provide an overview over alternative stable schemes. While the fat-tailedness of the alpha stable distributions makes it already an attractive candidate for modeling financial returns, the concept of the domains of attraction provides a further argument for its use in finance, as under the relaxation of the assumption of a finite variance of the continuously arriving return innovations the resulting return distribution at lower frequencies is generally an alpha stable distribution.

Although the alpha stable distribution is well established in financial economics and econometrics, there still exists some confusion about the naming convention and parameterization. Popular terms for the alpha stable distribution are the *stable Paretian*, *Lévy stable* or simply *stable* laws. The parameterization of the distribution in turn varies mostly with its application. For instance, to numerically integrate the characteristic function, it is preferable to have a continuous parameterization in all parameters.

The numerical integration of the alpha stable distributions is important, since with the exception of a few special cases, its pdf is unavailable in closed form. However, the characteristic function of the standard parameterization is given by

$$\mathbb{E}\left[\exp\left(itX\right)\right] = \begin{cases} \exp\left(-c^\alpha \left|t\right|^\alpha \left(1 - i\beta \, \text{sign}\left(t\right) \tan \frac{\pi\alpha}{2}\right) \right. \\ \left. \qquad\qquad + \, i\tau t\right) & \alpha \neq 1 \\ \exp\left(-c \left|t\right| \left(1 + i\beta \frac{2}{\pi} \, \text{sign}\left(t\right) \ln \left(\left|t\right|\right)\right) \right. \\ \left. \qquad\qquad + \, i\tau t\right) & \alpha = 1 \, , \end{cases}$$

(12)

where i is the imaginary unit, sign $(\cdot)$ denotes the sign function, which is defined as

$$\text{sign}(x) = \begin{cases} -1 & x < 0 \\ 0 & x = 0 \\ 1 & x > 0\,, \end{cases}$$

$0 < \alpha \leq 2$ denotes the *shape parameter*, *characteristic exponent* or *index of stability*, $-1 \leq \beta \leq 1$ is the skewness parameter, and $\tau \in \mathbb{R}$ and $c \geq 0$ are the location and scale parameters, respectively.

Figure 4 highlights the impact of the parameters α and β. β controls the skewness of the distribution. The shape parameter α controls the behavior of the tails of the distribution and therefore the degree of leptokurtosis. For $\alpha < 2$ moments only up to (and excluding) α exist, and for $\alpha > 1$ we have $\mathbb{E}[X] = \tau$. In general, for $\alpha \in (0, 1)$ and $\beta = 1$ ($\beta = -1$) the support of the distribution is the set (τ, ∞) (or $(-\infty, \tau)$) rather than the whole real line. In the

following we call this stable distribution with $\alpha \in (0, 1)$, $\tau = 0$ and $\beta = 1$ the positive alpha stable distribution.

Moreover, for $\alpha < 2$ the stable law has asymptotic power tails,

$$\bar{F}(x) = P(X > x) \cong c^\alpha d (1 + \beta) x^{-\alpha}$$
$$f_S(x, \alpha, \beta, c, \tau) \cong \alpha c^\alpha d (1 + \beta) x^{-\alpha+1}$$

with $d = \sin\left(\frac{\pi\alpha}{2}\right) \Gamma(\alpha)/\pi$.

For $\alpha = 2$ the stable law is equivalent to the normal law with variance $2c^2$, for $\alpha = 1$ and $\beta = 0$ the Cauchy distribution is obtained, and for $\alpha = 1/2$, $\beta = 1$ and $\tau = 0$ the stable law is equivalent to the Lévy distribution, with support over the positive real line.

An additional property of the stable laws is that they are closed under convolution for constant α, i. e., for two independent alpha stable rvs $X_1 \sim S(\alpha, \beta_1, c_1, \tau_1)$ and $X_2 \sim S(\alpha, \beta_2, c_2, \tau_2)$ with common shape parameter α we
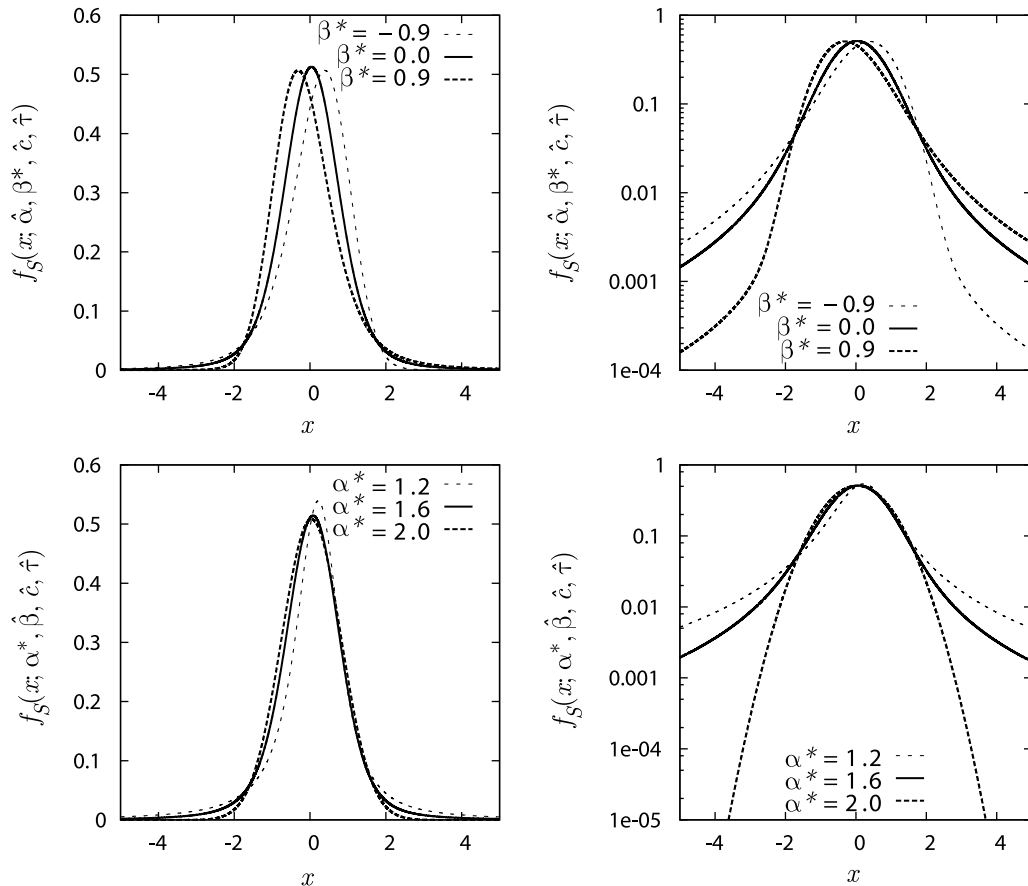
Financial Economics, Fat-Tailed Distributions, Figure 4
Density function (pdf) of the alpha stable distribution for different parameter vectors. The *right panel* plots the log-densities to better visualize the tail behavior. The *upper (lower) section* present the pdf for different values of β (α)

have

$$X_1 + X_2 \sim S\left(\alpha, \frac{\beta_1 c_1^\alpha + \beta_2 c_2^\alpha}{c_1^\alpha + c_2^\alpha}, \left(c_1^\alpha + c_2^\alpha\right)^{1/\alpha}, \tau_1 + \tau_2\right)$$

and

$$aX_1 + b \sim \begin{cases} S\left(\alpha, \operatorname{sign}(a)\beta, |a| c, a\tau + b\right) & \alpha \neq 1 \\ S\left(1, \operatorname{sign}(a)\beta, |a| c, a\tau + b - \frac{2}{\pi}\beta ca \log|a|\right) & \\ & \alpha = 1 . \end{cases}$$

These results can be extended to n stable rvs. The closedness under convolution immediately implies the infinite divisibility of the stable law. As such every stable law corresponds to a Lévy process. A more detailed analysis of alpha stable processes in the context of Lévy processes is given in [192,193].

The computation and estimation of the alpha stable distribution is complicated by the aforementioned non-existence of a closed form pdf. As a consequence, a number of different approximations for evaluating the density have been proposed, see e. g. [65,175]. On the basis of these approximations, parameter estimation is facilitated using for example the maximum-likelihood estimator, see [66], or other estimation methods. As maximum-likelihood estimation relies on computationally demanding numerical integration methods, the early literature focused on alternative estimation methods. The most important methods include the quantile estimation suggested by [77,163], which is still heavily applied in order to obtain starting values for more sophisticated estimation procedures, as well as the characteristic function approach proposed by [127, 131,186]. However, based on its nice asymptotic properties and presently available computational power, the maximum-likelihood approach is preferable.

Many financial applications also involve the simulation of a return series. In derivative pricing, for example, the computation of an expectation is oftentimes impossible as the financial instrument is generally a highly nonlinear function of asset returns. A common way to alleviate this problem is to apply Monte Carlo integration, which in turn requires quasi rvs drawn from the respective return distribution, i. e. the alpha stable distribution. A useful simulation algorithm for alpha stable rvs is proposed by [49], which is a generalization of the algorithm of [120] to the non-symmetric case. A random variable X distributed according to the stable law, $S(\alpha, \beta, c, \tau)$, can be generated as follows:

1. Draw a rv U, uniformly distributed over the interval $(-\pi/2, \pi/2)$, and an (independent) exponential rv E with unit mean,

2. if $\alpha \neq 1$, compute

$$X = cS \frac{\sin(\alpha(U + B))}{\cos^{1/\alpha}(U)}$$
$$\cdot \left(\frac{\cos(U - \alpha(U + B))}{E}\right)^{(1-\alpha)/\alpha} + \tau$$

where

$$B := \frac{\arctan\left(\beta \tan\left(\frac{\pi\alpha}{2}\right)\right)}{\alpha}$$
$$S := \left(1 + \beta^2 \tan^2\left(\frac{\pi\alpha}{2}\right)\right)^{1/(2\alpha)}$$

for $\alpha = 1$ compute

$$X = c\frac{2}{\pi}\left(\left(\frac{\pi}{2} + \beta U\right)\tan(U) - \beta \log\left(\frac{\frac{\pi}{2}E\cos(U)}{\frac{\pi}{2} + \beta U}\right)\right)$$
$$+ \frac{2}{\pi}\beta c \log(c) + \tau .$$

Interestingly, for $\alpha = 2$ the algorithm collapses to the Box–Muller method [42] to generate normally distributed rvs.

As further discussed in Subsect. "The Generalized Hyperbolic Distribution", the mixing of normal distributions allows one to derive interesting distributions having support over the real line and exhibiting heavier tails than the Gaussian. While generally any distribution with support over the positive real line can be used as the mixing distribution for the variance, transformations of the positive alpha stable distribution are often used in financial modeling.

In this context the symmetric alpha stable distributions have a nice representation. In particular, if $X \sim S(\alpha^*, 0, c, 0)$ and A (independent of X) is an α/α^* positive alpha stable rv with scale parameter $\cos^{\alpha^*/\alpha}\left(\frac{\pi\alpha}{2\alpha^*}\right)$ then

$$Z = A^{1/\alpha^*}X \sim S(\alpha, 0, c, 0) .$$

For $\alpha^* = 2$ this property implies that every symmetric alpha stable distribution, i. e. an alpha stable distribution with $\beta = 0$, can be viewed as being conditionally normally distributed, i. e., it can be represented as a continuous variance normal mixture distribution.

Generally, the tail behavior of the resulting mixture distribution is completely determined by the (positive) tails of the variance mixing distribution. In the case of the positive alpha stable distribution this implies that the resulting symmetric stable distribution exhibits very heavy tails, in fact the second moment does not exist. As the literature is controversial on the adequacy of such heavy tails (see Sect. "Empirical Evidence About the Tails"), transformations of the positive alpha stable distribution are oftentimes considered to weight down the tails. The method of

exponential tilting is very useful in this context. In a general setup the exponential tilting of a rv X with respect to a rv Y (defined on the same probability space) defines a new rv $\tilde{X}$, whose pdf can be written as

$$f_{\tilde{X}}(x;\theta) = f_X(x)\frac{\mathbb{E}\left[\exp\left(\theta Y\right)|X=x\right]}{\mathbb{E}\left[\exp\left(\theta Y\right)\right]},$$

where the parameter θ determines the "degree of dampening". The exponential tilting of a rv X with respect to itself, known as *Esscher transformation*, is widely used in financial economics and mathematics, see e.g. [88]. In this case the resulting pdf is given by

$$\begin{aligned} f_{\tilde{X}}(x;\theta) &= \frac{\exp\left(\theta x\right)}{\mathbb{E}\left[\exp\left(\theta X\right)\right]} f_X(x) \\ &= \exp\left(\theta x - K\left(\theta\right)\right) f_X(x), \end{aligned} \quad (13)$$

with $K(\cdot)$ denoting the cumulant generating function, $K(\theta) := \log\left(\mathbb{E}\left[\exp\left(\theta X\right)\right]\right)$.

The *tempered stable* (TS) laws are given by an application of the Esscher transform (13) to a positive alpha stable law. Note that the *Laplace transform* $\mathbb{E}[\exp(-tX)]$, $t \geq 0$, of a positive alpha stable rv is given by $\exp(-\delta (2t)^\alpha)$, where $\delta = c^\alpha/(2^\alpha \cos(\frac{\pi\alpha}{2}))$. Thus, with $\theta = -(1/2)\gamma^{1/\alpha} \leq 0$, the pdf of the tempered stable law is given by

$$\begin{aligned} f_{TS}(x;\alpha,\delta,\gamma) &= \frac{\exp\left(-\frac{1}{2}\gamma^{1/\alpha}x\right)}{\mathbb{E}\left[\exp\left(-\frac{1}{2}\gamma^{1/\alpha}X\right)\right]} f_S(x;\alpha,1,c(\delta,\alpha),0) \\ &= \exp\left(\delta\gamma - \frac{1}{2}\gamma^{1/\alpha}x\right) f_S(x;\alpha,1,c(\delta,\alpha),0) \end{aligned}$$

with $0 < \alpha < 1, \delta > 0$, and $\gamma \geq 0$.

A generalization of the TS distribution was proposed by [22]. This class of *modified stable* (MS) laws can be obtained by applying the following transformation

$$f_{MS}(x,\alpha,\lambda,\delta,\gamma) = c(\alpha,\lambda,\delta,\gamma)\, x^{\lambda+\alpha} f_{TS}(x;\alpha,\delta,\gamma), \quad (14)$$

with $\lambda \in \mathbb{R}$, $\gamma \vee (-\lambda) > 0$ and $c(\alpha,\lambda,\delta,\gamma)$ is a norming constant. For a more detailed analysis, we refer to [22]. Note that the terms "modified stable" or "tempered stable distribution" are not unique in the literature. Very often the so-called truncated Lévy flights/processes (see for example [56,130,157]) are also called TS processes (or corresponding distributions). These distributions are obtained by applying a smooth downweighting of the large jumps (in terms of the Lévy density).

The MS distribution is a quite flexible distribution defined over the positive real line and nests several important distributions. For instance, for $\alpha = 1/2$, the MS distribu-

tion reduces to the *generalized inverse Gaussian* (GIG) distribution, which is of main interest in Subsect. "The Generalized Hyperbolic Distribution".

Note that in contrast to the unrestricted MS distribution, the pdf of the GIG distribution is available in closed form and can be straightforwardly obtained by applying the above transformation. In particular, for $\alpha = 1/2$, the positive alpha stable distribution is the Lévy distribution with closed form pdf given by

$$f_S(x;1/2,1,c,0) = \sqrt{\frac{c}{2\pi}}\frac{\exp\left(-\frac{c}{2x}\right)}{x^{3/2}}.$$

Applying the Esscher transformation (13) with $\theta = -(1/2)\gamma^2$ yields the pdf of the inverse Gaussian (or Wald) distribution

$$f_{IG}(x;\delta,\gamma) = \frac{\delta}{\sqrt{2\pi}}x^{-3/2}\exp\left(\delta\gamma - \left(\delta^2 x^{-1} + \gamma^2 x\right)/2\right),$$

where $\delta > 0$ and $\gamma \geq 0$. Applying the transformation (14) delivers the pdf of the GIG distribution,

$$\begin{aligned} f_{GIG}(x;\lambda,\delta,\gamma) &= \frac{(\gamma/\delta)^\lambda}{2K_\lambda(\delta\gamma)}x^{\lambda-1} \\ &\quad \cdot \exp\left(-\left(\delta^2 x^{-1} + \gamma^2 x\right)/2\right), \quad (15) \end{aligned}$$

with $K_\lambda(\cdot)$ being the *modified Bessel function of the third kind* and of order $\lambda \in \mathbb{R}$. Note that this function is oftentimes called the modified Bessel function of the second kind or Macdonald function. Nevertheless, one representation is given by

$$K_\lambda(x) = \frac{1}{2}\int_0^\infty y^{\lambda-1}\exp\left(-\frac{1}{2}x\left(y + y^{-1}\right)\right)\mathrm{d}y.$$

The parameters of the GIG distribution are restricted to satisfy the following conditions:

$$\begin{aligned} \delta \geq 0 \text{ and } \gamma > 0 &\quad \text{if } \lambda > 0 \\ \delta > 0 \text{ and } \gamma > 0 &\quad \text{if } \lambda = 0 \quad (16) \\ \delta > 0 \text{ and } \gamma \geq 0 &\quad \text{if } \lambda < 0. \end{aligned}$$

Importantly, in contrast to the positive alpha stable law, all moments exist and are given by

$$\mathbb{E}\left[X^r\right] = \left(\frac{\delta}{\gamma}\right)^r\frac{K_{\lambda+r}(\delta\gamma)}{K_\lambda(\delta\gamma)} \quad (17)$$

for all $r > 0$. For a very detailed analysis of the GIG distribution we refer to [117]. The GIG distribution nests several positive distributions as special cases and as limit distributions. Since all of these distributions belong to a special class of the generalized hyperbolic distribution, we proceed with a discussion of the latter, thus providing a broad framework for the discussion of many important distributions in finance.

The Generalized Hyperbolic Distribution

The mixing of normal distributions is well suited for financial modeling, as it allows construction of very flexible distributions. For example, the *normal variance-mean mixture*, given by

$$X = \mu + \beta V + \sqrt{V} Z , \tag{18}$$

with Z being normally distributed and V a positive random variable, generally exhibits heavier tails than the Gaussian distribution. Moreover, this mixture possesses interesting properties, for an overview see [26]. First, similarly to other mixtures, the normal variance-mean mixture is a conditional Gaussian distribution with conditioning on the volatility states, which is appealing when modeling financial returns. Second, if the mixing distribution, i. e. the distribution of V, is infinitely divisible, then X is likewise infinitely divisible. This implies that there exists a Lévy process with support over the whole real line, which is distributed at time $t = 1$ according to the law of X. As the theoretical properties of Lévy processes are well established in the literature (see, e. g., [24,194]), this result immediately suggests to formulate financial models in terms of the corresponding Lévy process.

Obviously, different choices for the distribution of V result in different distributions of X. However, based on the above property, an infinitely divisible distribution is most appealing. For the MS distributions discussed in Subsect. "Alpha Stable and Related Distributions", infinite divisibility is not yet established, although [22] strongly surmise so. However, for some special cases infinite divisibility has been shown to hold. The most popular is the GIG distribution yielding the *generalized hyperbolic* (GH) distribution for X. The latter distribution was introduced by [17] for modeling the distribution of the size of sand particles. The infinite divisibility of the GIG distribution was shown by [20].

To derive the GH distribution as a normal variance-mean mixture, let $V \sim \text{GIG}(\lambda, \delta, \gamma)$ as in (15), with $\gamma = \sqrt{\alpha^2 - \beta^2}$, and $Z \sim \text{N}(0, 1)$ independent of V. Applying (18) yields the GH distributed rv $X \sim \text{GH}(\lambda, \alpha, \beta, \mu, \delta)$ with pdf

$$f_{\text{GH}}(x; \lambda, \alpha, \beta, \mu, \delta)$$
$$= \frac{(\delta\gamma)^\lambda (\delta\alpha)^{1/2 - \lambda}}{\sqrt{2\pi}\delta K_\lambda (\delta\gamma)} \left(1 + \frac{(x - \mu)^2}{\delta^2}\right)^{\lambda/2 - 1/4}$$
$$\cdot K_{\lambda - 1/2} \left(\alpha\delta\sqrt{1 + \frac{(x - \mu)^2}{\delta^2}}\right) \exp(\beta(x - \mu))$$

for $\mu \in \mathbb{R}$ and

$$\delta \geq 0 \text{ and } |\beta| < \alpha \quad \text{if } \lambda > 0$$
$$\delta > 0 \text{ and } |\beta| < \alpha \quad \text{if } \lambda = 0$$
$$\delta > 0 \text{ and } |\beta| \leq \alpha \quad \text{if } \lambda < 0 ,$$

which are the induced parameter restrictions of the GIG distribution given in (16).

Note that, based on the mixture representation (18), the existing algorithms for generating GIG distributed rvs can be used to draw rvs from the GH distribution, see [12, 60].

For $|\beta + u| < \alpha$, the moment generating function of the GH distribution is given by

$$\mathbb{E}\left[\exp(uX)\right] = \exp(\mu u) \left(\frac{\alpha^2 - \beta^2}{\alpha^2 - (\beta + u)^2}\right)^{\lambda/2}$$
$$\cdot \frac{K_\lambda \left(\delta\sqrt{\alpha^2 - (\beta + u)^2}\right)}{K_\lambda \left(\delta\sqrt{\alpha^2 - \beta^2}\right)} . \tag{19}$$

As the moment generating function is infinitely differentiable in the neighborhood of zero, moments of all orders exist and have been derived in [25]. In particular, the mean and the variance of a GH rv X are given by

$$\mathbb{E}[X] = \mu + \frac{\beta\delta}{\sqrt{\alpha^2 - \beta^2}} \frac{K_{\lambda+1}\left(\delta\sqrt{\alpha^2 - \beta^2}\right)}{K_\lambda\left(\delta\sqrt{\alpha^2 - \beta^2}\right)}$$
$$= \mu + \beta\mathbb{E}[X_{\text{GIG}}]$$

$$\mathbb{V}[X] = \frac{\delta}{\sqrt{\alpha^2 - \beta^2}} \frac{K_{\lambda+1}\left(\delta\sqrt{\alpha^2 - \beta^2}\right)}{K_\lambda\left(\delta\sqrt{\alpha^2 - \beta^2}\right)} + \frac{\beta^2\delta^2}{\alpha^2 - \beta^2}$$
$$\cdot \left(\frac{K_{\lambda+2}\left(\delta\sqrt{\alpha^2 - \beta^2}\right)}{K_\lambda\left(\delta\sqrt{\alpha^2 - \beta^2}\right)} - \frac{K_{\lambda+1}^2\left(\delta\sqrt{\alpha^2 - \beta^2}\right)}{K_\lambda^2\left(\delta\sqrt{\alpha^2 - \beta^2}\right)}\right)$$
$$= \mathbb{E}[X_{\text{GIG}}] + \beta^2\mathbb{V}[X_{\text{GIG}}] ,$$

with $X_{\text{GIG}} \sim \text{GIG}(\lambda, \delta, \gamma)$ denoting a GIG distributed rv. Skewness and kurtosis can be derived in a similar way using the third and fourth derivative of the moment generating function (19). However, more information on the tail behavior is given by

$$f_{\text{GH}}(x; \lambda, \alpha, \beta, \mu, \delta) \cong |x|^{\lambda-1} \exp((\mp\alpha + \beta) x) ,$$

which shows that the GH distribution exhibits semi-heavy tails, see [22].

The moment generating function (19) also shows that the GH distribution is generally not closed under convolution. However, for $\lambda \in \{-1/2, 1/2\}$, the modified Bessel

function of the third kind satisfies

$$K_{-\frac{1}{2}}(x) = K_{\frac{1}{2}}(x) = \sqrt{\frac{\pi}{2x}} \exp(-x) ,$$

yielding the following form of the moment generating function for $\lambda = -1/2$

$$\mathbb{E}\left[\exp(uX)|\lambda = -1/2\right]$$
$$= \exp(\mu u) \frac{\exp\left(\delta\sqrt{\alpha^2 - \beta^2}\right)}{\exp\left(\delta\sqrt{\alpha^2 - (\beta + u)^2}\right)} ,$$

which is obviously closed under convolution. This special class of the GH distribution is called normal inverse Gaussian distribution and is discussed in more detail in Subsect. "The Normal Inverse Gaussian Distribution". The closedness under convolution is an attractive property of this distribution as it facilitates forecasting applications.

Another very popular distribution that is nested in the GH distribution is the hyperbolic distribution given by $\lambda = 1$ (see the discussion in Subsect. "The Hyperbolic Distribution"). Its popularity is primarily based on its pdf, which can (except for the norming constant) be expressed in terms of elementary functions allowing for a very fast numerical evaluation. However, given the increasing computer power and the slightly better performance in reproducing the unconditional return distribution, the normal inverse Gaussian distribution is now the most often used subclass.

Interestingly, further well-known distributions can be expressed as limiting cases of the GH distribution, when certain of its parameters approach their extreme values. To this end, the following alternative parametrization of the GH distribution turns out to be useful. Setting $\xi = 1/\sqrt{1 + \delta\sqrt{\alpha^2 - \beta^2}}$ and $\chi = \xi\beta/\alpha$ renders the two parameters invariant under location and scale transformations. This is an immediate result of the following property of the GH distribution. If $X \sim \text{GH}(\lambda, \alpha, \beta, \mu, \delta)$, then

$$a + bX \sim \text{GH}\left(\lambda, \frac{\alpha}{|b|}, \frac{\beta}{|b|}, a + b\mu, \delta|b|\right) .$$

Furthermore, the parameters are restricted by $0 \leq |\chi| < \xi < 1$, implying that they are located in the so-called *shape triangle* introduced by [27]. Figure 5 highlights the impact of the parameters in the GH distribution in terms of χ, ξ and λ. Obviously, χ controls the skewness and ξ the tailedness of the distribution. The impact of λ is not so clear-cut. The lower panels in Fig. 5 depict the pdfs for different values of λ whereby the first two moments and the values of χ and ξ are kept constant to show the "partial" influence.

As pointed out by [69], the limit distributions can be classified by the values of ξ and χ as well as by the values ϱ and ζ of a second location and scale invariant parametrization of the GH, given by $\varrho = \beta/\alpha$ and $\zeta = \delta\sqrt{\alpha^2 - \beta^2}$, as follows:

- $\xi = 1$ and $-1 \leq \chi \leq 1$: The resulting limit distributions depend here on the values of λ. Note, that in order to reach the boundary either $\delta \to 0$ or $|\beta| \to \alpha$.
  - For $\lambda > 0$ and $|\beta| \to \alpha$ no limit distribution exists, but for $\delta \to 0$ the GH distribution converges to the distribution of a variance gamma rv (see Subsect. "The Variance Gamma Distribution"). However, note that $|\beta| < \alpha$ implies $|\chi| < \xi$ and so the limit distribution is not valid in the corners. For these cases, the limit distributions are given by $\xi = |\chi|$ and $0 < \xi \leq 1$, i. e. the next case.
  - For $\lambda = 0$ there exists no proper limit distribution.
  - For $\lambda < 0$ and $\delta \to 0$ no proper distribution exists but for $\beta \to \pm\alpha$ the limit distribution is given in [185] with pdf

$$\frac{2^{\lambda+1}\left(\delta^2 + (x - \mu)^2\right)^{(\lambda-1/2)/2}}{\sqrt{2\pi}\,\Gamma(-\lambda)\,\delta^{2\lambda}\alpha^{\lambda-1/2}}$$
$$\cdot K_{\lambda-1/2}\left(\alpha\sqrt{\delta^2 + (x - \mu)^2}\right)$$
$$\cdot \exp(\pm\alpha(x - \mu)) , \qquad (20)$$

which is the limit distribution of the corners, since $\beta = \pm\alpha$ is equivalent to $\chi = \pm\xi$. This distribution was recently called the GH skew t distribution by [1]. Assuming additionally that $\alpha \to 0$ and $\beta = \varrho\alpha \to 0$ with $\varrho \in (-1, 1)$ yields the limit distribution in between

$$\frac{\Gamma(-\lambda + 1/2)}{\sqrt{\pi\delta^2}\,\Gamma(-\lambda)}\left(1 + \frac{(x - \mu)^2}{\delta^2}\right)^{\lambda-1/2} ,$$

which is the scaled and shifted Student's t distribution with -2λ degrees of freedom, expectation μ and variance $4\lambda^2\nu/(\nu - 2)$, for more details see Subsect. "The Student t Distribution".

- $\xi = |\chi|$ and $0 < \xi \leq 1$: Except for the upper corner the limit distribution of the right boundary can be derived for

$$\beta = \alpha - \frac{\phi}{2}; \quad \alpha \to \infty; \quad \delta \to 0; \quad \alpha\delta^2 \to \tau$$

with $\phi > 0$ and is given by the μ-shifted GIG distribution $\text{GIG}(\lambda, \sqrt{\tau}, \sqrt{\phi})$. The distribution for the left boundary is the same distribution but mirrored at $x = 0$. Note that the limit behavior does not depend
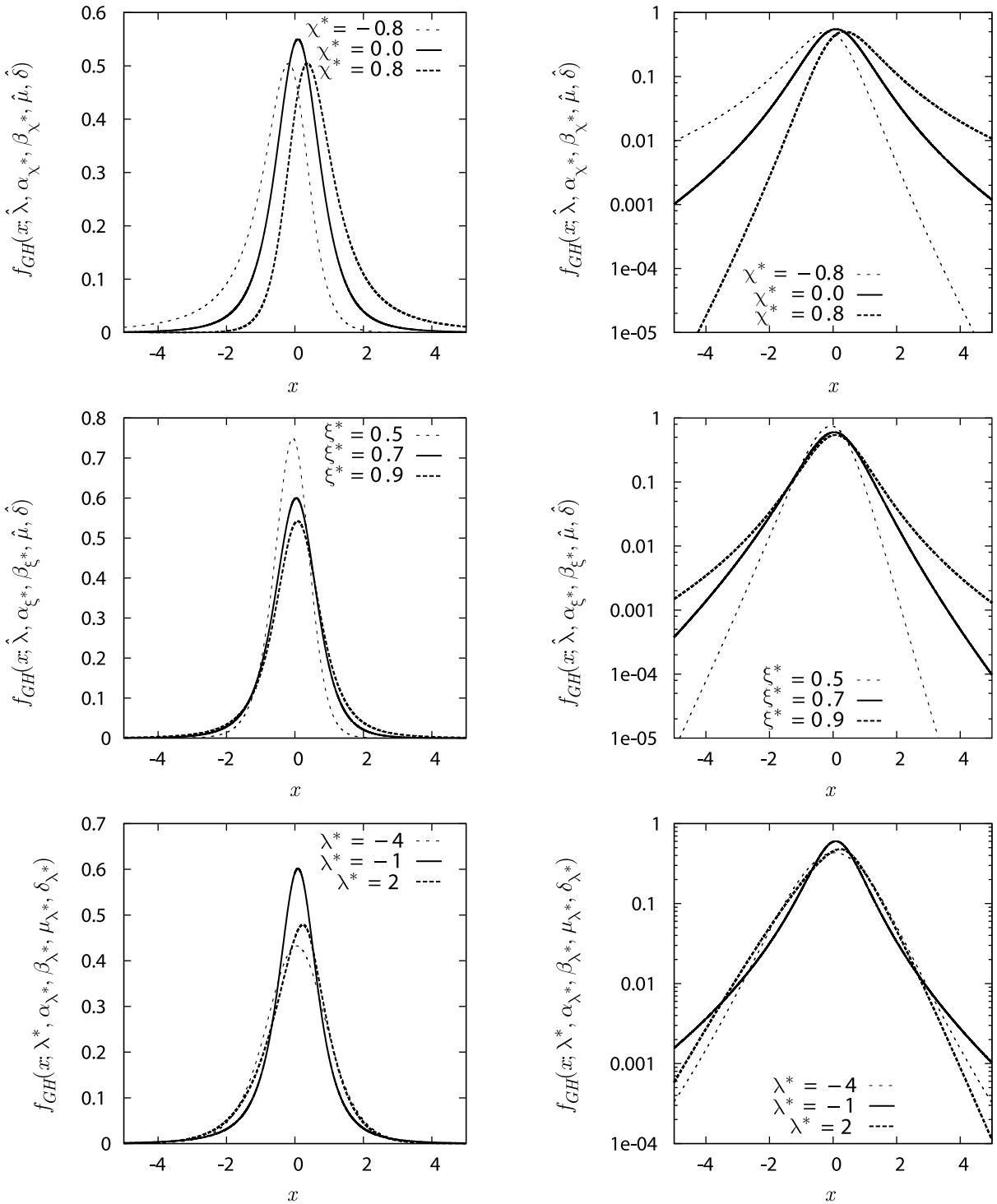
Financial Economics, Fat-Tailed Distributions, Figure 5
Density function (pdf) of the GH distribution for different parameter vectors. The *right panel* plots the log-densities to better visualize the tail behavior. The *upper and middle section* present the pdf for different values of χ and ξ. Note that these correspond to different values of α and β. The *lower panel* highlights the influence of λ if the first two moments, as well as χ and ξ, are held fixed. This implies that α, β, μ and δ have to be adjusted accordingly

on λ. However, to obtain the limit distributions for the left and right upper corners we have to distinguish for different values of λ. Recall that for the regime $\xi = 1$ and $-1 \leq \chi \leq 1$ the derivation was not possible.

– For $\lambda > 0$ the limit distribution is a gamma distribution.
– For $\lambda = 0$ no limit distribution exists.
– For $\lambda < 0$ the limit distribution is the reciprocal gamma distribution.

• $\xi = \chi = 0$: This is the case for $\alpha \to \infty$ or $\delta \to \infty$. If only $\alpha \to \infty$ then the limit distribution is the Dirac measure concentrated in μ. If in addition $\delta \to \infty$ and $\delta/\alpha \to \sigma^2$ then the limit distribution is a normal distribution with mean $\mu + \beta\sigma^2$ and variance σ^2.

As pointed out by [185] applying the unrestricted GH distribution to financial data results in a very flat likelihood function especially for λ. This characteristic is illustrated in Fig. 6, which plots the maximum of the log likelihood for different values of λ using our sample of the S&P500 index returns. This implies that the estimate of λ is generally associated with a high standard deviation. As a consequence, rather than using the GH distribution directly, the finance literature primarily predetermines the value of λ, resulting in specific subclasses of the GH distribution, which are discussed in the sequel. However, it is still interesting to derive the general results in terms of the GH distribution (or the corresponding Lévy process) directly and to restrict only the empirical application to a subclass. For example [191] derived a diffusion process with GH marginal distribution, which is a generalization of the result of [33], who proposed a diffusion process with hyperbolic marginal distribution.

The Hyperbolic Distribution Recall, that the *hyperbolic* (HYP) distribution can be obtained as a special case of the GH distribution by setting $\lambda = 1$. Thus, all properties of the GH law can be applied to the HYP case. For instance the pdf of the HYP distribution is straightforwardly given by (19) setting $\lambda = 1$

$$
\begin{aligned}
f_H(x;\alpha,\beta,\mu,\delta) &:= f_{GH}(x;1,\alpha,\beta,\mu,\delta) \\
&= \frac{\sqrt{\alpha^2 - \beta^2}}{2\alpha\delta K_1\left(\delta\sqrt{\alpha^2 - \beta^2}\right)} \\
&\quad \cdot \exp\left(-\alpha\sqrt{\delta^2 + (x-\mu)^2}\right. \\
&\qquad \left. +\beta(x-\mu)\right),
\end{aligned}
\tag{21}
$$

where $0 \leq |\beta| < \alpha$ are the shape parameter and $\mu \in \mathbb{R}$ and $\delta > 0$ are the location and scale parameter, respectively.

The distribution was applied to stock return data by [70,71,132] while [33] derived a diffusion model with marginal distribution belonging to the class of HYP distributions.

The Normal Inverse Gaussian Distribution The *normal inverse Gaussian* (NIG) distribution is given by the GH distribution with $\lambda = -1/2$ and has the following pdf

$$
\begin{aligned}
f_{NIG}(x;\alpha,\beta,\mu,\delta) &:= f_{GH}\left(x;-\frac{1}{2},\alpha,\beta,\mu,\delta\right) \\
&= \frac{\alpha\delta K_1\left(\alpha\sqrt{\delta^2 + (x-\mu)^2}\right)}{\pi\sqrt{\delta^2 + (x-\mu)^2}} \\
&\quad \cdot \exp\left(\delta\gamma + \beta(x-\mu)\right)
\end{aligned}
\tag{22}
$$
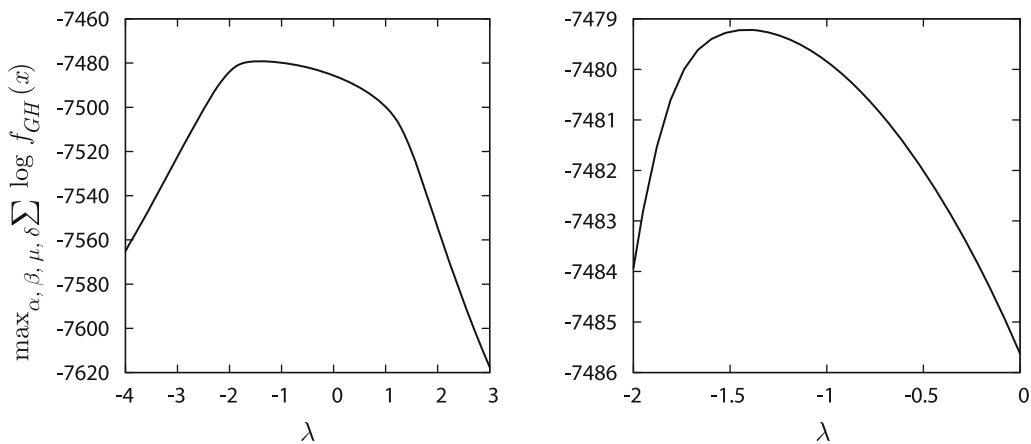
Financial Economics, Fat-Tailed Distributions, Figure 6
Partially maximized log likelihood, estimated maximum log likelihood values of the GH distribution for different values of λ

with $0 < |\beta| \leq \alpha$, $\delta > 0$ and $\mu \in \mathbb{R}$. The moments of a NIG distributed rv can be obtained from the moment generating function of the GH distribution (19) and are given by

$$\mathbb{E}[X] = \mu + \frac{\delta\beta}{\sqrt{\alpha^2 - \beta^2}} \quad \text{and} \quad \mathbb{V}[X] = \frac{\delta\alpha^2}{\sqrt{\alpha^2 - \beta^2}^3}$$

$$\mathbb{S}[X] = 3\frac{\beta}{\alpha\sqrt{\delta\sqrt{\alpha^2 - \beta^2}}} \quad \text{and} \quad \mathbb{K}[X] = 3\frac{\alpha^2 + 4\beta^2}{\delta\alpha^2\sqrt{\alpha^2 - \beta^2}} \ .$$

This distribution was heavily applied in financial economics for modeling the unconditional as well as the conditional return distribution, see e.g. [18,21,185]; as well as [10,19,114], respectively. Recently, [57] used the NIG distribution for modeling realized variance and found improved forecast performance relative to a Gaussian model. A more realistic modeling of the distributional properties is not only important for risk management or forecasting, but also for statistical inference. For example the efficient method of moments, proposed by [87] requires the availability of a highly accurate auxiliary model, which provide the objective function to estimate a more structural model. Recently, [39] provided such an auxiliary model, which uses the NIG distribution and realized variance measures.

Recall that for $\lambda = -1/2$ the mixing distribution is the inverse Gaussian distribution, which facilitates the generation of rvs. Hence, rvs with NIG distribution can be generated in the following way:

1. Draw a chi-square distributed rv C with one degree of freedom and a uniformly distributed rv over the interval $(0, 1)$ U
2. Compute

$$X_1 = \frac{\delta}{\gamma} + \frac{1}{2\delta\gamma}\left(\frac{\delta C}{\gamma} - \sqrt{4\delta^3 C/\gamma + \delta^2 C^2/\gamma^2}\right)$$

3. If $U < \delta/(\gamma(\delta/\gamma + X_1))$ return X_1 else return $\delta^2/(\gamma^2 X_1)$.

As pointed out by [187] the main difference between the HYP and NIG distribution: "Hyperbolic log densities, being hyperbolas, are strictly concave everywhere. Therefore they cannot form any sharp tips near $x = 0$ without loosing too much mass in the tails ... In contrast, NIG log densities are concave only in an interval around $x = 0$, and convex in the tails." Moreover, [19] concludes, "It is, moreover, rather typical that asset returns exhibit tail behavior that is somewhat heavier than log linear, and this

further strengthens the case for the NIG in the financial context".

The Student t Distribution Next to the alpha stable distribution *Student's t* (t thereafter) distribution has the longest history in financial economics. One reason is that although the non-normality of asset returns is widely accepted, there still exists some discussion on the exact tail behavior. While the alpha stable distribution implies extremely slowly decreasing tails for $\alpha \neq 2$, the t distribution exhibits power tails and existing moments up to (and excluding) ν. As such, the t distribution might be regarded as the strongest competitor to the alpha stable distribution, shedding also more light on the empirical tail behavior of returns. The pdf for the scaled and shifted t distribution is given by

$$f_t(x; \nu, \mu, \sigma) = \frac{\Gamma((\nu + 1)/2)}{\sqrt{\nu\pi}\,\Gamma(\nu/2)\,\sigma}$$
$$\cdot \left(1 + \frac{1}{\nu}\left(\frac{x - \mu}{\sigma}\right)^2\right)^{-(\nu+1)/2} \quad (23)$$

for $\nu > 0$, $\sigma > 0$ and $\mu \in \mathbb{R}$. For $\mu = 0$ and $\sigma = 1$ the well-known standard t distribution is obtained. The shifted and scaled t distribution can also be interpreted as a mean-variance mixture (18) with a reciprocal gamma distribution as a mixing distribution. The mean, variance, and kurtosis (3) are given by μ, $\sigma^2\nu/(\nu - 2)$, and $3(\nu - 2)/(\nu - 4)$, provided that $\nu > 1$, $\nu > 2$, and $\nu > 4$, respectively. The tail behavior is

$$f_t(x, \nu, \mu, \sigma) \cong cx^{-\nu-1} \ .$$

The t distribution is one of the standard non-normal distributions in financial economics, see e.g. [36,38,184]. However, as the unconditional return distribution may exhibit skewness, a skewed version of the t distribution might be more adequate in some cases. In fact, several skewed t distributions were proposed in the literature, for a short overview see [1]. The following special form of the pdf was considered in [81,102]

$$f_{t,\text{FS}}(x; \nu, \mu, \sigma, \beta)$$
$$= \frac{2\beta}{\beta^2 + 1}\frac{\Gamma\left(\frac{\nu+1}{2}\right)}{\Gamma(\nu/2)\sqrt{\pi\nu}\sigma}$$
$$\cdot \left(1 + \frac{1}{\nu}\left(\frac{x-\mu}{\sigma}\right)^2\left(\frac{1}{\beta^2}\mathcal{I}(X \geq \mu) + \beta^2\mathcal{I}(x < \mu)\right)\right)^{-\frac{\nu+1}{2}}$$

with $\beta > 0$. For $\beta = 1$ the pdf reduces to the pdf of the usual symmetric scaled and shifted t distribution. Another

skewed t distribution was proposed by [116] with pdf

$$
\begin{aligned}
&f_{t,\mathrm{JF}}\left(x, \nu, \mu, \sigma, \beta\right) \\
&= \frac{\Gamma\left(\nu+\beta\right) 2^{1-\nu-\beta}}{\Gamma\left(\nu/2\right) \Gamma\left(\nu/2+\beta\right) \sqrt{\nu+\beta}\,\sigma} \\
&\quad \cdot \left(1 + \frac{\frac{x-\mu}{\sigma}}{\sqrt{\nu+\beta+\left(\frac{x-\mu}{\sigma}\right)^2}}\right)^{(\nu+1)/2} \\
&\quad \cdot \left(1 - \frac{\frac{x-\mu}{\sigma}}{\sqrt{\nu+\beta+\left(\frac{x-\mu}{\sigma}\right)^2}}\right)^{\beta+(\nu+1)/2}
\end{aligned}
$$

for $\beta > -\nu/2$. Again, the usual t distribution can be obtained as a special case for $\beta = 0$. A skewed t distribution in terms of the pdf and cdf of the standard t distribution $f_t\left(x; \nu, 0, 1\right)$ and $F_t\left(x; \nu, 0, 1\right)$ is given by [13,43]

$$
\begin{aligned}
&f_{t,\mathrm{AC}}\left(x; \nu, \mu, \sigma, \beta\right) \\
&= \frac{2}{\sigma} f_t\left(\frac{x-\mu}{\sigma}, \nu, 0, 1\right) \\
&\quad \cdot F_t\left(\beta\left(\frac{x-\mu}{\sigma}\right)\sqrt{\frac{\nu+1}{\nu+\left(\frac{x-\mu}{\sigma}\right)^2}}, \nu+1, 0, 1\right)
\end{aligned}
$$

for $\beta \in \mathbb{R}$.

Alternatively, a skewed t distribution can also be obtained as a limit distribution of the GH distribution. Recall that for $\lambda < 0$ and $\beta \to \alpha$ the limit distribution is given by (20) as

$$
\begin{aligned}
&f_{t,\mathrm{GH}}\left(x; \lambda, \mu, \delta, \alpha\right) \\
&= \frac{2^{\lambda+1}\left(\delta^2+(x-\mu)^2\right)^{(\lambda-1/2)/2}}{\sqrt{2\pi}\,\Gamma\left(-\lambda\right) \delta^{2\lambda}\alpha^{\lambda-1/2}} \\
&\quad \cdot K_{\lambda-1/2}\left(\alpha\sqrt{\delta^2+(x-\mu)^2}\right) \cdot \exp\left(\alpha\left(x-\mu\right)\right).
\end{aligned}
$$

for $\alpha \in \mathbb{R}$. The symmetric t distribution is obtained for $\alpha \to 0$. The distribution was introduced by [185] and a more detailed examination was recently given in [1].

The Variance Gamma Distribution The *variance gamma* (VG) distribution can be obtained as a mean-variance mixture with gamma mixing distribution. Note that the gamma distribution is obtained in the limit from the GIG distribution for $\lambda > 0$ and $\delta \to 0$. The pdf of the VG distribution is given by

$$
\begin{aligned}
&f_{\mathrm{VG}}\left(x; \mu, \alpha, \beta, \lambda\right) := \lim_{\delta \to 0} f_{\mathrm{GH}}\left(x; \lambda, \alpha, \beta, \delta, \mu\right) \\
&= \frac{\gamma^{2\lambda}\left|x-\mu\right|^{\lambda-1/2} K_{\lambda-1/2}\left(\alpha\left|x-\mu\right|\right)}{\sqrt{\pi}\,\Gamma\left(\lambda\right)\left(2\alpha\right)^{\lambda-1/2}} \exp\beta\left(x-\mu\right).
\end{aligned}
$$

$$\tag{24}$$

Note, the usual parameterization of the VG distribution

$$
\begin{aligned}
&f_{\mathrm{VG}}^*\left(x; \sigma^*, \theta^*, \nu^*, \mu^*\right) \\
&= \frac{2\exp\left(\theta^*\left(x-\mu^*\right)/\sigma^{*2}\right)}{\nu^{*1/\nu^*}\sqrt{2\pi\sigma^{*2}}\,\Gamma\left(1/\nu^*\right)} \left(\frac{\left(x-\mu^*\right)^2}{2\sigma^{*2}/\nu^*+\theta^{*2}}\right)^{\frac{1}{2\nu^*}-\frac{1}{4}} \\
&\quad \cdot K_{\frac{1}{\nu^*}-\frac{1}{2}}\left(\frac{\sqrt{\left(x-\mu^*\right)^2\left(2\sigma^{*2}/\nu^*+\theta^{*2}\right)}}{\sigma^{*2}}\right)
\end{aligned}
$$

is different from the one used here, however the parameters can be transformed between these representations in the following way

$$
\sigma^* = \sqrt{\frac{2\lambda}{\alpha^2-\beta^2}}; \quad \theta^* = \frac{2\beta\lambda}{\alpha^2-\beta^2};
$$

$$
\nu^* = \frac{1}{\lambda}; \qquad \mu^* = \mu.
$$

This distribution was introduced by [149,150,151]. For $\lambda = 1$ (the HYP case) we obtain a skewed, shifted and scaled Laplace distribution with pdf

$$
\begin{aligned}
&f_{\mathrm{L}}\left(x; \alpha, \beta, \mu\right) \\
&:= \lim_{\delta \to 0} f_{\mathrm{GH}}\left(x; 1, \alpha, \beta, \delta, \mu\right) \\
&= \frac{\alpha^2-\beta^2}{2\alpha}\exp\left(-\alpha\left|x-\mu\right|+\beta\left(x-\mu\right)\right).
\end{aligned}
$$

A generalization of the VG distribution to the so-called CGMY distribution was proposed by [48].

The Cauchy Distribution Setting $\lambda = -1/2$, $\beta \to 0$ and $\alpha \to 0$ the GH distribution converges to the *Cauchy* distribution with parameters μ and δ. Since the Cauchy distribution belongs to the class of symmetric alpha stable ($\alpha = 1$) and symmetric t distributions ($\nu = 1$) we refer to Subsect. "Alpha Stable and Related Distributions" and "The Student t Distribution" for a more detailed discussion.

The Normal Distribution For $\alpha \to \infty$, $\beta = 0$ and $\delta = 2\sigma^2$ the GH distribution converges to the *normal* distribution with mean μ and variance σ^2.

Finite Mixtures of Normal Distributions

The density of a (finite) mixture of k normal distributions is given by a linear combination of k Gaussian *component*

densities, i. e.,

$$f_{\text{NM}}(x; \theta) = \sum_{j=1}^{k} \lambda_j \phi\left(x; \mu_j, \sigma_j^2\right),$$
$$\phi\left(x; \mu, \sigma^2\right) = \frac{1}{\sqrt{2\pi}\sigma} \exp\left(-\frac{(x-\mu)^2}{2\sigma^2}\right),$$

(25)

where $\theta = (\lambda_1, \ldots, \lambda_{k-1}, \mu_1, \ldots, \mu_k, \sigma_1^2, \ldots, \sigma_k^2)$, $\lambda_k = 1 - \sum_{j=1}^{k-1} \lambda_j$, $\lambda_j > 0$, $\mu_j \in \mathbb{R}$, $\sigma_j^2 > 0$, $j = 1, \ldots, k$, and $(\mu_i, \sigma_i^2) \neq (\mu_j, \sigma_j^2)$ for $i \neq j$. In (25), the λ_j, μ_j, and σ_j^2 are called the *mixing weights*, *component means*, and *component variances*, respectively.

Finite mixtures of normal distributions have been applied as early as 1886 in [174] to model leptokurtic phenomena in astrophysics. A voluminous literature exists, see [165] for an overview. In our discussion, we shall focus on a few aspects relevant for applications in finance. In this context, (25) arises naturally when the component densities are interpreted as different *market regimes*. For example, in a two-component mixture ($k = 2$), the first component, with a relatively high mean and small variance, may be interpreted as the bull market regime, occurring with probability λ_1, whereas the second regime, with a lower expected return and a greater variance, represents the bear market. This (typical) pattern emerges for the S&P500 returns, see Table 1. Clearly (25) can be generalized to accommodate non-normal component densities; e. g., [104] model stock returns using mixtures of generalized error distributions of the form (40). However, it may be argued that in this way much of the original appeal of (25), i. e., within-regime normality along with CLT arguments, is lost.

The moments of (25) can be inferred from those of the normal distribution, with mean and variance given by

$$\mathbb{E}[X] = \sum_{j=1}^{k} \lambda_j \mu_j, \text{ and}$$
$$\mathbb{V}[X] = \sum_{j=1}^{k} \lambda_j \left(\sigma_j^2 + \mu_j^2\right) - \left(\sum_{j=1}^{k} \lambda_j \mu_j\right)^2,$$

(26)

respectively. The class of finite normal mixtures is very flexible in modeling the leptokurtosis and, if existent, skewness of financial data. To illustrate the first property, consider the *scale normal mixture*, where, in (25), $\mu_1 = \mu_2 = \cdots = \mu_k := \mu$. In fact, when applied to financial return data, it is often found that the market regimes differ mainly in their variances, while the component means are rather close in value, and often their differences are not significant statistically. This reflects the observation that excess kurtosis is a much more pronounced (and ubiqui-

Financial Economics, Fat-Tailed Distributions, Table 1
Maximum-likelihood parameter estimates of the iid model

Distri-bution	Parameters					Loglik
	$\hat\lambda$	$\hat\mu$	$\hat\alpha$	$\hat\beta$	$\hat\delta$	
GH	−1.422	0.087	0.322	−0.046	1.152	−7479.2
	(0.351)	(0.018)	(0.222)	(0.022)	(0.139)	
	$\hat\mu$	$\hat\delta$	$\hat\lambda$	$\hat\alpha$		
t_{GH}	0.084	1.271	3.445	−0.041		−7479.7
	(0.018)	(0.052)	(0.181)	(0.021)		
	$\hat\nu$	$\hat\mu$	$\hat\sigma$	$\hat\beta$		
t_{JF}	3.348	0.098	0.684	0.091		−7480.0
	(0.179)	(0.025)	(0.012)	(0.049)		
	$\hat\nu$	$\hat\mu$	$\hat\sigma$	$\hat\beta$		
t_{AC}	3.433	0.130	0.687	−0.123		−7480.1
	(0.180)	(0.042)	(0.013)	(0.068)		
	$\hat\nu$	$\hat\mu$	$\hat\sigma$	$\hat\beta$		
t_{FS}	3.432	0.085	0.684	0.972		−7480.3
	(0.180)	(0.020)	(0.012)	(0.017)		
Sym-metric t	$\hat\nu$	$\hat\mu$	$\hat\sigma$			
	3.424	0.056	0.684			−7481.7
	(0.179)	(0.011)	(0.012)			
	$\hat\mu$	$\hat\alpha$	$\hat\beta$	$\hat\delta$		
NIG	0.088	0.784	−0.048	0.805		−7482.0
	(0.018)	(0.043)	(0.022)	(0.028)		
	$\hat\mu$	$\hat\alpha$	$\hat\beta$	$\hat\delta$		
HYP	0.090	1.466	−0.053	0.176		−7499.5
	(0.018)	(0.028)	(0.023)	(0.043)		
	$\hat\mu$	$\hat\alpha$	$\hat\beta$	$\hat\lambda$		
VG	0.092	1.504	−0.054	1.115		−7504.2
	(0.013)	(0.048)	(0.019)	(0.054)		
Alpha stable	$\hat\alpha$	$\hat\beta$	$\hat c$	$\hat\tau$		
	1.657	−0.094	0.555	0.036		−7522.5
	(0.024)	(0.049)	(0.008)	(0.015)		
Finite mixture ($k = 2$)	$\hat\lambda_1$	$\hat\mu_1$	$\hat\mu_2$	$\hat\sigma_1^2$	$\hat\sigma_2^2$	
	0.872	0.063	−0.132	0.544	4.978	−7580.8
	(0.018)	(0.012)	(0.096)	(0.027)	(0.530)	
	$\hat\mu$	$\hat\sigma$				
Cauchy	0.060	0.469				−7956.6
	(0.010)	(0.008)				
	$\hat\mu$	$\hat\sigma$				
Normal	0.039	1.054				−8168.9
	(0.014)	(0.010)				

Shown are maximum likelihood estimates for iid models with different assumptions about the distribution of the innovations. Standard errors are given in parentheses. "Loglik" is the value of the maximized log likelihood function.

tous) property of asset returns than skewness. In the scale mixture case, the density is symmetric, but with higher peaks and thicker tails than the normal with the same mean and variance. To see this, note that $\sum_j (\lambda_j/\sigma_j) >$

$(\sum_j \lambda_j \sigma_j^2)^{-1/2} \Leftrightarrow (\sum_j \lambda_j \sigma_j^2)^{1/2} > [\sum_j(\lambda_j/\sigma_j)]^{-1}$. But $(\sum_j \lambda_j \sigma_j^2)^{1/2} > \sum_j \lambda_j \sigma_j > [\sum_j(\lambda_j/\sigma_j)]^{-1}$ by Jensen's and the arithmetic-harmonic mean inequality, respectively. This shows $f_{\mathrm{NM}}(\mu;\theta) > \phi(\mu;\mu,\sum_j \lambda_j \sigma_j^2)$, i. e., peakedness. Tailedness follows from the observation that the difference between the mixture and the mean-variance equivalent normal density is asymptotically dominated by the component with the greatest variance. Moreover, the densities of the scale mixture and the mean-variance equivalent Gaussian intersect exactly two times on both sides of the mean, so that the scale mixture satisfies the density crossing condition in Finucan's theorem mentioned in Sect. "Definition of the Subject" and observed in the center panel of Fig. 1. This follows from the fact that, if $a_1, \ldots, a_n$ and $\gamma_1 < \cdots < \gamma_n$ are real constants, and N is the number of real zeros of the function $\varphi(x) = \sum_i a_i e^{\gamma_i x}$, then $W - N$ is a non-negative even integer, where W is the number of sign changes in the sequence $a_1, \ldots, a_n$ [183]. Skewness can be incorporated into the model when the component means are allowed to differ. For example, if, in the two-component mixture, the high-variance component has both a smaller mean and mixing weight, then the distribution will be skewed to the left.

Because of their flexibility and the aforementioned economic interpretation, finite normal mixtures have been frequently used to model the unconditional distribution of asset returns [40,44,129,179], and they have become rather popular since the publication of Hamilton's [101] paper on Markov-switching processes, where the mixing weights are assumed to be time-varying according to a k-state Markov chain; see, e. g., [200] for an early contribution in this direction.

However, although a finite mixture of normals is a rather flexible model, its tails decay eventually in a Gaussian manner, and therefore, according to the discussion in Sect. "Empirical Evidence About the Tails", it may often not be appropriate to model returns at higher frequencies unconditionally. Nevertheless, when incorporated into a GARCH structure (see Sect. "Volatility Clustering and Fat Tails"), it provides a both useful and intuitively appealing framework for modeling the *conditional* distribution of asset returns, as in [5,96,97]. These papers also provide a discussion of alternative interpretations of the mixture model (25), as well as an overview over the extensive literature.

Empirical Comparison

In the following we empirically illustrate the adequacy of the various distributions discussed in the previous sections for modeling the unconditional return distribution. Table 1 presents the estimation results for the S&P500 index assuming iid returns. The log likelihood values clearly indicate the inadequacy of the normal, Cauchy and stable distributions. This is also highlighted in the upper panel of Fig. 7, which clearly shows that the tails of the Cauchy and stable distributions are too heavy, whereas those of the normal distribution are too weak. To distinguish the other distributions in more detail, the lower left panel is an enlarged display of the shadowed box in the upper panel. It illustrates nicely that the two component mixture, VG and HYP distribution exhibit semiheavy tails, which are probably a little bit to weak for an adequate modeling as is indicated by the log likelihood values. Similarly, the two-component finite normal mixture, although much better than the normal, cannot keep up with most of the other models, presumably due to its essentially Gaussian tails. Although the pdf of the NIG distribution lies somewhere in between the pdfs of the HYP and the different t distributions, the log likelihood value clearly indicates that this distribution is in a statistical sense importantly closer to the t distributions. A further distinction between the other distributions including all kinds of t distributions and the GH distribution is nearly impossible, as can be seen from the lower right plot, which is an enlarged display of the lower left panel. The log likelihood values also do not allow for a clear distinction. Note also that the symmetric t distribution performs unexpectedly well. In particular, its log likelihood is almost indistinguishable from those of the skewed versions. Also note that, for all t distributions, the estimated tail index, ν, is close to 3.5, which is in accordance with the results from semiparametric tail estimation in Sect. "Empirical Evidence About the Tails".

The ranking of the distributions in terms of the log likelihood depends of course heavily on the dataset, and different return series may imply different rankings. However, Table 1 also highlights some less data-dependent results, which are more or less accepted in the literature, e. g., the tails of the Cauchy and stable distributions are too heavy, and those of the HYP and VG are too light for the unconditional distribution. This needs of course no longer be valid in a different modeling setup. Especially in a GARCH framework the conditional distribution don't need to imply such heavy tails because the model itself imposes fatter tails.

In Sect. "Application to Value-at-Risk", the comparison of the models will be continued on the basis of their ability the measure the Value-at-Risk, an important concept in risk management.
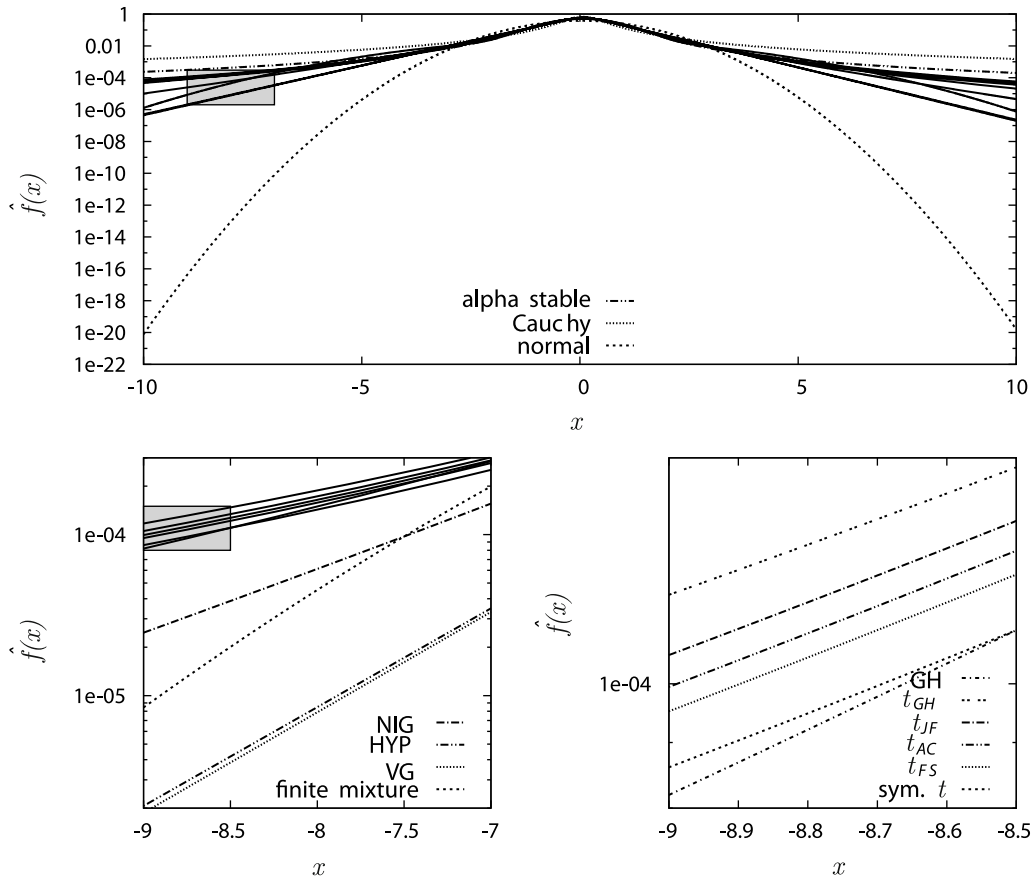
Financial Economics, Fat-Tailed Distributions, Figure 7
Plot of the estimated pdfs of the different return distributions assuming iid returns

Volatility Clustering and Fat Tails

It has long been known that the returns of most financial assets, although close to being unpredictable, exhibit significant dependencies in measures of volatility, such as absolute or squared returns. Moreover, the empirical results based on the recent availability of more precise volatility measures, such as the *realized volatility*, which is defined as the sum over the squared intradaily high-frequency returns (see, e. g., [7] and [23]), also point towards the same direction. In particular, the realized volatility has been found to exhibit strong persistence in its autocorrelation function, which shows a hyperbolic decay indicating the presence of long memory in the volatility process. In fact, this finding as well as other stylized features of the realized volatility have been observed across different data sets and markets and are therefore by now widely acknowledged and established in the literature. For a more detailed and originating discussion on the stylized facts of the high-fre-

quency based volatility measures for stock returns and exchange returns we refer to [8,9], respectively.

The observed dependence of time-varying pattern of the volatility is usually referred to as *volatility clustering*. It is also apparent in the top panel of Fig. 1 and was already observed by Mandelbrot [155], who noted that "large changes tend to be followed by large changes—of either sign—and small changes tend to be followed by small changes". It is now well understood that volatility clustering can explain at least part of the fat-tailedness of the unconditional return distribution, even if the *conditional* distribution is Gaussian. This is also supported by the recent observation that if the returns are scaled by the realized volatility then the distribution of the resulting series is approximately Gaussian (see [9] and [8]). To illustrate, consider a time series $\{\epsilon_t\}$ of the form

$$\epsilon_t = \eta_t \sigma_t \,, \tag{27}$$

where $\{\eta_t\}$ is an iid sequence with mean zero and unit

variance, with η_t being independent of σ_t, so that σ_t^2 is the conditional variance of ϵ_t. With respect to the kurtosis measure $\mathbb{K}$ in (3), it has been observed by [108], and earlier by [31] in a different context, that, as long as σ_t^2 is not constant, Jensen's inequality implies $\mathbb{E}[\epsilon_t^4] = \mathbb{E}[\eta_t^4]\mathbb{E}[\sigma_t^4] > \mathbb{E}[\eta_t^4]\mathbb{E}^2[\sigma_t^2]$, so that the kurtosis of the unconditional distribution exceeds that of the innovation process. Clearly, $\mathbb{K}$ provides only limited information about the actual shape of the distribution, and more meaningful results can be obtained by specifying the dynamics of the conditional variance, σ_t^2. A general useful result [167] for analyzing the tail behavior of processes such as (27) is that, if ξ_t and σ_t are independent non-negative random variables with σ_t regularly varying, i. e., $P(\sigma_t > x) = L(x)x^{-\alpha}$ for some slowly varying L, and $\mathbb{E}[\xi_t^{\alpha+\delta}] < \infty$ for some $\delta > 0$, then $\xi_t\sigma_t$ is likewise regularly varying with tail index α, namely,

$$P(\xi_t\sigma_t > x) \cong \mathbb{E}[\xi^\alpha] P(\sigma_t > x) \quad \text{as } x \to \infty . \quad (28)$$

Arguably the most popular model for the evolution of σ_t^2 in (27) is the generalized autoregressive conditional heteroskedasticity process of orders p and q, or GARCH(p, q), as introduced by [37,73], which specifies the conditional variance as

$$\sigma_t^2 = \alpha_0 + \sum_{i=1}^{q} \alpha_i \epsilon_{t-i}^2 + \sum_{i=1}^{p} \beta_i \sigma_{t-i}^2 . \quad (29)$$

The case $p = 0$ in (29) is referred to as an ARCH(q) process, which is the specification considered in [73]. To make sure that the conditional variance remains positive for all t, appropriate restrictions have to be imposed on the parameters in (29), i. e., α_i, $i = 0, \dots, q$, and β_i, $i = 1, \dots, p$. It is clearly sufficient to assume that α_0 is positive and all the other parameters are non-negative, as in [37], but these conditions can be relaxed substantially if $p, q > 0$ and $p + q > 2$ [173].

(27) and (29) is covariance stationary iff

$$P(z) = z^m - \sum_{i=1}^{m} (\alpha_i + \beta_i) z^{m-i} = 0 \Rightarrow |z| < 1 , \quad (30)$$

where $m = \max\{p, q\}$, and $\alpha_i = 0$ for $i > q$, and $\beta_i = 0$ for $i < p$, which boils down to $\sum_i \alpha_i + \sum_i \beta_i < 1$ in case the non-negativity restrictions of [37] are imposed. The situation $\sum_i \alpha_i + \sum_i \beta_i = 1$ is referred to as an integrated GARCH (IGARCH) model, and in applications it is often found that the sum is just below unity. This indicates a high degree of volatility persistence, but the interpretation of this phenomenon is not so clear-cut [166]. If (30) holds, the unconditional variance of the process de-

fined by (27) and (29) is given by

$$\mathbb{E}[\epsilon_t^2] = \frac{\alpha_0}{1 - \sum_{i=1}^{q} \alpha_i - \sum_{i=1}^{p} \beta_i} . \quad (31)$$

In practice, the GARCH(1,1) specification is of particular importance, and it will be the focus of our discussion too, i. e., we shall concentrate on the model (27) with

$$\sigma_t^2 = \alpha_0 + (\alpha_1 \eta_{t-1}^2 + \beta_1) \sigma_{t-1}^2 ,$$
$$\alpha_0 > 0 , \quad \alpha_1 > 0 , \quad 1 > \beta_1 \geq 0 . \quad (32)$$

The case $\alpha_1 = 0$ corresponds to a model with constant variance, which is of no interest in the current discussion.

An interesting property of the GARCH process is that its unconditional distribution is fat-tailed even with light-tailed (e. g., Gaussian) innovations, i. e., the distributional properties of the returns will not reflect those of the innovation (news) process. This has been known basically since [37,73], who showed that, even with normally distributed innovations, (G)ARCH processes do not have all their moments finite. For example, for the GARCH(1,1) model, [37] showed that, with $m \in \mathbb{N}$, the unconditional $(2m)$th moment of ϵ_t in (27) is finite if and only if

$$\mathbb{E}[(\alpha_1 \eta_t^2 + \beta_1)^m] < 1 , \quad (33)$$

which, as long as $\alpha_1 > 0$, will eventually be violated for all practically relevant distributions. The argument in [37] is based on the relation

$$\mathbb{E}[\sigma_t^{2m}] = \sum_{i=0}^{m} \binom{m}{i} \alpha_0^i \mathbb{E}[(\alpha_1 \eta_{t-1}^2 + \beta_1)^{m-i}]$$
$$\mathbb{E}[\sigma_{t-1}^{2(m-i)}] , \quad (34)$$

which follows from (32). The coefficient of $\mathbb{E}[\sigma_{t-1}^{2m}]$ on the right-hand side of (34) is just the expression appearing in (33), and consequently the $(2m)$th unconditional moment cannot be finite if this exceeds unity. The heavy-tailedness of the GARCH process is sometimes also exemplified by means of its unconditional kurtosis measure (3), which is finite for the GARCH(1,1) model with Gaussian innovations iff $3\alpha_1^2 + 2\alpha_1\beta_1 + \beta_1^2 < 1$. Writing (34) down for $m = 2$, using (31) and substituting into (3) gives

$$\mathbb{K}[\epsilon_t] = \frac{3[1 - (\alpha_1 + \beta_1)^2]}{1 - (\alpha_1 + \beta_1)^2 - 2\alpha_1^2} > 3 ,$$

as $\mathbb{E}[\epsilon_t^4] = 3\mathbb{E}[\sigma_t^4]$. [73] notes that "[m]any statistical procedures have been designed to be robust to large errors, but … none of this literature has made use of the

fact that temporal clustering of outliers can be used to predict their occurrence and minimize their effects. This is exactly the approach taken by the ARCH model". Conditions for the existence of and expressions for higher-order moments of the GARCH(p, q) model can be found in [50,105,122,139]. The relation between the conditional and unconditional kurtosis of GARCH models was investigated in [15], see also [47] for related results.

A more precise characterization of the tails of GARCH processes has been developed by applying classical results about the tail behavior of solutions of stochastic difference equations as, for example, in [124]. We shall continue to concentrate on the GARCH(1,1) case, which admits relatively explicit results, and which has already been written as a first-order stochastic difference equation in (32). Iterating this,

$$\sigma_t^2 = \sigma_0^2 \prod_{i=1}^{t} \left(\alpha_1 \eta_{t-i}^2 + \beta_1 \right)$$
$$+ \alpha_0 \left[1 + \sum_{k=1}^{t-1} \prod_{i=1}^{k} \left(\alpha_1 \eta_{t-i}^2 + \beta_1 \right) \right]. \quad (35)$$

Nelson [171] has shown that the GARCH(1,1) process (32) has a strictly stationary solution, given by

$$\sigma_t^2 = \alpha_0 \left[1 + \sum_{k=1}^{\infty} \prod_{i=1}^{k} \left(\alpha_1 \eta_{t-i}^2 + \beta_1 \right) \right], \quad (36)$$

if and only if

$$\mathbb{E} \left[\log \left(\alpha_1 \eta_t^2 + \beta_1 \right) \right] < 0 . \quad (37)$$

The keynote of the argument in [171] is the application of the strong law of large numbers to the terms of the form $\prod_{i=1}^{k} (\alpha_1 \eta_{t-i}^2 + \beta_1) = \exp\{\sum_{1}^{k} \log (\alpha_1 \eta_{t-i}^2 + \beta_1)\}$ in (35), revealing that (35) converges almost surely if (37) holds. Note that $\mathbb{E} \left[\log \left(\alpha_1 \eta_t^2 + \beta_1 \right) \right] < \log \mathbb{E} \left[\alpha_1 \eta_t^2 + \beta_1 \right]$ $= \log (\alpha_1 + \beta_1)$, i. e., stationary GARCH processes need not be covariance stationary. Using (36) and standard moment inequalities, [171] further established that, in case of stationarity, $\mathbb{E} \left[|\epsilon_t|^p \right]$, $p > 0$, is finite if and only if $\mathbb{E}[(\alpha_1 \eta_t^2 + \beta_1)^{p/2}] < 1$, which generalizes (33) to noninteger moments. It may now be supposed, and, building on the results of [90,124], has indeed been established by [167], that the tails of the marginal distribution of ϵ_t generated by a GARCH(1,1) process decay asymptotically in a Pareto-type fashion, i. e.,

$$P(|\epsilon_t| > x) \cong cx^{-\alpha} \quad \text{as } x \to \infty , \quad (38)$$

where the tail index α is the unique positive solution of the equation

$$h(\alpha) := \mathbb{E} \left[\left(\alpha_1 \eta_t^2 + \beta_1 \right)^{\alpha/2} \right] = 1 . \quad (39)$$

This follows from (28) along with the result that the tails of σ_t^2 and σ_t are asymptotically Paretian with tail indices $\alpha/2$ and α, respectively. For a discussion of technical conditions, see [167]. [167] also provides an expression for the constant c in (38), which is difficult to calculate explicitly, however. For the ARCH(1) model with Gaussian innovations, (39) becomes $(2\alpha_1)^{\alpha/2} \Gamma \left[(\alpha + 1)/2 \right] / \sqrt{\pi} = 1$, which has already been obtained by [63] and was foreshadowed in the work of [168]. The results reported above have been generalized in various directions, with qualitatively similar conclusions. The GARCH(p, q) case is treated in [29], while [140,141] consider various extensions of the standard GARCH(1,1) model.

Although the *unconditional* distribution of a GARCH model with Gaussian innovations has genuinely fat tails, it is often found in applications that the tails of empirical return distributions are even fatter than those implied by fitted Gaussian GARCH models, indicating that the *conditional* distribution, i. e., the distribution of η_t in (27), is likewise fat-tailed. Therefore, it has become standard practice to assume that the innovations η_t are also heavy tailed, although it has been questioned whether this is the best modeling strategy [199]. The most popular example of a heavy tailed innovation distribution is certainly the t considered in Subsect. "The Student t Distribution", which was introduced by [38] into the GARCH literature. Some authors have also found it beneficial to let the degrees of freedom parameter ν in (23) be time-varying, thus obtaining time-varying conditional fat-tailedness [45].

In the following, we shall briefly discuss a few GARCH(1,1) estimation results for the S&P500 series in order to compare the tails implied by these models with those from the semiparametric estimation procedures in Sect. "Empirical Evidence About the Tails". As distributions for the innovation process $\{\eta_t\}$, we shall consider the Gaussian, t, and the generalized error distribution (GED), which was introduced by [172] into the GARCH literature, see [128] for a recent contribution and asymmetric extensions. It has earlier been used in an unconditional context by [94] for the S&P500 returns. The density of the GED with mean zero and unit variance is given by

$$f_{\text{GED}}(x; \nu) = \frac{\lambda \nu}{2^{1/\nu+1} \Gamma(1/\nu)} \exp\left(-\frac{|\lambda x|^{\nu}}{2} \right), \quad \nu > 0,$$
$$\quad (40)$$

where $\lambda = 2^{1/\nu} \sqrt{\Gamma(3/\nu)/\Gamma(1/\nu)}$. Parameter ν in (40) controls the thickness of the tails. For $\nu = 2$, we get the normal distribution, and a leptokurtic shape is obtained for $\nu < 2$. In the latter case, the tails of (40) are therefore thicker than those of the Gaussian, but they are not fat in

the Pareto sense. However, even if one argues for Pareto tails of return distributions, use of (40) may be appropriate as a conditional distribution in GARCH models, because the power law already accompanies the volatility dynamics. To make the estimates of the parameter α_1 in (32) comparable, we also use the unit variance version of the t, which requires multiplying X in (23) by $\sqrt{(\nu - 2)/\nu}$. Returns are modeled as $r_t = \mu + \epsilon_t$, where μ is a constant mean and ϵ_t is generated by (27) and (32). Parameter estimates, obtained by maximum-likelihood estimation, are provided in Table 2. In addition to the GARCH parameters in (32) and the shape parameters of the innovation distributions, Table 2 reports the log likelihood values and the implied tail indices, $\hat{\alpha}$, which are obtained by solving (39) numerically. First note that all the GARCH models have considerably higher likelihood values than the iid models in Table 1, which highlights the importance of accounting for conditional heteroskedasticity. We can also conclude that the Gaussian assumption is still inadequate as a conditional distribution in GARCH models, as both the t and the GED achieve significantly higher likelihood values, and their estimated shape parameters indicate pronounced non-normalities. However, the degrees of freedom parameter of the t, ν, is somewhat increased in comparison to Table 1, as part of the leptokurtosis is now explained by the GARCH effects.

Compared to the nonparametric tail estimates obtained in Sect. "Empirical Evidence About the Tails", the tail index implied by the Gaussian GARCH(1,1) model turns out to be somewhat too high, while those of the more flexible models are both between 3 and 4 and therefore more in line with what has been found in Sect. "Empirical Evidence About the Tails". However, for all three models, the confidence intervals for α, as obtained from 1,000 simulations from the respective estimated GARCH processes,

are rather wide, so that we cannot conclusively rule out the existence of the unconditional fourth (and even fifth) moment. The width of the confidence intervals reflects the fact that the implied tail indices are very sensitive to small variations in the underlying GARCH parameters. For example, if, in the GARCH model with conditional normality, we replace the estimate $\hat{\beta}_1 = 0.911$ with 0.9, the implied tail index is 7.31, and with $\beta_1 = 0.92$, we get $\alpha = 2.05$, which is close to an infinite variance. The situation is depicted in Fig. 8, showing $h(\alpha)$ in (39) for the different values of β_1. The shape of h follows generally from $h(0) = 1$, $h'(0) < 0$ by (37), $h'' > 0$, i.e., h is convex, and $\lim_{\alpha \to \infty} h(\alpha) = \infty$ as long as $P\left[(\alpha_1 \eta_t^2 + \beta_1) > 1\right] > 0$, so that $h(\alpha) = 1$ has a unique positive solution. Note that both 0.9 and 0.92 are covered by $0.911 \pm 2 \times 0.009$, i.e., a 95% confidence interval for β_1. This shows that the GARCH-implied tail indices are rather noisy.

Alternatively, we may avoid precise assumptions about the distribution of the innovation process $\{\eta_t\}$ and rely on quasi maximum-likelihood results [138]. That is, we estimate the innovations by $\hat{\eta}_t = \hat{\epsilon}_t / \hat{\sigma}_t$, $t = 1, \ldots, 5, 550$, where $\{\hat{\sigma}_t\}$ is the sequence of conditional standard deviations implied by the estimated Gaussian GARCH model, and then solve the sample analogue of (39), i.e., $T^{-1} \sum_{t=1}^{T} (\hat{\alpha}_1 \hat{\eta}_t^2 + \hat{\beta}_1)^{\alpha/2} = 1$, a procedure theoretically justified in [32]. Doing so, we obtain $\hat{\alpha} = 2.97$, so that we recover the "universal cubic law". However, the 95% confidence interval, calculated from 1,000 GARCH simulations, where the innovation sequences are obtained by sampling with replacement from the $\hat{\eta}_t$-series, is (1.73, 4.80), which is still reconcilable with a finite fourth moment, and even with an infinite second moment. These results clearly underline the caveat brought out by [72] (p. 349), that "[t]here is no free lunch when it comes to [tail index] estimation".

Financial Economics, Fat-Tailed Distributions, Table 2
GARCH parameter estimates

Distribution	$\hat{\mu}$	$\hat{\alpha}_0$	$\hat{\alpha}_1$	$\hat{\beta}_1$	$\hat{\nu}$	$\hat{\alpha}$	Loglik
Normal	0.059	0.012	0.080	0.911	—	4.70	−7271.7
	(0.011)	(0.002)	(0.008)	(0.009)		(3.20, 7.22)	
GED	0.063	0.007	0.058	0.936	1.291	3.95	−7088.2
	(0.010)	(0.002)	(0.007)	(0.008)	(0.031)	(2.52, 6.95)	
Symmetric t	0.063	0.006	0.051	0.943	6.224	3.79	−7068.1
	(0.010)	(0.002)	(0.006)	(0.007)	(0.507)	(2.38, 5.87)	

Shown are maximum-likelihood estimation results for GARCH(1,1) models, as given by (27) and (32), with different assumptions about the distribution of the innovations η_t in (27). Standard errors for the model parameters and 95% confidence intervals for the implied tail indices, $\hat{\alpha}$, are given in parentheses. "Loglik" is the value of the maximized log likelihood function.
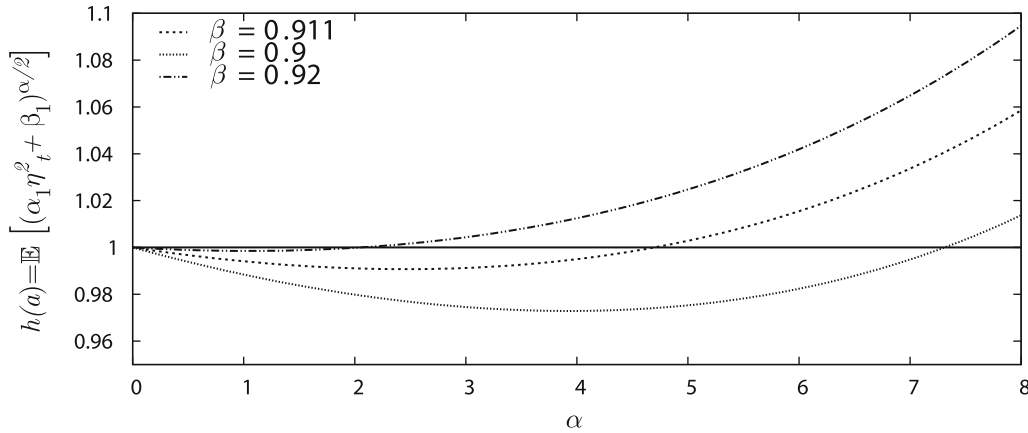
Financial Economics, Fat-Tailed Distributions, Figure 8
The figure displays the function $h(\alpha)$, as defined in (39), for Gaussian η_t, $\alpha = 0.0799$ and various values of β_1. Note that $\hat{\alpha}_1 = 0.0799$ and $\hat{\beta}_1 = 0.911$ are the maximum likelihood estimates for the S&P500 returns, as reported in Table 2

Application to Value-at-Risk

In this section, we compare the models discussed in Sects. "Some Specific Distributions" and "Volatility Clustering and Fat Tails" on an economic basis by employing the Value-at-Risk (VaR) concept, which is a widely used measure to describe the downside risk of a financial position both in industry and in academia [118]. Consider a time series of portfolio returns, r_t, and an associated series of ex-ante VaR measures with target probability ξ, $\mathrm{VaR}_t(\xi)$. The $\mathrm{VaR}_t(\xi)$ implied by a model $\mathcal{M}$ is defined by

$$\mathrm{Pr}_{t-1}^{\mathcal{M}}(r_t < -\mathrm{VaR}_t(\xi)) = \xi, \tag{41}$$

where $\mathrm{Pr}_{t-1}^{\mathcal{M}}(\cdot)$ denotes a probability derived from model $\mathcal{M}$ using the information up to time $t-1$, and the negative sign in (41) is due to the convention of reporting VaR as a positive number. For an appropriately specified model, we expect $100 \times \xi\%$ of the observed return values not to exceed the (negative of the) respective VaR forecast. Thus, to assess the performance of the different models, we examine the percentage shortfall frequencies,

$$U_\xi = 100 \times \frac{x}{T} = 100 \times \hat{\xi}, \tag{42}$$

where T denotes the number of forecasts evaluated, x is the observed shortfall frequency, i.e., the number of days for which $r_t < -\mathrm{VaR}_t(\xi)$, and $\hat{\xi} = x/T$ is the empirical shortfall probability. If $\hat{\xi}$ is significantly less (higher) than ξ, then model $\mathcal{M}$ tends to overestimate (underestimate) the risk of the position. In the present application, in order to capture even the more extreme tail region, we focus on the target probabilities $\xi = 0.001, 0.0025, 0.005, 0.01, 0.025,$ and 0.05.

To formally test whether a model correctly estimates the risk (according to VaR) inherent in a given financial position, that is, whether the empirical shortfall probability, $\hat{\xi}$, is statistically indistinguishable from the nominal shortfall probability, ξ, we use the likelihood ratio test [133]

$$\mathrm{LRT}_{\mathrm{VaR}} = -2\left\{ x \log\frac{\xi}{\hat{\xi}} + (T - x) \log\frac{1-\xi}{1-\hat{\xi}} \right\} \overset{\mathrm{asy}}{\sim} \chi^2(1). \tag{43}$$

On the basis of the first 1,000 return observations, we calculate one-day-ahead VaR measures based on parameter estimates obtained from an expanding data window, where the parameters are updated every day. Thus we get, for each model, 4,550 one-day-ahead out-of-sample VaR measures.

Table 3 reports the realized one-day-ahead percentage shortfall frequencies for the different target probabilities, ξ, as given above. The upper panel of the table shows the results for the unconditional distributions discussed in Sect. "Some Specific Distributions". The results clearly show that the normal distribution strongly *underestimates* ($\hat{\xi} > \xi$) the downside risk for the lower target probabilities, while the Cauchy as well as the alpha stable distributions tend to significantly *overestimate* ($\hat{\xi} < \xi$) the tails. This is in line with what we have observed from the empirical density plots presented in Fig. 7, which, in contrast to the out-of-sample VaR calculations, are based on estimates for the entire sample. Interestingly, the finite normal mixture distribution also tends to overestimate the risk at the lower VaR levels, leading to a rejection of cor-

Financial Economics, Fat-Tailed Distributions, Table 3
Backtesting Value-at-Risk measures

Unconditional Distributional Models						
Distribution	$U_{0.001}$	$U_{0.0025}$	$U_{0.005}$	$U_{0.01}$	$U_{0.025}$	$U_{0.05}$
GH	0.04	0.11**	0.24***	0.73*	2.70	5.89***
t_{GH}	0.07	0.11**	0.22***	0.75*	2.75	5.96***
t_{JF}	0.04	0.11**	0.31**	0.88	2.64	5.32
t_{AC}	0.04	0.11**	0.26**	0.84	2.48	5.16
t_{FS}	0.07	0.13*	0.33**	0.95	2.77	5.38
Symmetric t	0.07	0.15	0.31**	0.92	3.08**	6.35***
NIG	0.07	0.15	0.26**	0.70**	2.35	5.34
HYP	0.13	0.24	0.51	0.95	2.50	5.16
VG	0.13	0.24	0.51	0.92	2.46	5.10
Alpha stable	0.04	0.11**	0.33**	0.75*	2.44	4.90
Finite mixture ($k = 2$)	0.04	0.07***	0.11***	0.37***	2.99**	6.40***
Cauchy	0.00***	0.00***	0.00***	0.00***	0.09***	0.88***
Normal	0.48***	0.64***	0.97***	1.36**	2.44	4.02***

GARCH(1,1) Models						
Distribution	$U_{0.001}$	$U_{0.0025}$	$U_{0.005}$	$U_{0.01}$	$U_{0.025}$	$U_{0.05}$
Normal	0.40***	0.66***	0.92***	1.36**	2.95*	4.57
GED	0.20*	0.33	0.44	0.79	2.48	4.79
Symmetric t	0.11	0.26	0.40	0.92	2.86	5.45

The table shows the realized one-day-ahead percentage shortfall frequencies, U_ξ, for given target probabilities, ξ, as defined in (42). Asterisks *, ** and *** indicate significance at the 10%, 5% and 1% levels, respectively, as obtained from the likelihood ratio test (43).

rect coverage for almost all target probabilities. In contrast, the HYP distribution, whose empirical tails have been very close to those of the normal mixture in-sample (see Fig. 7), nicely reproduces the target probabilities, as does the VG distribution.

Similarly to the log likelihood results presented in Subsect. "Empirical Comparison" the Value-at-Risk evaluation does not allow for a clear distinction between the different t distributions, the GH and the NIG distribution. Similar to the Cauchy and the stable, they all tend to overestimate the more extreme target probabilities, while they imply too large shortfall probabilities at the five percent quantile.

The fact that most unconditional distributional models tend to overestimate the risk at the lower target probabilities may be due to our use of an expanding data window and the impact of the "Black Monday", where the index decreased by more than 20%, at the beginning of our sample period. In this regard, the advantages of accounting for time-varying volatility via a GARCH(1,1) structure may become apparent, as this model allows the more recent observations to have much more impact on the conditional density forecasts.

In fact, by inspection of the results for the GARCH models, as reported in the lower part of Table 3, it turns out that the GARCH(1,1) model with a normal distribution strongly underestimates the empirical shortfall probabilities at all levels except the largest (5%). However, considering a GED or t distribution for the return innovations within the GARCH model provides accurate estimates of downside risks.

To further discriminate between the GARCH processes and the iid models, tests for *conditional* coverage may be useful, which are discussed in the voluminous VaR literature (e. g., [53]).

Finally, we point out that the current application is necessarily of an illustrative nature. In particular, if the data generating process is not constant but evolves slowly over time and/or is subject to abrupt structural breaks, use of a rolling data window will be preferred to an expanding window.

Future Directions

As highlighted in the previous sections, there exists a plethora of different and well-established approaches

for modeling the tails of univariate financial time series. However, on the multivariate level the number of models and distributions is still very limited, although the joint modeling of multiple asset returns is crucial for portfolio risk management and allocation decisions. The problem is then to model the dependencies between financial assets. In the literature, this problem has been tackled, for example, by means of multivariate extensions of the mean-variance mixture (18) [19], multivariate GARCH models [30], regime-switching models [11], and copulas [51]. The problem is particularly intricate if the number of assets to be considered is large, and much work remains to understand and properly model their dependence structure.

It is also worth mentioning that the class of GARCH processes, due to its interesting conditional and unconditional distributional properties, has been adopted, for example, in the signal processing literature [3,52,55], and it is to be expected that it will be applied in other fields in the future.

Bibliography

Primary Literature

1. Aas K, Haff IH (2006) The generalized hyperbolic skew Student's t-distribution. J Financial Econom 4:275–309
2. Abhyankar A, Copeland LS, Wong W (1995) Moment condition failure in high frequency financial data: Evidence from the S&P 500. Appl Econom Lett 2:288–290
3. Abramson A, Cohen I (2006) State smoothing in Markov-switching time-frequency GARCH models. IEEE Signal Process Lett 13:377–380
4. Akgiray V, Booth GG (1988) The stable-law model of stock returns. J Bus Econ Stat 6:51–57
5. Alexander C, Lazar E (2006) Normal mixture GARCH(1,1). Applications to exchange rate modelling. J Appl Econom 21:307–336
6. Alexander SS (1961) Price movements in speculative markets: Trends or random walks. Ind Manag Rev 2:7–25
7. Andersen TG, Bollerslev T (1998) Answering the skeptics: Yes, standard volatility models do provide accurate forecasts. Int Econ Rev 39:885–905
8. Andersen TG, Bollerslev T, Diebold FX, Ebens H (2001) The distribution of realized stock return volatility. J Financial Econ 61:43–76
9. Andersen TG, Bollerslev T, Diebold FX, Labys P (2001) The distribution of realized exchange rate volatility. J Am Stat Assoc 96:42–55
10. Andersson J (2001) On the normal inverse Gaussian stochastic volatility model. J Bus Econ Stat 19:44–54
11. Ang A, Chen J (2002) Asymmetric correlations of equity portfolios. J Financial Econ 63:443–494
12. Atkinson AC (1982) The simulation of generalized inverse Gaussian and hyperbolic random variables. SIAM J Sci Stat Comput 3:502–515
13. Azzalini A, Capitanio A (2003) Distributions generated by perturbation of symmetry with emphasis on a multivariate skew t-distribution. J R Stat Soc Ser B 65:367–389
14. Bachelier L (1964) Theory of speculation. In: Cootner PH (ed) The random character of stock market prices. MIT Press, Cambridge, pp 17–75
15. Bai X, Russell JR, Tiao GC (2003) Kurtosis of GARCH and stochastic volatility models with non-normal innovations. J Econom 114:349–360
16. Balanda KP, MacGillivray HL (1988) Kurtosis: A critical review. Am Stat 42:111–119
17. Barndorff-Nielsen OE (1977) Exponentially decreasing distributions for the logarithm of particle size. Proc R Soc Lond Ser A 353:401–419
18. Barndorff-Nielsen OE (1988) Processes of normal inverse Gaussian type. Finance Stoch 2:41–68
19. Barndorff-Nielsen OE (1997) Normal inverse Gaussian distributions and stochastic volatility modelling. Scand J Stat 24:1–13
20. Barndorff-Nielsen OE, Halgreen C (1977) Infinite divisibility of the hyperbolic and generalized inverse Gaussian distributions. Probab Theory Relat Fields 38:309–311
21. Barndorff-Nielsen OE, Prause K (2001) Apparent scaling. Finance Stoch 5:103–113
22. Barndorff-Nielsen OE, Shephard N (2001) Normal modified stable processes. Theory Prob Math Stat 65:1–19
23. Barndorff-Nielsen OE, Shephard N (2002) Estimating quadratic variation using realized variance. J Appl Econom 17:457–477
24. Barndorff-Nielsen OE, Shephard N (2007) Financial volatility in continuous time. Cambridge University Press
25. Barndorff-Nielsen OE, Stelzer R (2005) Absolute moments of generalized hyperbolic distributions and approximate scaling of normal inverse Gaussian Lévy processes. Scand J Stat 32:617–637
26. Barndorff-Nielsen OE, Kent J, Sørensen M (1982) Normal variance-mean mixtures and z distributions. Int Stat Rev 50:145–159
27. Barndorff-Nielsen OE, Blæsild P, Jensen JL, Sørensen M (1985) The fascination of sand. In: Atkinson AC, Fienberg SE (eds) A celebration of statistics. Springer, Berlin, pp 57–87
28. Barnea A, Downes DH (1973) A reexamination of the empirical distribution of stock price changes. J Am Stat Assoc 68:348–350
29. Basrak B, Davis RA, Mikosch T (2002) Regular variation of GARCH processes. Stoch Process Appl 99:95–115
30. Bauwens L, Laurent S, Rombouts JVK (2006) Multivariate GARCH models: A survey. J Appl Econom 21:79–109
31. Beale EML, Mallows CL (1959) Scale mixing of symmetric distributions with zero mean. Ann Math Stat 30:1145–1151
32. Berkes I, Horváth L, Kokoszka P (2003) Estimation of the maximal moment exponent of a GARCH(1,1) sequence. Econom Theory 19:565–586
33. Bibby BM, Sørensen M (1997) A hyperbolic diffusion model for stock prices. Finance Stoch 1:25–41
34. Bingham NH, Goldie CM, Teugels JL (1987) Regular variation. Cambridge University Press
35. Blanchard OJ, Watson MW (1982) Bubbles, rational expectations, and financial markets. In: Wachtel P (ed) Crises in the economic and financial structure. Lexington Books, Lexington, pp 295–315

36. Blattberg RC, Gonedes NJ (1974) A comparison of the stable and Student distributions as statistical models for stock prices. J Bus 47:244–280

37. Bollerslev T (1986) Generalized autoregressive conditional heteroskedasticity. J Econom 31:307–327

38. Bollerslev T (1987) A conditionally heteroskedastic time series model for speculative prices and rates of return. Rev Econ Stat 69:542–547

39. Bollerslev T, Kretschmer U, Pigorsch C, Tauchen G (2007) A discrete-time model for daily S&P500 returns and realized variations: Jumps and leverage effects. J Econom (in press)

40. Boothe P, Glassman D (1987) The statistical distribution of exchange rates. J Int Econ 22:297–319

41. Bouchaud JP, Potters M (2000) Theory of financial risks. From statistical physics to risk management. Cambridge University Press, Cambridge

42. Box GEP, Muller ME (1958) A note on the generation of random normal deviates. Ann Math Stat 29:610–611

43. Branco MD, Dey DK (2001) A general class of multivariate skew-elliptical distributions. J Multivar Anal 79:99–113

44. Broca DS (2004) Mixture distribution models of Indian stock returns: An empirical comparison. Indian J Econ 84:525–535

45. Brooks C, Burke SP, Heravi S, Persand G (2005) Autoregressive conditional kurtosis. J Financial Econom 3:399–421

46. Campbell J, Lo AW, MacKinlay AC (1997) The econometrics of financial markets. Princeton University Press, Princeton

47. Carnero MA, Peña D, Ruiz E (2004) Persistence and kurtosis in GARCH and stochastic volatility models. J Financial Econom 2:319–342

48. Carr P, Geman H, Madan DB, Yor M (2002) The fine structure of asset returns: An empirical investigation. J Bus 75:305–332

49. Chambers JM, Mallows CL, Stuck BW (1976) A method for simulating stable random variables. J Am Stat Assoc 71:340–344

50. Chen M, An HZ (1998) A note on the stationarity and the existence of moments of the GARCH model. Stat Sin 8:505–510

51. Cherubini U, Luciano E, Vecchiato W (2004) Copula methods in finance. Wiley, New York

52. Cheung YM, Xu L (2003) Dual multivariate auto-regressive modeling in state space for temporal signal separation. IEEE Trans Syst Man Cybern B: 33:386–398

53. Christoffersen PF, Pelletier D (2004) Backtesting value-at-risk: A duration-based approach. J Financial Econom 2:84–108

54. Clark PK (1973) A subordinated stochastic process model with finite variance for speculative prices. Econometrica 41:135–155

55. Cohen I (2004) Modeling speech signals in the time-frequency domain using GARCH. Signal Process 84:2453–2459

56. Cont R, Tankov P (2004) Financial modelling with jump processes. Chapman & Hall, Boca Raton

57. Corsi F, Mittnik S, Pigorsch C, Pigorsch U (2008) The volatility of realized volatility. Econom Rev 27:46–78

58. Cotter J (2005) Tail behaviour of the euro. Appl Econ 37:827–840

59. Dacorogna MM, Müller UA, Pictet OV, de Vries CG (2001) Extremal forex returns in extremely large data sets. Extremes 4:105–127

60. Dagpunar JS (1989) An easily implemented generalised inverse Gaussian generator. Commun Stat Simul 18:703–710

61. Danielsson J, de Vries CG (1997) Tail index and quantile estimation with very high frequency data. J Empir Finance 4:241–257

62. Danielsson J, de Haan L, Peng L, de Vries CG (2001) Using a bootstrap method to choose the sample fraction in tail index estimation. J Multivar Anal 76:226–248

63. de Haan L, Resnick SI, Rootzén H, de Vries CG (1989) Extremal behaviour of solutions to a stochastic difference equation with applications to ARCH processes. Stoch Process Appl 32:213–234

64. de Vries CG (1994) Stylized facts of nominal exchange rate returns. In: van der Ploeg F (ed) The handbook of international macroeconomics. Blackwell, Oxford, pp 348–389

65. Doganoglu T, Mittnik S (1998) An approximation procedure for asymmetric stable Paretian densities. Comput Stat 13:463–475

66. DuMouchel WH (1973) On the asymptotic normality of the maximum-likelihood estimate when sampling from a stable distribution. Ann Stat 1:948–957

67. DuMouchel WH (1983) Estimating the stable index α in order to measure tail thickness: A critique. Ann Stat 11:1019–1031

68. Dyson FJ (1943) A note on kurtosis. J R Stat Soc 106:360–361

69. Eberlein E, von Hammerstein EA (2004) Generalized hyperbolic and inverse Gaussian distributions: Limiting cases and approximation of processes. In: Dalang RC, Dozzi M, Russo F (eds) Seminar on stochastic analysis, random fields and applications IV. Birkhäuser, Basel, pp 221–264

70. Eberlein E, Keller U (1995) Hyperbolic distributions in finance. Bernoulli 1:281–299

71. Eberlein E, Keller U, Prause K (1998) New insights into smile, mispricing, and value at risk: The hyperbolic model. J Bus 71:371–405

72. Embrechts P, Klüppelberg C, Mikosch T (1997) Modelling extremal events for insurance and finance. Springer, Berlin

73. Engle RF (1982) Autoregressive conditional heteroscedasticity with estimates of the variance of United Kingdom inflation. Econometrica 50:987–1006

74. Fama EF (1963) Mandelbrot and the stable Paretian hypothesis. J Bus 36:420–429

75. Fama EF (1965) The behavior of stock market prices. J Bus 38:34–105

76. Fama EF (1976) Foundations of finance. Basic Books, New York

77. Fama EF, Roll R (1971) Parameter estimates for symmetric stable distributions. J Am Stat Assoc 66:331–338

78. Farmer JD, Lillo F (2004) On the origin of power-law tails in price fluctuations. Quantit Finance 4:C7–C11

79. Feller W (1950) An introduction to probability theory and its applications I. Wiley, New York

80. Feller W (1971) An introduction to probability theory and its applications II. Wiley, New York

81. Fernández C, Steel MFJ (1998) On Bayesian modeling of fat tails and skewness. J Am Stat Assoc 93:359–371

82. Finucan HM (1964) A note on kurtosis. J R Stat Soc Ser B 26:111–112

83. Gabaix X, Gopikrishnan P, Plerou V, Stanley E (2007) A unified econophysics explanation for the power-law exponents of stock market activity. Physica A 382:81–88

84. Gabaix X, Gopikrishnan P, Plerou V, Stanley HE (2003) A theory of power-law distributions in financial market fluctuations. Nature 423:267–270

85. Gabaix X, Gopikrishnan P, Plerou V, Stanley HE (2006) Institutional investors and stock market volatility. Quart J Econ 121:461–504

86. Galbraith JW, Zernov S (2004) Circuit breakers and the tail index of equity returns. J Financial Econom 2:109–129

87. Gallant AR, Tauchen G (1996) Which moments to match? Econom Theory 12:657–681

88. Gerber HU, Shiu ESW (1994) Option pricing by Esscher transforms. Trans Soc Actuar 46:99–140

89. Ghose D, Kroner KF (1995) The relationship between GARCH and symmetric stable processes: Finding the source of fat tails in financial data. J Empir Finance 2:225–251

90. Goldie CM (1991) Implicit renewal theory and tails of solutions of random equations. Ann Appl Probab 1:126–166

91. Gopikrishnan P, Meyer M, Amaral LAN, Stanley HE (1998) Inverse cubic law for the distribution of stock price variations. Eur Phys J B 3:139–140

92. Gopikrishnan P, Plerou V, Amaral LAN, Meyer M, Stanley HE (1999) Scaling of the distribution of fluctuations of financial market indices. Phys Rev E 60:5305–5316

93. Gourieroux C, Jasiak J (1998) Truncated maximum likelihood, goodness of fit tests and tail analysis. Working Paper, CREST

94. Gray JB, French DW (1990) Empirical comparisons of distributional models for stock index returns. J Bus Finance Account 17:451–459

95. Gut A (2005) Probability: A graduate course. Springer, New York

96. Haas M, Mittnik S, Paolella MS (2004) Mixed normal conditional heteroskedasticity. J Financial Econom 2:211–250

97. Haas M, Mittnik S, Paolella MS (2004) A new approach to Markov-switching GARCH models. J Financial Econom 2: 493–530

98. Hagerman RL (1978) More evidence on the distribution of security returns. J Finance 33:1213–1221

99. Hall JA, Brorsen W, Irwin SH (1989) The distribution of futures prices: A test of the stable Paretian and mixture of normals hypotheses. J Financial Quantit Anal 24:105–116

100. Hall P (1982) On some simple estimates of an exponent of regular variation. J R Stat Soc Ser B 44:37–42

101. Hamilton JD (1989) A new approach to the economic analysis of nonstationary time series and the business cycle. Econometrica 57:357–384

102. Hansen BE (1994) Autoregressive conditional density estimation. Int Econ Rev 35:705–730

103. Harrison P (1996) Similarities in the distribution of stock market price changes between the eighteenth and twentieth centuries. J Bus 71:55–79

104. Hazan A, Landsman Z, Makov UE (2003) Robustness via a mixture of exponential power distributions. Comput Stat Data Anal 42:111–121

105. He C, Teräsvirta T (1999) Fourth moment structure of the GARCH(p, q) process. Econom Theory 15:824–846

106. Hill BM (1975) A simple general approach to inference about the tail of a distribution. Ann Stat 3:1163–1174

107. Hols MCAB, de Vries CG (1991) The limiting distribution of extremal exchange rate returns. J Appl Econom 6:287–302

108. Hsieh DA (1989) Modeling heteroskedasticity in daily foreign-exchange rates. J Bus Econ Stat 7:307–317

109. Hsu DA, Miller RB, Wichern DW (1974) On the stable Paretian behavior of stock-market prices. J Am Stat Assoc 69:108–113

110. Huisman R, Koedijk KG, Kool CJM, Palm F (2001) Tail-index estimates in small samples. J Bus Econ Stat 19:208–216

111. Huisman R, Koedijk KG, Kool CJM, Palm F (2002) The tail-fatness of FX returns reconsidered. Economist 150:299–312

112. Hyung N, de Vries CG (2005) Portfolio diversification effects of downside risk. J Financial Econom 3:107–125

113. Jansen DW, de Vries CG (1991) On the frequency of large stock returns: Putting booms and busts into perspective. Rev Econ Stat 73:18–24

114. Jensen MB, Lunde A (2001) The NIG-S&ARCH model: A fat-tailed, stochastic, and autoregressive conditional heteroskedastic volatility model. Econom J 4:319–342

115. Jondeau E, Rockinger M (2003) Testing for differences in the tails of stock-market returns. J Empir Finance 10:559–581

116. Jones MC, Faddy MJ (2003) A skew extension of the t-distribution, with applications. J R Stat Soc Ser B 65:159–174

117. Jørgensen B (1982) Statistical properties of the generalized inverse gaussian distribution. Springer, Berlin

118. Jorion P (2006) Value at risk: The new benchmark for controlling derivatives risk. McGraw-Hill, New York

119. Kaizoji T, Kaizoji M (2003) Empirical laws of a stock price index and a stochastic model. Adv Complex Syst 6:303–312

120. Kanter M (1975) Stable densities under change of scale and total variation inequalities. J Am Stat Assoc 3:697–707

121. Kaplansky I (1945) A common error concerning kurtosis. J Am Stat Assoc 40:259

122. Karanasos M (1999) The second moment and the autocovariance function of the squared errors of the GARCH model. J Econom 90:63–76

123. Kearns P, Pagan A (1997) Estimating the density tail index for financial time series. Rev Econ Stat 79:171–175

124. Kesten H (1973) Random difference equations and renewal theory for products of random matrices. Acta Math 131:207–248

125. Koedijk KG, Schafgans MMA, de Vries CG (1990) The tail index of exchange rate returns. J Int Econ 29:93–108

126. Koedijk KG, Stork PA, de Vries CG (1992) Differences between foreign exchange rate regimes: The view from the tails. J Int Money Finance 11:462–473

127. Kogon SM, Williams DB (1998) Characteristic function based estimation of stable distribution parameters. In: Adler RJ, Feldman RE, Taqqu MS (eds) A practical guide to heavy tails: Statistical techniques and applications. Birkhäuser, Basel, pp 311–335

128. Komunjer I (2007) Asymmetric power distribution: Theory and applications to risk management. J Appl Econom 22:891–921

129. Kon SJ (1984) Models of stock returns: A comparison. J Finance 39:147–165

130. Koponen I (1995) Analytic approach to the problem of convergence of truncated Lévy flights towards the Gaussian stochastic process. Phys Rev E 52:1197–1199

131. Koutrouvelis IA (1980) Regression-type estimation of the parameters of stable laws. J Am Stat Assoc 75:918–928

132. Küchler U, Neumann K, Sørensen M, Streller A (1999) Stock returns and hyperbolic distributions. Math Comput Model 29:1–15

133. Kupiec PH (1995) Techniques for verifying the accuracy of risk management models. J Deriv 3:73–84

134. Laherrère J, Sornette D (1998) Stretched exponential distributions in nature and economy: "fat tails" with characteristic scales. Eur Phys J B 2:525–539

135. Lau AHL, Lau HS, Wingender JR (1990) The distribution of stock returns: New evidence against the stable model. J Bus Econ Stat 8:217–223

136. Leadbetter MR, Lindgren G, Rootzén H (1983) Extremes and related properties of random sequences and processes. Springer, New York

137. LeBaron B (2001) Stochastic volatility as a simple generator of apparent financial power laws and long memory. Quantit Finance 1:621–631

138. Lee SW, Hansen BE (1994) Asymptotic theory for the GARCH(1,1) quasi-maximum likelihood estimator. Econom Theory 10:29–52

139. Ling S, McAleer M (2002) Necessary and sufficient moment conditions for the GARCH(r, s) and asymmetric power GARCH(r, s) models. Econom Theory 18:722–729

140. Liu JC (2006) On the tail behaviors of a family of GARCH processes. Econom Theory 22:852–862

141. Liu JC (2006) On the tail behaviors of Box–Cox transformed threshold GARCH(1,1) processes. Stat Probab Lett 76:1323–1330

142. Longin FM (1996) The asymptotic distribution of extreme stock market returns. J Bus 69:383–408

143. Longin FM (2005) The choice of the distribution of asset returns: How extreme value theory can help? J Bank Finance 29:1017–1035

144. Loretan M, Phillips PCB (1994) Testing the covariance stationarity of heavy-tailed time series. J Empir Finance 1:211–248

145. Lux T (1996) The stable Paretian hypothesis and the frequency of large returns: An examination of major German stocks. Appl Financial Econ 6:463–475

146. Lux T (2000) On moment condition failure in German stock returns: An application of recent advances in extreme value statistics. Empir Econ 25:641–652

147. Lux T (2001) The limiting extremal behaviour of speculative returns: An analysis of intra-daily data from the Frankfurt stock exchange. Appl Financial Econ 11:299–315

148. Lux T, Sornette D (2002) On rational bubbles and fat tails. J Money Credit Bank 34:589–610

149. Madan DB, Carr PP, Chang EC (1998) The variance gamma process and option pricing. Eur Finance Rev 2:79–105

150. Madan DB, Milne F (1991) Option pricing with v. g. martingale components. Math Finance 1:39–55

151. Madan DB, Seneta E (1990) The variance gamma (v. g.) model for share market returns. J Bus 63(4):511–524

152. Malevergne Y, Pisarenko V, Sornette D (2005) Empirical distributions of stock returns: Between the stretched exponential and the power law? Quantit Finance 5:379–401

153. Malevergne Y, Pisarenko V, Sornette D (2006) On the power of generalized extreme value (GEV) and generalized Pareto distribution (GDP) estimators for empirical distributions of stock returns. Appl Financial Econ 16:271–289

154. Mandelbrot B (1963) New methods in statistical economics. J Polit Econ 71:421–440

155. Mandelbrot B (1963) The variation of certain speculative prices. J Bus 36:394–419

156. Mandelbrot B (1967) The variation of some other speculative prices. J Bus 40:393–413

157. Mantegna RN, Stanley HE (1994) tic process with ultraslow convergence to a Gaussian: The truncated Lévy flight. Phys Rev Lett 73:2946–2949

158. Marsaglia G, Marshall AW, Proschan F (1965) Moment crossings as related to density crossings. J R Stat Soc Ser B 27:91–93

159. Mason DM (1982) Laws of large numbers for sums of extreme values. Ann Probab 10:754–764

160. Matia K, Amaral LAN, Goodwin SP, Stanley HE (2002) Different scaling behaviors of commodity spot and future prices. Phys Rev E 66:045103

161. Matia K, Pal M, Salunkay H, Stanley HE (2004) Scale-dependent price fluctuations for the Indian stock market. Europhys Lett 66:909–914

162. McCulloch JH (1997) Measuring tail thickness to estimate the stable index α: A critique. J Bus Econ Stat 15:74–81

163. McCulloch JH (1986) Simple consistent estimators of stable distribution parameters. Commun Stat Simul 15:1109–1136

164. McDonald JB (1996) Probability distributions for financial models. In: Maddala GS, Rao CR (eds) Handbook of statistics 14: Statistical methods in finance. Elsevier, Amsterdam, pp 427–461

165. McLachlan GJ, Peel D (2000) Finite mixture models. Wiley, New York

166. Mikosch T, Stărică C (2004) Nonstationarities in financial time series, the long-range dependence, and the IGARCH effects. Rev Econ Stat 86:378–390

167. Mikosch T, Stărică C (2000) Limit theory for the sample autocorrelations and extremes of a GARCH(1,1) process. Ann Stat 28:1427–1451

168. Milhøj A (1985) The moment structure of ARCH processes. Scand J Stat 12:281–292

169. Mittnik S, Rachev ST (1993) Modeling asset retunrs with alternative stable distributions. Econom Rev 12:261–330

170. Mizuno T, Kurihara S, Takayasu M, Takayasu H (2003) Analysis of high-resolution foreign exchange data of USD-JPY for 13 years. Physica A 324:296–302

171. Nelson DB (1990) Stationarity and persistence in the GARCH(1,1) model. Econom Theory 6:318–334

172. Nelson DB (1991) Conditional heteroskedasticity in asset returns: A new approach. Econometrica 59:347–370

173. Nelson DB, Cao CQ (1992) Inequality constraints in the univariate GARCH model. J Bus Econ Stat 10:229–235

174. Newcomb S (1980) A generalized theory of the combination of observations so as to obtain the best result. In: Stigler SM (ed) American contributions to mathematical statistics in the nineteenth century, vol. 2. Arno, New York, pp. 343–366

175. Nolan JP (1997) Numerical calculation of stable densities and distribution functions. Commun Stat Stoch Models 13:759–774

176. Officer RR (1972) The distribution of stock returns. J Am Stat Assoc 67:807–812

177. Omran MF (1997) Moment condition failure in stock returns: UK evidence. Appl Math Finance 4:201–206

178. Osborne MFM (1959) Brownian motion in the stock market. Oper Res 7:145–173

179. Peiró A (1994) The distribution of stock returns: International evidence. Appl Financial Econ 4:431–439

180. Perry PR (1983) More evidence on the nature of the distribution of security returns. J Financial Quantit Anal 18:211–221

181. Plerou V, Gopikrishnan P, Amaral LAN, Meyer M, Stanley HE (1998) Scaling of the distribution of price fluctuations of individual companies. Phys Rev E 60:6519–6529

182. Plerou V, Gopikrishnan P, Gabaix X, Stanley E (2004) On the origin of power-law fluctuations in stock prices. Quantit Finance 4:C11–C15

183. Pólya G, Szegö G (1976) Problems and theorems in analysis II. Springer, Berlin

184. Praetz PD (1972) The distribution of share price changes. J Bus 45:49–55

185. Prause K (1999) The generalized hyperbolic model: Estimation, financial derivatives, and risk measures. Ph D thesis, Albert-Ludwigs-Universität Freiburg i. Br.

186. Press SJ (1972) Estimation in univariate and multivariate stable distributions. J Am Stat Assoc 67:842–846

187. Raible S (2000) Lévy processes in finance: Theory, numerics, and empirical facts. Ph D thesis, Albert-Ludwigs-Universität Freiburg i. Br.

188. Resnick SI (1987) Extreme values, regular variation, and point processes. Springer, New York

189. Resnick SI (1997) Heavy tail modeling and teletraffic data. Ann Stat 25:1805–1849

190. Resnick SI, Stărică C (1998) Tail index estimation for dependent data. Ann Appl Probab 8:1156–1183

191. Rydberg TH (1999) Generalized hyperbolic diffusion processes with applications in finance. Math Finance 9:183–201

192. Samorodnitsky G, Taqqu MS (1994) Stable non-gaussian random processes: Stochastic models with infinite variance. Chapman & Hall, New York

193. Sato KI (1999) Lévy processes and infinitely divisible distributions. Cambridge University Press, Cambridge

194. Schoutens W (2003) Lévy processes in finance: Pricing financial derivatives. Wiley, New York

195. Seneta E (1976) Regularly varying functions. Springer, Berlin

196. Sigman K (1999) Appendix: A primer on heavy-tailed distributions. Queueing Syst 33:261–275

197. Silva AC, Prange RE, Yakovenko VM (2004) Exponential distribution of financial returns at mesoscopic time lags: A new stylized fact. Physica A 344:227–235

198. Silverman BW (1986) Density estimation for statistics and data analysis. Chapman & Hall, New York

199. Tsiakas I (2006) Periodic stochastic volatility and fat tails. J Financial Econom 4:90–135

200. Turner CM, Startz R, Nelson CR (1989) A Markov model of heteroskedasticity, risk, and learning in the stock market. J Financial Econ 25:3–22

201. Werner T, Upper C (2004) Time variation in the tail behavior of bund future returns. J Future Markets 24:387–398

202. Weron R (2001) Levy-stable distributions revisited: Tail index > 2 does not exclude the Levy-stable regime. Int J Mod Phys C 12:209–223

Books and Reviews

Bollerslev T, Engle RF, Nelson DB (1994) ARCH models. In: Engle RF, McFadden DL (eds) Handbook of econometrics IV. Elsevier, Amsterdam, pp 2959–3038

Borak S, Härdle W, Weron R (2005) Stable distributions. In: Cizek P, Härdle W, Weron R (eds) Statistical tools for finance and insurance. Springer, Berlin, pp 21–44

Janicki A, Weron A (1993) Simulation and chaotic behavior of α-stable stochastic processes. Dekker, New York

McCulloch JH (1996) Financial applications of stable distributions. In: Maddala GS, Rao CR (eds) Handbook of statistics 14: Statistical methods in finance. Elsevier, Amsterdam, pp 393–425

Mikosch T (2004) Modeling dependence and tails of financial time series. In: Finkenstädt B, Rootzén H (eds) Extreme values in finance, telecommunications, and the environment. Chapman & Hall, Boca Raton, pp 185–286

Mittnik S, Rachev ST, Paolella MS (1998) Stable paretian modeling in finance: Some empirical and theoretical aspects. In: Adler RJ, Feldman RE, Taqqu MS (eds) A practical guide to heavy tails: Statistical techniques and applications. Birkhäuser, Basel, pp 79–110

Nelsen RB (2006) An introduction to copulas, 2nd edn. Springer, New York

Pagan A (1996) The econometrics of financial markets. J Empir Finance 3:15–102

Palm FC (1996) GARCH models of volatility. In: Maddala GS, Rao CR (eds) Handbook of statistics 14. Elsevier, Amsterdam, pp 209–240

Rachev ST, Mittnik S (2000) Stable paretian models in finance. Wiley, Chichester

Zolotarev VM (1986) One-dimensional stable distributions. AMS, Providence

Financial Economics, Non-linear Time Series in

TERENCE C. MILLS[1], RAPHAEL N. MARKELLOS[1,2]
[1] Department of Economics, Loughborough University, Loughborough, UK
[2] Department of Management Science and Technology, Athens University of Economics and Business, Athens, Greece

Article Outline

Glossary

Arbitrage The possibility of producing a riskless profit by exploiting price differences between identical or linked assets.

Market efficiency A market is called efficient when all available information is reflected accurately, instantly and fully in the prices of traded assets. Depending on the definition of the available information set, private, public or that contained in historical prices, market efficiency is considered as strong, semi-strong or weak, respectively. The market price of an asset in an efficient market is an unbiased estimate of its true value. Sys-

tematic excess profits, which cannot be justified on the basis of the underlying risk, are not possible in such a market.

Martingale The term was originally used to describe a particular gambling strategy in which the stake is doubled following a losing bet. In probability theory it refers to a stochastic process that is a mathematical model of 'fair play'. This has been one of the most widely assumed processes for financial prices. It implies that the best forecast for tomorrow's price is simply today's price or, in other words, that the expected difference between any two successive prices is zero. Assuming a positive (negative) expected difference leads to the more general and realistic class of submartingale (supermartingale) processes. The martingale process implies that price differences are serially uncorrelated and that univariate linear time series models of prices have no forecasting value. However, martingales do not preclude the potential usefulness of nonlinear models in predicting the evolution of higher moments, such as the variance. The efficient market hypothesis is often incorrectly equated to the so-called random walk hypothesis, which roughly states that financial prices are martingales.

Option A call (put) option is a contractual agreement which gives the holder the right to buy (sell) a specified quantity of the underlying asset, within a specified period of time, at a price that is agreed when the contract is executed. Options are derivative assets since their value is based upon the variation in the underlying, which is typically the price of some asset such as a stock, commodity, bond, etc. Other basic types of derivatives include futures, forwards and swaps. An option is real, in contrast to financial, when the corresponding right refers to some business decision, such as the right to build a factory.

Portfolio theory The study of how resources should be optimally allocated between alternative investments on the basis of a given time investment horizon and a set of preferences.

Systematic risk Reflects the factors affecting all securities or firms in an economy. It cannot be reduced by diversification and it is also known as market risk. In the context of one of the most popular financial models, the Capital Asset Pricing Model (CAPM), systematic risk is measured by the beta coefficient.

Unsystematic risk This is the part of risk that is unique to a particular security or firm and can be reduced through diversification. This risk cannot be explained on the basis of fluctuations in the market as whole and it is also known as residual or idiosyncratic risk.

Volatility A measure of overall risk for an asset or portfolio which represents the sum of systematic and unsystematic risk. While several different approaches have been proposed for approximating this unobservable variable, the simplest one is based on the annualized standard deviation estimated using a historical sample of daily returns.

Definition of the Subject

Financial economics is the branch of economic science that deals with how groups of agents, such as households, firms, investors, creditors and economies as a whole, allocate and exchange financial resources in the context of markets. A wide variety of problems and applications fall within this broad subject area, including asset pricing, portfolio optimization, market efficiency, capital budgeting, interest and exchange rate modeling, risk management, forecasting and trading, market microstructure and behavioral finance. It is a highly quantitative and empirical discipline which draws its theoretical foundations and tools primarily from economics, mathematics and econometrics. Academic research in this area has flourished over the past century in line with the growing importance of financial markets and assets for the everyday life of corporations and individuals (for a historical overview of financial economics, see [69]). Consequently, at least 6 out of the 39 Nobel prizes in Economics have been awarded for research undertaken in areas related to financial economics. The close relationship between finance and time series analysis became widely apparent when Sir Clive W.J. Granger and Robert F. Engle III jointly received the 2003 Nobel Prize. Their work in time series econometrics has had a profound impact both on academic research and on the practice of finance. In particular, the ARCH model, first proposed by Engle [27] for modeling the variability of inflation, is today one of the most well known and important applications of a nonlinear time series model in finance. We should also acknowledge the Nobel Prize received by Robert C. Merton and Myron S. Scholes in 1993 for their pioneering work in the 1970s on pricing financial derivatives. In particular, they, along with Fischer Black, developed an analytical framework and simple mathematical formulae for pricing derivative assets, such as options and warrants, which have highly nonlinear payoff functions. Their work was the first step in the development of the derivatives industry and the whole risk management culture and practice in finance.

The close link between finance and nonlinear time series analysis is by no means accidental, being a consequence of four main factors. First, financial time series

have always been considered ideal candidates for data-hungry nonlinear models. The fact that organized financial markets and information brokers (e. g., newspapers, data vendors, analysts, etc.) have been around for many years has meant that an abundance of high quality historical data exists. Most of this data is in the form of time series and usually spans several decades, sometimes exceeding a century. Furthermore, asset prices can now be collected at ultra-high frequencies, often less than a minute, so that sample sizes may run into millions of observations. Second, the poor forecasting performance of linear models allied to the prospect of obtaining large financial gains by 'beating the market' on the basis of superior forecasts produced by nonlinear time series models has provided a natural motive for researchers from several disciplines. Third, developments in the natural sciences since the 1980s with respect to chaos theory, nonlinear dynamics and complexity have fueled a 'nonlinearist' movement in finance and have motivated a new research agenda on relevant theories, models and testing procedures for financial time series. Underlying this movement was the concern that the apparent unpredictability of financial time series may simply be due to the inadequacy of standard linear models. Moreover, it was also thought that the irregular fluctuations in financial markets may not be the result of propagated exogenous random shocks but, rather, the outcome of some, hopefully low-dimensional, chaotic system (see the entry by Shintani on ▶ Financial Forecasting, Sensitive Dependence). Fourth, and most importantly, although the bulk of financial theory and practice is built upon affine models, a wealth of theoretical models and supporting empirical evidence has been published suggesting that the nature of some financial problems may be inherently nonlinear. Two prime examples are the time-varying and asymmetric nature of financial risk and the highly nonlinear relationships that arise in situations involving financial options and other derivatives.

Introduction

Traditionally, theorists and empirical researchers in finance and economics have had rather different views concerning nonlinearity. Theorists have shown some interest in nonlinearities and have used them in a variety of different ways, such as first order conditions, multimodality, ceilings and floors, regime switching, multiple equilibria, peso problems, bandwagon effects, bubbles, prey-predator dynamics, time-varying parameters, asymmetries, discontinuities and jump behavior, non-additivity, non-transitivity, etc. Theories and structural models that have nonlinear elements can be found in most areas of finance

and economics (selective reviews with a focus mainly on economics are given by Brock and de Lima [19], Lorenz (see Chaps. 1–3 and 6 in [53]), Mullineux and Peng [67], Rosser [74]; other sources include Chap. 3 and pp. 114–147 in [40], [75]). Prominent examples include the noise-trader models of exchange rate determination [34,35], the target-zone exchange rate models [36,50] and the imperfect knowledge models [37]. Nonlinearities find their natural place in the theory of financial derivatives (for overviews, see [46,62]) and real options [26], where pay-off functions and relationships between pricing variables are inherently highly nonlinear. The popularity of nonlinearities is limited by the prevailing equilibrium theory assumptions (convexity and continuity conditions, concavity of utility and production functions, constant returns to scale, intertemporally independent tastes and technology, rational aggregate expectations and behavior, etc.) which invariably lead to linear relationships.

For many years, nonlinearities were not a serious consideration when attempting to build empirical models. Alfred Marshall, one of the great pioneers of mathematical economics, epitomized the culture against nonlinear models when saying that "*natura non facit saltum*", or nature dislikes jumps. Although he contemplated the possibility of multiple equilibria and switching behavior and understood that this situation would entail a tendency for stable and unstable equilibria to alternate, he dismissed it as deriving "*from the sport of imagination rather than the observation of facts*". Correspondingly, in empirical and theoretical finance the mainstream approach has been to transform any nonlinearities to linearized forms using Taylor series expansions which excluded second-and higher-order terms. Since the 1990s, however, there has been a significant turn in favor of nonlinear modeling in finance. In addition to the reasons advanced earlier, this development has also been the result of advances in econometric estimation and of the widespread availability of cheap computer power. Some of the basic nonlinear models and relevant theories that have been used in finance will be discussed in the subsequent section (for a comprehensive review of the linear and nonlinear time series models used in finance, see [22,63]).

Basic Nonlinear Financial Time Series Models

Most of the theoretical and empirical research in financial economics has typically hypothesized that asset price time series are unit root stochastic processes with returns that are serially unpredictable. For many years it was thought that this unpredictability was necessary in order to ensure that financial markets function properly according to

the Efficient Market Hypothesis (EMH; see the reviews by Fama [32,33]). Within this framework, a market is considered efficient with respect to a specific information set, an asset pricing model and a data generating process, respectively. For example, a general condition of efficiency is that market prices fully, correctly and instantaneously reflect the available information set. This is sometimes formalized as the Random Walk Hypothesis (RWH), which predicts that prices follow random walks with price changes that are unforecastable on the basis of past price changes. An even milder condition is that trading on the information set does not allow profits to be made at a level of risk that is inconsistent with the underlying asset pricing model. Although initially the EMH and RWH were thought to be an unavoidable consequence of the widely accepted paradigm of rational expectations, this was later refuted by a series of studies showing that random walk behavior was neither a necessary nor sufficient condition for rationally determined financial prices. Market efficiency has profound practical economic implications insofar as financial prices serve both as ways of integrating and distributing available information and as asset allocation devices.

One of the simplest models of financial prices that can be derived on the basis of unpredictability is the martingale process:

$$p_t = p_{t-1} + \varepsilon_t \qquad (1)$$

where p_t is the price of an asset observed at time t and ε_t is the martingale increment or martingale difference. The martingale has the following properties: a) $E(|p_t|) < \infty$ for each t, b) $E(p_t|\Im_s) = p_s$ whenever $s \leqslant t$, where $\Im_s$ is the σ-algebra comprized of events determined by observations over the interval $[0, t]$, so that $\Im_s \subseteq \Im_t$ when $s \leqslant t$. The martingale possesses the Markov property since the differences $\Delta p_t = p_t - p_{t-1} = \varepsilon_t$ are unpredictable on the basis of past differences. By successive backward substitution in (1) we can express the current price as the accumulation of all past errors. In financial terms, errors can be thought to be the result of unexpected information or news. By restricting the differences ε_t to be identically and independently distributed (iid) we obtain what is often called the random walk process. The random walk is a term and assumption which is widely employed in finance. It was first used by Karl Pearson in a letter to Nature in 1905 trying to describe a mosquito infestation in a forest. Soon after, Pearson compared the process to the walk of an intoxicated man, hence the graphical term "drunkard's walk".

By representing the random walk in continuous time with a growth rate μ, as is often useful when dealing with derivatives, we obtain the generalized Wiener process (also called Brownian motion or diffusion):

$$\mathrm{d}p_t = \mu \mathrm{d}t + \sigma \mathrm{d}w_t \qquad (2)$$

where $\mathrm{d}w_t$ is a standard normal random variable. The parameters μ and σ are referred to in finance as the drift and volatility of the process, respectively. Another point worth mentioning is that in both discrete and continuous time the analysis is typically undertaken using logarithmically transformed prices. This precludes the paradoxical possibility of obtaining negative prices while, at the same time, regularizing the statistical behavior of the data. Assuming that prices are lognormally distributed means that logarithmic returns are normally distributed and can be calculated as $\log p_t - \log p_{t-1}$ or $\log(p_t/p_{t-1})$. These represent continuously compounded returns and are approximately equal to simple percentage returns.

Random walks, along with continuous-time mathematical finance, were formally introduced in 1900 by Louis Bachelier in his brilliant doctoral dissertation *Théorie de la Spéculation*. Under the supervision of the great Henri Poincaré, who first realized the possibility of chaotic motion, Bachelier developed the mathematical framework of random walks in continuous time in order to describe the unpredictable evolution of stock prices and to build the first option pricing model (biographical details of Bachelier are given in [58]). Random walks were independently discovered by Albert Einstein in 1905 and, of course, have since played a fundamental role in physics and mathematics. They were later rigorously treated, along with forecasting and nonlinear modeling, by Norbert Wiener, the father of cybernetics. Several important deviations from the Bachelierian random walk and normal distribution paradigm were developed several decades later by Benoit Mandelbrot and his co-authors (for an overview see [59], and the references given therein). This research developed around the generalized Central Limit Theorem (CLT), the stable family of distributions, long-term dependence processes, scaling and fractals. Indeed, it is clear that Mandelbrot views his research as similar to that of Bachelier in that both were inspired by finance and both found great applications later in physics or, to use Mandelbrot's words, both were cases of the "unexpected historical primacy of financial economics over physics" (see p. 174 in [59]).

Much of the motivation behind nonlinear time series modeling in finance has to do with certain empirical characteristics, or stylized facts, which have been observed over the years across many financial assets, markets and time periods. Since these characteristics were not always consistent with a linear data generating process, nonlinear models seemed to be a reasonable explanation. In particular,
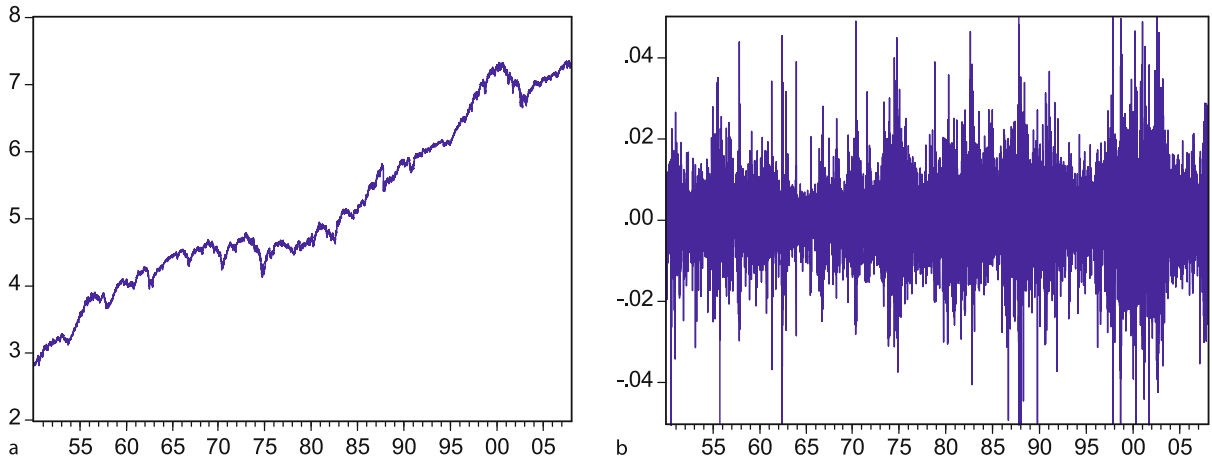
Financial Economics, Non-linear Time Series in, Figure 1
Daily S&P 500 log Index Prices *(left)* **and Returns** *(right)* **(3/1/1950–14/12/2007) (Returns are trimmed to ±5% in order to improve the readability of the graph)**

starting with Mandelbrot and others in the 1960s, several empirical studies have reported that financial assets typically have daily returns exhibiting:

- Nonnormality: skewed and leptokurtic (fat-tailed and high-peaked) unconditional distributions.
- Jump behavior: discontinuous variations that result in extreme observations.
- Volatility clustering: large (small) returns in magnitude tend to be followed by large (small) returns of either sign.
- Unpredictability: zero or weak serial autocorrelations in returns.

In order to illustrate these characteristics, we investigate the empirical behavior of daily logarithmic prices and returns (simply referred to as prices and returns hereafter) for the S&P 500 index. The series is publicly available from *Yahoo Finance*. The empirical analysis is undertaken using the econometric software packages *EViews* 5.0 by Quantitative Micro Software and *Time Series Modelling* 4.18 by James Davidson, respectively. The sample consists of 14,582 closing (end of the day) prices covering the period 3/1/1950–14/12/2007. The index is calculated as a weighted average of the common stock prices for the 500 largest firms traded on the New York Stock Exchange (NYSE) and is adjusted for dividends and splits. The S&P 500 is often used as a proxy for the so-called market portfolio and as a measure of the overall performance of the US stock market.

The prices depicted in the left part of Fig. 1 exhibit the upward drifting random walk behavior which is so rou-

tinely observed in financial time series. This is consistent with the fact that the series could not be predicted on the basis of past values using a member of the ARIMA class of linear models (as popularized by Box and Jenkins [16]). More specifically, the best fit was offered by an ARIMA(1,1,1) model, although with a disappointingly low R-squared statistic of just 0.64% (absolute t-statistics in brackets):

$$\Delta p_t = \underset{(3.9175)}{0.0003} - \underset{(3.3030)}{0.3013} \, \Delta p_{t-1} + \underset{(4.2664)}{0.3779} \, \varepsilon_{t-1} + \varepsilon_t \, . \quad (3)$$

Such weak linear serial predictabilities are often found at high sampling frequencies and are usually explained by market microstructures. They do not represent true predictabilities but, rather, result from specific market mechanisms and trading systems (see the survey by Biais et al. [12]).

A close examination of the return series, presented in the right part of Fig. 1, suggests the presence of large, discontinuous variations. In particular, we can count 26 daily returns which, in absolute value, exceed five standard deviations. This implies that such extreme events occur with a probability of 0.18% or, on average, almost once every two years (assuming 250 trading days in each calendar year). Under a normal distribution, such 'five-standard deviation' events should be extremely rare, with a probability of occurrence of only 0.00003%, or less than 3 days in every 40,000 years! The fat tails of the return distribution are also reflected in the kurtosis coefficient of 37.3, which is much larger than the value of 3 that corresponds to a normal distribution. In terms of asymmetry, the distribution

is skewed to the left with the relevant coefficient estimated at -1.3.

Clearly, the normal distribution provides a poor approximation of reality here: the distribution of the errors in the random process described by (1) should be allowed to follow some non-Gaussian, fat-tailed and possibly skewed distribution (see the entry by Haas and Pigorsch on ▶ Financial Economics, Fat-Tailed Distributions). Various distributions having these properties have been proposed, including the Student-t, the mixture of normals, double Weibull, generalized beta, Tukey's $g \times h$, generalized exponential, asymmetric scale gamma, etc. (see [48,71]). Although some of the non-Gaussian distributions that have been proposed have many desirable properties, empirical evidence regarding their appropriateness for describing financial returns has been inconclusive. Moreover, these distributions often bring with them acute mathematical problems in terms of representation, tractability, estimation, mixing and asymptotics. A distribution that has received considerable attention is the stable family (also known as the stable Paretian, Pareto–Lévy, or Lévy flight), which was initially proposed by Mandelbrot [54,55]. Stable distributions are highly flexible, have the normal and Cauchy as special cases, and can represent 'problematic' empirical densities that exhibit asymmetry and leptokurtosis. Furthermore, they are consistent with stochastic behavior that is characterized by discontinuities or jumps. From a theoretical point of view, stable distributions are particularly appealing since they are the limiting class of distributions in the generalized CLT, which applies to scaled sums of iid random variables with infinite variances. Stable distributions also exhibit invariance under addition, a property that is important for financial data, which are usually produced as the result of time aggregation. For a comprehensive discussion of these distributions, see [54,55,64,65,71,76].

In terms of the conditional distribution, it is evident from the graph of returns that the variance is not homogeneous across time, as one would expect for an iid process. In line with this observation, the autocorrelation of squared or absolute returns suggest the presence of strong dependencies in higher moments, something that in turn is indicative of conditional heteroskedasticity (see Fig. 3 below). On the basis of the above, it appears that the simple random walk model is far too restrictive and that the more general martingale process provides a better approximation to the data. Unlike the random walk, the martingale rules out any dependence in the conditional expectation of Δp_{t+1} on the information available at t, while allowing dependencies involving higher conditional moments of Δp_{t+1}. This property of martingales is very useful for explaining clusters in variances, since it allows persistence (correlation) in the conditional variances of returns.

It should be noted that much of the empirical work on nonlinear financial time series has involved modeling time varying variances. This concentration on the variance stems from it being the most widely used measure of risk, which, in orthodox finance theory, is the sole determinant of the expected return of any asset. Knowing the expected return enables the opportunity cost of any investment or asset to be estimated and, ultimately, to have a fair price put on its value by discounting all future revenues against the expected return. Variance was introduced in the pathbreaking research of Nobel Laureate Harry Markowitz in the 1950s on investment portfolio selection, which laid the basis for what is known today as modern portfolio theory. The main innovation of Markowitz was that he treated portfolio selection as a tractable, purely quantitative problem of utility maximization under uncertainty, hence the term 'quant analysis'. Markowitz assumed that economic agents face a choice over two-dimensional indifference curves of investment preferences for risk and return. Under some additional assumptions, he obtained a solution to this problem and described the preferences of homogeneous investors in a normative manner using the mean and variance of the probability distribution of single period returns: such investors should optimize their portfolios on the basis of a 'mean-variance' efficiency criterion, which yields the investment with the highest expected return for a given level of return variance.

Let us now turn to some of the processes that have been used to model regularities in variance. For example, consider the GARCH(1,1) process, which has become very popular for modeling the conditional variance, σ_t^2, as a deterministic function of lagged variances and squared errors (see the entry by Hafner on ▶ GARCH Modeling):

$$\sigma_t^2 = \alpha_0 + \alpha_1 \varepsilon_{t-1}^2 + \beta_1 \sigma_{t-1}^2 \qquad (4)$$

where the ε_t are, in general, the residuals from a fitted conditional mean equation. This specification corresponds to a single-lagged version of the GARCH(p, q) (Generalized Autoregressive Conditional Heteroskedasticity) model proposed by Bollerslev [13] and can easily be modified to include additional lagged squared errors and variances. The GARCH model is an extension of the ARCH process originally proposed by Engle [27] and has served as the basis for the development of an extensive family of related models. For a review of this huge literature see, among others, [10,14,15,52,79], and Chap. 5 in [63]. Multivariate extensions of GARCH processes have also been proposed, but bring several computational and estimation problems (see [9,21]).

Two alternative approaches to modeling conditional variances in finance are extreme value estimators (see [17]) and realized variance (see [8]). Extreme value estimators depend on opening, closing, high and low prices during the trading day. Although they perform relatively well in terms of efficiency and are easy to estimate, they are quite badly biased. Realized variances are considered to be very accurate and are easily estimated as the sum of squared returns within a fixed time interval. Their limitation is that they require high frequency data at the intradaily level, which can be strongly affected by market microstructures and may not always be readily available. Rather than focusing on just the conditional variance, models have also been proposed for higher moments, such as conditional skewness and kurtosis (e. g., [43,47]).

To illustrate the application of some of the most popular GARCH parameterizations, consider again the S&P 500 return series. Using Maximum Likelihood (ML) estimation with t-student errors, the following 'GARCH(1,1)-in-Mean' (GARCH-M) model was obtained (absolute z-statistics appear in brackets):

$$\Delta p_t = \underset{(10.0994)}{0.0785} \sigma_t + \varepsilon_t$$

$$\sigma_t^2 = \underset{(6.5616)}{5.77 \cdot 10^{-7}} + \underset{(16.3924)}{0.0684} \varepsilon_{t-1}^2 + \underset{(218.0935)}{0.9259} \sigma_{t-1}^2 \,.$$

In this model, originally proposed by Engle et al. [30], returns are positively related to the conditional standard deviation, σ_t. This is a particularly useful specification since it is directly consistent with Markowitz's theory about the positive relation between expected return and risk. In particular, the slope coefficient in the conditional mean equation can be interpreted as a relative risk aversion parameter, measuring how investors are compensated by higher returns for bearing higher levels of risk.

It is instructive to show in Fig. 2 both the estimated GARCH(1,1)-M conditional standard deviations and the standardized residuals, ε_t/σ_t. On the one hand, the model clearly produces mean reversion in volatility, which resembles the empirical behavior observed in the original series. Although the estimated conditional variance process appears to be highly persistent, it is nevertheless stationary since the sufficiency condition is satisfied because $\alpha_1 + \beta_1 = 0.0684 + 0.9259 = 0.9943 < 1$. On the other hand, the standardized residuals have a far more homogeneous conditional volatility than the original series and more closely resemble a white noise process. Moreover, the standardized residuals are closer to a normal distribution, with a kurtosis coefficient of 7.7, almost five times smaller than that of the original return series.

Careful inspection of the relationship between returns and conditional variance often reveals an asymmetric relationship. Threshold GARCH (TGARCH) and Exponential GARCH (EGARCH) are two of the specifications often used to model this commonly encountered nonlinearity. These models were estimated using the ML approach and the following conditional variance specifications were obtained.

TGARCH

$$\sigma_t^2 = \underset{(8.5110)}{7.50 \cdot 10^{-7}} + \underset{(6.3766)}{0.0259} \varepsilon_{t-1}^2$$

$$+ \underset{(13.1058)}{0.0865} \varepsilon_{t-1}^2 \, g + \underset{(224.9279)}{0.9243} \sigma_{t-1}^2$$
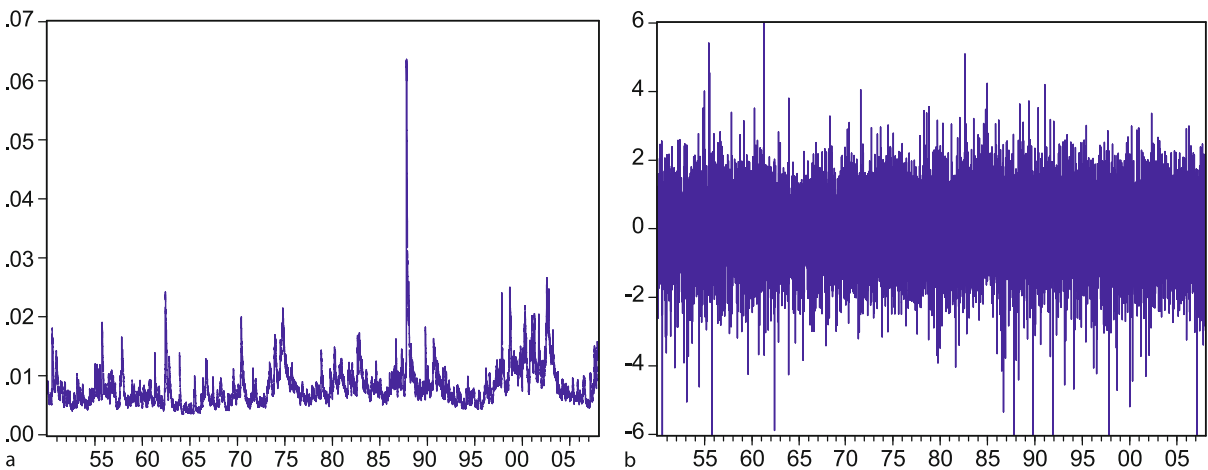
Financial Economics, Non-linear Time Series in, Figure 2
GARCH(1,1)-M standard deviations *(left)* **and standardized residuals** *(right)* **(Residuals are trimmed to ±6 standard deviations in order to improve the readability of the graph)**

EGARCH

$$\log(\sigma_t^2) = \underset{(13.7858)}{-0.2221} + \underset{(16.9612)}{0.1209} \left| \varepsilon_{t-1}/\sigma_{t-1} \right|$$

$$\underset{(15.9599)}{- 0.0690}\,\varepsilon_{t-1}/\sigma_{t-1} + \underset{(703.7161)}{0.9864}\,\log(\sigma_{t-1}^2)\,.$$

In the TGARCH model, the threshold parameter is defined as $g = 1$ if $\varepsilon_{t-1} < 0$ and 0 otherwise. Standard GARCH models, such as the GARCH-M estimated previously, assume that positive and negative errors (or news) have a symmetric effect on volatility. In the TGARCH and EGARCH models, news has an asymmetric effect on volatility depending on its sign. Specifically, in the TGARCH model news will have differential impacts on volatility depending on the signs and sizes of the coefficients on ε_{t-1}^2 and $\varepsilon_{t-1}^2 \cdot g$: good news ($\varepsilon_{t-1} > 0$) has an impact of 0.0259, while bad news ($\varepsilon_{t-1} < 0$) has a stronger impact of $0.0259 + 0.0865 = 0.1124$. Since the coefficient of $\varepsilon_{t-1}^2 \cdot g$ is positive (0.0865), bad news tends to increase volatility, producing what is known as the 'leverage' effect. This was first observed in the 1970s and postulates that negative returns will usually reduce the stock price and market value of the firm, which in turn means an increase in leverage, i. e. a higher debt to equity ratio, and ultimately an increase in volatility. In the EGARCH model, forecasts are guaranteed to be positive since logarithms of the conditional variance are modeled. Since the sign of the coefficient on $\varepsilon_{t-1}/\sigma_{t-1}$ is non-zero and negative we can conclude that the effect of news on volatility is asymmetric and that a leverage effect is present.

Inspection of the autocorrelation functions (ACFs) in Fig. 3 for the returns and absolute returns of the S&P 500, the latter being a proxy for volatility, suggests very different behavior of the two series. While returns have an ACF

that is typical of a white noise process, the autocorrelations of the absolute returns die out very slowly and become negative only after 798 lags! It turns out that many financial series have such extremely persistent or long-memory behavior. This phenomenon was first described by Mandelbrot [56,57] in the context of the 'Hurst effect' and was latter defined as fractional Brownian motion (see the relevant review by Brock [18]). Hosking [45] and Granger and Joyeux [39] modeled long-memory by extending the ARIMA class of processes to allow for fractional unit roots (for reviews, see [3,11,73,82]). The ARFIMA(p, d, q) model uses a fractional difference operator based on a binomial series expansion of the parameter d for any value between -0.5 and 0.5:

$$\Delta^d = 1 - dB + \frac{d(d-1)}{2!}B^2 - \frac{d(d-1)(d-2)}{3!}B^3 + \dots \tag{5}$$

where B is the backshift (or lag) operator with $B^m x_t = x_{t-m}$. In a similar fashion, investigating the existence of long-memory in the conditional variance of the returns could be undertaken in the context of a Fractional GARCH model (see [4]).

In our S&P 500 example, we have shown that nonlinearities enter through the conditional variance process and do so in an asymmetric manner. A natural question to ask is whether nonlinearities also exist in the conditional mean. Consider, for example, a generalization of the linear ARMA process

$$\Delta p_t = f(\Delta p_{t-i}, \varepsilon_{t-i}) + \varepsilon_t \tag{6}$$

where $f()$ is a nonlinear function and $\Delta p_{t-i}, \varepsilon_{t-i}$ are lagged price differences and errors, respectively. A wide
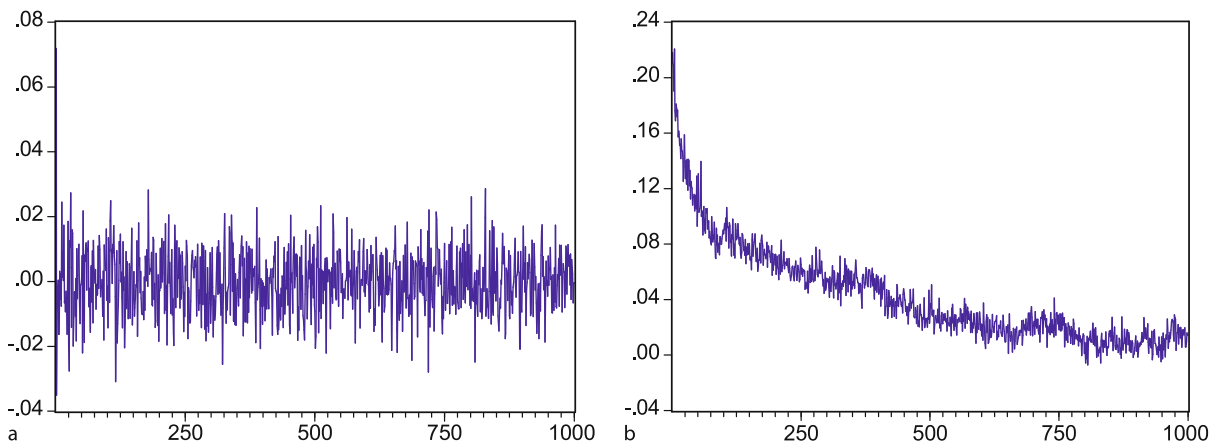
Financial Economics, Non-linear Time Series in, Figure 3
Autocorrelation function of S&P 500 simple *(left)* and absolute returns *(right)*

variety of testing procedures have been proposed for examining the possibility of nonlinearities in the conditional mean process (for reviews see the relevant sections in [40], and [63]). Here we use the BDS test of the null hypothesis of serial independence, which has been widely applied and has been shown to have good power against a variety of nonlinear alternatives (see [20]). The test is inspired by chaos theory and phase space analysis and is based on the concept of the correlation integral. Specifically, the test relies on the property that, for an iid series, the probability of the distance between any two points being no greater than a predefined distance (ε) should be constant. A joint probability can also be calculated for sets comprising multiple pairs of points chosen by moving through consecutive sequences of observations in the sample. The number of consecutive data points used in such a set is called the (embedding) dimension and may be chosen by the user. Brock et al. [20] constructed an asymptotically normally distributed test statistic for the constancy of the distance ε between points. When this test was applied to the residuals from an MA(1)-EGARCH(1,1) model fitted to the S&P 500 returns, it was always insignificant across a variety of dimensions, implying that any nonlinear dependencies in the returns are due solely to GARCH effects.

An agnostic, yet often convenient, way to approximate the unknown nonlinear function (6) is to consider some nonparametric estimator (see [72]). Although several nonparametric estimators have been used with mixed success, one of the most popular is based on the neural network family of models (see [80]). A rich variety of parametric nonlinear functions have also been proposed in finance. A convenient and intuitive way of introducing nonlinearity is to allow 'regime switching' or 'time-variation' in the parameters of the data generating process (for a review see [70]). Three of the most popular approaches in this category are the Markov switching, the Threshold Autoregressive (TAR) and the Smooth Transition (STAR) models. In the first approach (for a popular implementation, see [41,42]), the model parameters switch according to a multiple (typically two) unobserved state Markov process. In TAR models (see [81], for a comprehensive description), nonlinearities are captured using piecewise autoregressive linear models over a number of different states. For example, consider the simple two regime case:

$$x_t = \begin{cases} \omega_1 + \sum_{i=1}^{p} \varphi_{1i} x_{t-i+1} + \sigma_1 \varepsilon_t, & s_{t-d} < c \\ \omega_2 + \sum_{i=1}^{p} \varphi_{2i} x_{t-i+1} + \sigma_2 \varepsilon_t, & s_{t-d} \geqslant c \end{cases} \quad (7)$$

where c is the threshold value, s_t is a threshold variable, d is a delay parameter assumed to be less than or equal to p, and the ε_t are iid standard normal variates assumed to

be independent of lagged s_ts. The threshold variable is often determined by a linear combination of the lagged x_ts, in which case we obtain the Self Exciting TAR (SETAR) model. This has become a popular parameterization in finance since it can produce different dynamic behavior across regimes with characteristics such as asymmetry, limit cycles, jumps and time irreversibility (recall the TGARCH model introduced earlier, which has a related specification). STAR models allow a smooth switch between regimes using a smooth transition function. Transition functions that have been considered include the cumulative distribution of the standard normal, the exponential (ESTAR) and the logistic (LSTAR).

It is instructive to see how regime switching can be applied in the context of asset pricing models (for a comprehensive treatment of asset pricing, see [24]). The best known and most influential framework, which builds upon Markowitz's portfolio theory, is the Capital Asset Pricing Model (CAPM) proposed by Sharpe, Lintner, Black and others. The CAPM can be expressed as a single-period equilibrium model:

$$E(r_i) = r_{\mathrm{f}} + \beta_i \left[E(r_{\mathrm{m}}) - r_{\mathrm{f}} \right] \quad (8)$$

where $E(r_i)$ is the expected return on asset i, $E(r_{\mathrm{m}})$ is the expected return on the market portfolio, r_{f} is the risk-free interest rate, and the slope β_i is the so-called beta coefficient of asset i, measuring its systematic risk. Empirical implementations and tests of the CAPM are usually based on the 'excess market' and 'market model' regressions, respectively

$$r_{i,t} - r_{\mathrm{f},t} = r_{\mathrm{f},t} + \beta_i \left[r_{\mathrm{m},t} - r_{\mathrm{f},t} \right] + \varepsilon_{i,t} \quad (9)$$

and

$$r_{i,t} = \alpha_i + \beta_i r_{\mathrm{m},t} + \varepsilon_{i,t} . \quad (10)$$

The variance of the residuals $\varepsilon_{i,t}$ reflects the unsystematic risk in asset i. In practice the CAPM is typically estimated using ordinary least squares regression with five years of monthly data. A wealth of empirical evidence has been published showing that the basic assumptions of the CAPM regressions with respect to parameter stability and residual iid-ness are strongly refuted (see [60]). In particular, betas have been found to be persistent but unstable over time due to factors such as stock splits, business cycle conditions, market maturity and other political and economic events. In order to demonstrate the modeling of time-varying betas in the CAPM, consider first the simple market model regression for the stock returns of Tiffany & Co (listed on the New York Stock Exchange) against S&P

500 returns:

$$r_t = \underset{(17.5396)}{1.4081} r_{\mathrm{m},t} + \varepsilon_t, \quad R^2 = 28.75\% \,.$$

The regression was estimated using weekly returns from 30/12/1987 to 14/12/2007, a total of 1,044 observations. The R^2 statistic denotes the proportion of total risk that can be explained by the model and which is thus systematic. The beta coefficient is significantly higher than unity, suggesting that the stock is 'aggressive' in that it carries more risk than the market portfolio. Allowing the beta coefficient to switch according to a Markov process produces the following two-regime market model:

$$r_t = \begin{cases} \text{Regime 1:} & \underset{(1.9220)}{0.4797} r_{\mathrm{m},t} + \varepsilon_t \\ \text{Regime 2:} & \underset{(10.6381)}{1.9434} r_{\mathrm{m},t} + \varepsilon_t \end{cases} \quad R^2 = 41.63\% \,.$$

The explanatory power of the model has increased significantly and the stock is now characterized by both passive ($\beta = 0.4797 < 1$) and aggressive ($\beta = 1.9434 > 1$) systematic risk behavior regimes. The Markov transition probabilities $P(i \mid j)$, $j = 1, 2$, were estimated as $P(1|1) = 0.6833$, $P(1|2) = 0.3167$, $P(2|1) = 0.2122$ and $P(2|2) = 0.7878$. The smoothed probabilities for regime 1 are depicted in Fig. 4 and are seen to be rather volatile, so that the returns switch regimes rather frequently. For a discussion of the threshold CAPM see [2].

Another important category of models allows for nonlinear relationships between persistent financial time series. The most popular framework here is that of cointegration, which deals with variables that are individually nonstationary but have some joint stationary representation. For example, consider the linear combination of two unit root ($I(1)$) processes x_t and y_t

$$x_t = a + y_t + \varepsilon_t \,. \tag{11}$$

In general, ε_t will also be $I(1)$. However, as shown by Engle and Granger [29], if ε_t is actually $I(0)$, then x_t and y_t are said to be (linearly) cointegrated and will have an error-correction representation which, for example, could take the form

$$\Delta x_t = -\gamma \varepsilon_{t-1} + u_t \tag{12}$$

where $-\gamma$ denotes the strength of reversal to the equilibrium cointegrating relationship through the error-correction term, i. e., the lagged residual from the cointegrating regression (11). The finance literature has considered nonlinear generalizations of both the cointegrating regression (11) and the error-correction model (12) (see the entry by Escribano et al. on ► Econometrics: Non-linear Cointegration). Nonlinear error-correction mechanisms can be accommodated rather straightforwardly within the cointegration analysis framework, with the residuals from a linear cointegration relationship entering a nonlinear error-

Financial Economics, Non-linear Time Series in, Figure 4
Tiffany stock Markov switching market model smoothed probabilities for Regime 1 of 2

correction model. It has been shown that such nonlinearities may arise simply because of complex relationships between variables (see pp. 59–61 in [40]). Justifications in terms of finance theory have been based on factors such as arbitrage in the presence of transaction costs, heterogeneity among arbitrageurs, existence of partial adjustment models and market segmentation, agents' maximizing or minimizing behavior, constraints on central bank intervention, and intertemporal choice behavior under asymmetric adjustment costs. While almost all the different nonlinear specifications discussed previously have also been applied in error-correction modeling, threshold models hold a prominent position, as they allow large errors from equilibrium, i. e., those above some threshold, to be corrected while small errors are ignored (see, for example, [6]). The use of nonlinearities directly within the cointegrating relationship is not as straightfoward and brings several conceptual and estimation problems (see [63]).

Returning to the bivariate market model setting, it has been found that cointegrating relationships do exist between stock prices and index levels (see [60]). In our example, the logarithms of Tiffany's stock prices are cointegrated with S&P 500 logarithmic price levels. The following asymmetric error correction model was then estimated:

$$r_t = -0.0168 \, \varepsilon_{t-1} \, g + u_t$$
$$ \underset{(3.0329)}{}$$

where g is the heaviside function defined previously with $g = 1$ if $\varepsilon_{t-1} < 0$ and 0 otherwise, ε_{t-1} being obtained from the cointegrating regression.

Several studies have shown that empirical characteristics and regularities, such as those discussed previously are very unlikely to remain stable if the sampling frequency of the data changes. For example, we find that if the S&P 500 returns are estimated at an annual frequency using the first available January price, then their distribution becomes approximately Gaussian with skewness and kurtosis coefficients estimated at -0.4 and 2.7, respectively. The annual prices are highly predictable using an ARIMA(2,1,2) process with an impressive adjusted R-squared value of 15.7%. Moreover, standard tests of heteroskedasticity suggest that the variance of annual returns can be assumed to be constant! In contrast, for very high sampling frequencies, say at the intradaily or tick-by-tick level, the data behave in a different manner and are characterized by strong seasonalities, e. g., variances and volumes follow an inverse J shape throughout the trading day (see the review by Goodhart and O'Hara [38], and the discussion in [28]).

Finally, let us now turn our discussion to models in a continuous time setting. As previously mentioned, the analysis of derivatives provides a natural setting for nonlinear modeling since it deals with the pricing of assets with highly nonlinear payoff functions. For example, under the widely used Black–Scholes option pricing model (see [46], for a thorough description), stock prices are lognormally distributed and follow a Wiener process. The Black–Scholes model allows for highly nonlinear relationships between the pricing variables and parameters, as shown in Fig. 5.

Another popular use of continuous time processes is in modeling the autonomous dynamics of processes such as interest rates and the prices of stocks and commodities. A generic stochastic differential equation that can be used to nest alternative models is the following:

$$dS_t = \mu(S_t, t) \, dt + \sigma(S_t, t) \, dW_t + y(S_t, t) \, dq_t \quad (13)$$

where S_t is the price at time t, dW_t is a standard Wiener process, $\mu(S_t, t)$ is the drift, and $\sigma(S_t, t)$ is the diffusion coefficient. Both the drift and diffusion coefficients are assumed to be functions of time and price, respectively. A jump component is also allowed by incorporating a Poisson process, dq_t, with a constant arrival parameter λ, i. e., $\Pr\{dq_t = 1\} = \lambda dt$ and $\Pr\{dq_t = 0\} = 1 - \lambda dt$: y is the jump amplitude, also a function of time and price. dW_t, dq_t and y are assumed to be mutually independent processes. Several nonlinear models can be obtained by combining various assumptions for the components $\mu(S_t, t)$, $\sigma(S_t, t)$ and $y(S_t, t)$. For example, consider the following processes.

Mean Reverting Square-Root Process (MRSRP)

$$dS_t = \kappa(\theta - S_t) \, dt + \sigma \sqrt{S_t} dW_t \quad (14)$$

Constant Elasticity of Variance (CEV)

$$dS_t = \kappa(\theta - S_t) \, dt + \sigma S_t^\gamma \, dW_t \quad (15)$$

Geometric Wiener Process augmented by Jumps (GWPJ)

$$dS_t = (\mu - \lambda \mu_j) \, S_t dt + \sigma S_t dW_t + (e^y - 1) \, S_t dq_t \quad (16)$$

MRSRP augmented by Jumps (MRSRPJ)

$$dS_t = \kappa(\theta - S_t) \, dt + \sigma \sqrt{S_t} dW_t + y dq_t \, . \quad (17)$$

Model (14) has been widely used in modeling interest rates (e. g., [1,23,25]) and stochastic volatility (e. g., [44,68]). Process (16) is often used for representing the dynamics of stock prices and indices (e. g., [61]). Model (17) has been recently employed by several researchers for modeling volatility, because it allows rapid changes in volatility during times of market stress (e. g., [31]). While process (16) has a proportional structure, with μ being the expected return of the asset per unit of time and σ its
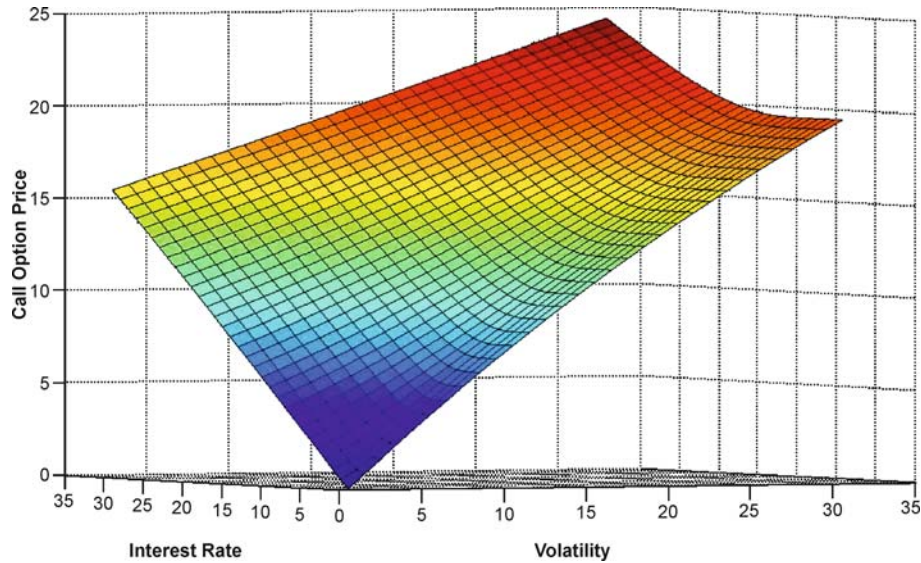
Financial Economics, Non-linear Time Series in, Figure 5
Call option prices, volatility and interest rate in the Black–Scholes model (Call option prices were estimated using the Black–Scholes model assuming a strike price of 50, 1 year time to maturity and a zero dividend yield)

volatility, the other processes have mean reverting drifts. In Eqs. (14), (15) and (17) κ is the speed of mean reversion, θ is the unconditional long-run mean, and σ the volatility of the price process. In Eq. (15), γ is a free parameter to be estimated that determines the dependence of the diffusion component on the current level of S. In Eqs. (16) and (17), λ is the average number of jumps per year and y is the jump size, which can be drawn from a normal or a double exponential distribution (see [49]).

An alternative way of representing the conditional variance is to use a stochastic volatility model, in which volatility is driven by its own noise (see the entry by Andersen and Benzoni on ▶ Stochastic Volatility). Stochastic volatility models are advantageous in that they are very flexible and have representations in both discrete and continuous time. The square root volatility model (also known as a scalar affine diffusion), proposed by Heston [44], is one of the most popular models in this area and is represented by the stochastic processes

$$
\begin{aligned}
\mathrm{d}\log(p_t) &= (\mu - 0.5\sigma_t)\,\mathrm{d}t + \sqrt{V_t}\mathrm{d}W_{1t} \\
\mathrm{d}V_t &= (\alpha - \beta\sigma_t)\,\mathrm{d}t + \sigma_V\sqrt{V_t}\mathrm{d}W_{2t}
\end{aligned}
\tag{18}
$$

where V_t is the instantaneous (latent) stochastic volatility, which is assumed to follow a mean reverting square root process. The parameter k measures the speed of mean reversion, while θ is the unconditional long run mean. $\mathrm{d}W_{1t}$ and $\mathrm{d}W_{2t}$ are Brownian motions with instantaneous correlation $\rho\mathrm{d}t$.

Future Directions

The coverage in this essay has, unavoidably, been far from exhaustive. The realm of relevant nonlinear models and theories in finance is extremely rich and is developing fast (a useful review of new developments is [66]). By transcending the representative agent framework and by extending the standard notion of rationality, researchers are now allowing for interactions between heterogeneous groups of investors using agent based models (for an overview of these fascinating developments, see [51] and the entry on ▶ Finance, Agent Based Modeling in by Manzan). While such approaches can reproduce stylized facts such as volatility clustering and long-term dependencies, it remains to be seen how they can be standardized and applied to the solution of specific problems by academics and practitioners.

Bibliography

1. Aït-Sahalia Y (1999) Transition densities for interest rate and other nonlinear diffusions. J Finance 54:1361–1395
2. Akdeniz L, Altay-Salih A, Caner M (2003) Time varying betas help in asset pricing: The threshold CAPM. Stud Nonlinear Dyn Econom 6:1–16
3. Baillie RT (1996) Long memory processes and fractional integration in econometrics. J Econom 73:5–59
4. Baillie RT, Bollerslev T, Mikkelson HO (1996) Fractionally integrated generalized autoregressive conditional heteroskedasticity. J Econom 74:3–30

5. Bakshi G, Ju N, Yang H (2006) Estimation of continuous-time models with an application to equity volatility dynamics. J Financ Econom 82:227–249

6. Balke NS, Fomby TB (1997) Threshold cointegration. Int Econom Rev 38:627–645

7. Barkley Rosser J Jr (1999) On the complexities of complex economic dynamics. J Econom Perspect 13:169–192

8. Barndorff-Nielsen OE, Graversen SE, Shephard N (2004) Power variation and stochastic volatility: A review and some new results. J Appl Probab 41:133–143

9. Bauwens L, Laurent S, Rombouts JVK (2006) Multivariate GARCH models: A survey. J Appl Econom 21:79–109

10. Bera AK, Higgins ML (1993) On ARCH models: Properties, estimation and testing. J Econom Surv 7:305–366

11. Beran JA (1992) Statistical methods for data with long-range dependence. Stat Sci 7:404–427

12. Biais B, Glosten L, Spatt C (2005) Market microstructure: A survey of microfoundations, empirical results, and policy implications. J Financ Mark 8:217–264

13. Bollerslev T (1986) Generalised autoregressive conditional heteroskedasticity. J Econom 31:307–27

14. Bollerslev T, Chou RY, Kroner KF (1992) ARCH modelling in finance: A review of the theory and empirical evidence. J Econom 52:5–59

15. Bollerslev T, Engle RF, Nelson DB (1994) ARCH Models. In: Engle RF, McFadden DL (eds) Handbook of Econometrics, vol 4. New York, North-Holland, pp 2959–3038

16. Box GEP, Jenkins GM (1976) Time Series Analysis: Forecasting and Control. Rev. Edn., Holden Day, San Francisco

17. Brandt MW, Diebold FX (2006) A no-arbitrage approach to range-based estimation of return covariances and correlations. J Bus 79:61–74

18. Brock WA (1999) Scaling in economics: A reader's guide. Ind Corp Change 8:409–446

19. Brock WA, de Lima PJF (1996) Nonlinear time series, complexity theory, and finance. In: Maddala GS, Rao RS (eds) Handbook of Statistics, vol 14. Elsevier, Amsterdam, pp 317–361

20. Brock WA, Dechert WD, Scheinkman JA, LeBaron B (1996) A test for independence based on the correlation dimension. Econom Rev 15:197–235

21. Brooks C (2006) Multivariate Volatility Models. In: Mills TC, Patterson K (eds) Palgrave Handbook of Econometrics, vol 1. Econometric Theory. Palgrave Macmillan, Basingstoke, pp 765–783

22. Campbell JY, Lo AW, MacKinlay AC (1997) The Econometrics of Financial Markets. Princeton University Press, New Jersey

23. Chan KC, Karolyi A, Longstaff FA, Sanders AB (1992) An empirical comparison of alternative models of the short-term interest rate. J Finance 47:1209–1227

24. Cochrane JH (2005) Asset Pricing. Princeton University Press, Princeton

25. Cox JC, Ingersoll JE, Ross SA (1985) A theory of the term structure of interest rates. Econometrica 53:385–408

26. Dixit AK, Pindyck RS (1994) Investment under Uncertainty. Princeton University Press, Princeton

27. Engle RF (1982) Autoregressive conditional heteroskedasticity with estimates of the variance of UK inflation. Econometrica 50:987–1008

28. Engle RF (2000) The econometrics of ultra-high frequency data. Econometrica 68:1–22

29. Engle RF, Granger CWJ (1987) Cointegration and error correction: representation, estimation and testing. Econometrica 55:251–276

30. Engle RF, Lilien DM, Robins RP (1987) Estimating time varying risk premia in the term structure: the ARCH-M model. Econometrica 55:391–408

31. Eraker B, Johannes M, Polson N (2003) The impact of jumps in volatility and returns. J Finance 53:1269–1300

32. Fama EF (1991) Efficient capital markets, vol II. J Finance 26:1575–1617

33. Fama EF (1998) Market efficiency, long-term returns, and behavioural finance. J Financ Econom 49:283–306

34. Frankel FA, Froot KA (1987) Using survey data to test propositions regarding exchange rate expectations. Am Econom Rev 77:33–153

35. Frankel FA, Froot KA (1988) Chartists, fundamentalists and the demand for dollars. Greek Econom Rev 10:49–102

36. Froot KA, Obstfeld M (1991) Exchange-rate dynamics under stochastic regime shifts – A unified approach. J Int Econom 31:203–229

37. Goldberg MD, Frydman R (1996) Imperfect knowledge and behaviour in the foreign exchange market. Econom J 106: 869–893

38. Goodhart CAE, O'Hara M (1997) High frequency data in financial markets: Issues and applications. J Empir Finance 4:73–114

39. Granger CWJ, Joyeux R (1980) An introduction to long memory time series models and fractional differencing. J Time Ser Anal 1:15–29

40. Granger CWJ, Teräsvirta T (1993) Modelling Nonlinear Economic Relationships. Oxford University Press, New York

41. Hamilton JD (1989) A new approach to the economic analysis of nonstationary time series and the business cycle. Econometrica 57:357–384

42. Hamilton JD (1990) Analysis of time series subject to changes in regime. J Econom 45:39–70

43. Hansen BE (1994) Autoregressive conditional density estimation. Int Econom Rev 35:705–730

44. Heston SL (1993) A closed-form solution for options with stochastic volatility with applications to bond and currency options. Rev Financ Stud 6:327–343

45. Hosking JRM (1981) Fractional differencing. Biometrika 68:165–176

46. Hull JC (2005) Options, Futures and Other Derivatives, 6th edn. Prentice Hall, Upper Saddle River

47. Jondeau E, Rockinger M (2003) Conditional volatility, skewness, and kurtosis: existence, persistence, and comovements. J Econom Dyn Control 27:1699–1737

48. Kon S (1984) Models of stock returns – a comparison. J Finance 39:147–65

49. Kou SG (2002) A jump-diffusion model for option pricing. Management Sci 48:1086–1101

50. Krugman PR (1991) Target zones and exchange-rate dynamics. Q J Econom 106:669–682

51. LeBaron B (2006) Agent-based Computational Finance. In: Tesfatsion L, Judd K (eds) Handbook of Computational Economics. North-Holland, Amsterdam, pp 1187–1232

52. Li WK, Ling S, McAleer M (2002) Recent theoretical results for time series models with GARCH errors. J Econom Surv 16: 245–269

53. Lorenz HW (1989) Nonlinear Dynamical Economics and Chaotic Motion. Springer, New York

F

54. Mandelbrot BB (1963) New methods in statistical economics. J Political Econom 71:421–440

55. Mandelbrot BB (1963) The variation of certain speculative prices. J Bus 36:394–419

56. Mandelbrot BB (1969) Long-run linearity, locally Gaussian process, H-spectra, and infinite variances. Int Econom Rev 10:82–111

57. Mandelbrot BB (1972) Statistical methodology for nonperiodic cycles: From the covariance to R/S analysis. Ann Econom Soc Measurement 1/3:259–290

58. Mandelbrot BB (1989) Louis Bachelier. In: The New Palgrave: Finance. Macmillan, London, pp 86–88

59. Mandelbrot BB (1997) Three fractal models in finance: Discontinuity, concentration, risk. Econom Notes 26:171–211

60. Markellos RN, Mills TC (2003) Asset pricing dynamics. Eur J Finance 9:533–556

61. Merton RC (1976) Option prices when underlying stock returns are discontinuous. J Financ Econom 3:125–144

62. Merton RC (1998) Applications of option-pricing theory: Twenty-five years later. Am Econom Rev 88:323–347

63. Mills TC, Markellos RN (2008) The Econometric Modelling of Financial Time Series, 3rd edn. Cambridge University Press, Cambridge

64. Mittnik S, Rachev ST (1993) Modeling asset returns with alternative stable distributions. Econom Rev 12:261–330

65. Mittnik S, Rachev ST (1993) Reply to comments on "Modeling asset returns with alternative stable distributions" and some extensions. Econom Rev 12:347–389

66. Mizrach B (2008) Nonlinear Time Series Analysis. In: Blume L, Durlauf S (eds) The New Palgrave Dictionary of Economics, 2nd edn. Macmillan, London, pp 4611–4616

67. Mullineux A, Peng W (1993) Nonlinear business cycle modelling. J Econom Surv 7:41–83

68. Pan J (2002) The jump-risk premia implicit in options: Evidence from an integrated time-series study. J Financ Econom 63:3–50

69. Poitras G (2000) The Early History of Financial Economics. Edward Elgar, Cheltenham, pp 1478–1776

70. Potter S (1999) Nonlinear time series modelling: An introduction. J Econom Surv 13:505–528

71. Rachev ST, Menn C, Fabozzi FJ (2005) Fat Tailed and Skewed Asset Distributions. Wiley, New York

72. Racine JS, Ullah A (2006) Nonparametric Econometrics. In: Mills TC, Patterson K (eds) Palgrave Handbook of Econometrics, vol 1: Econometric Theory. Palgrave Macmillan, Basingstoke, pp 1001–1034

73. Robinson PM (2003) Long Memory Time Series. In: Robinson PM (ed) Time Series with Long Memory. Oxford University Press, London, pp 4–32

74. Rosser JB Jr (1991) From Catastrophe to Chaos: A General Theory of Economic Discontinuities. Kluwer Academic, Norwell

75. Rostow WW (1993) Nonlinear Dynamics: Implications for economics in historical perspective. In: Day RH, Chen P (eds) Nonlinear Dynamics and Evolutionary Economics. Oxford University Press, Oxford

76. Samorodnitsky G, Taqqu MS (1994) Stable Non-Gaussian Random Processes. Chapman and Hall, New York

77. Schumpeter JA (1939) Business Cycles. McGraw-Hill, New York

78. Schwartz E (1997) The stochastic behavior of commodity prices: Implications for valuation and hedging. J Finance 52:923–973

79. Teräsvirta T (2006) An Introduction to Univariate GARCH Models. In: Andersen TG, Davis RA, Kreiss JP, Mikosch T (eds) Handbook of Financial Time Series. Springer, New York

80. Teräsvirta T, Medeiros MC, Rech G (2006) Building neural network models for time series: A statistical approach. J Forecast 25:49–75

81. Tong H (1990) Nonlinear Time Series: A Dynamical Systems Approach. Oxford University Press, Oxford

82. Velasco C (2006) Semiparametric Estimation of Long-Memory Models. In: Mills TC, Patterson K (eds) Palgrave Handbook of Econometrics, vol 1: Econometric Theory. Palgrave MacMillan, Basingstoke, pp 353–95

Financial Economics, Return Predictability and Market Efficiency

STIJN VAN NIEUWERBURGH[1], RALPH S. J. KOIJEN[2]
[1] Department of Finance, Stern School of Business, New York University, New York, USA
[2] Department of Finance, Tilburg University, Tilburg, The Netherlands

Article Outline

Glossary

Stock return The stock return in this entry refers to the return on the portfolio of all stocks that are traded on the three largest equity markets in the US: the NYSE, NASDAQ, and AMEX. The return is measured as the price of the stock at the end of the year plus the dividends received during the year divided by the price at the beginning of the year. The return of each stock is weighted by its market capitalization when forming the portfolio. The source for the data is CRSP.

Dividend-price ratio and dividend yield The dividend-price ratio of a stock is the ratio of the dividends received during the year divided by the price of the stock at the *end* of the year. The dividend yield, instead, is the ratio of the dividends received during the year divided by the price of the stock at the *beginning* of the year. The stock return is the sum of the dividend yield and the capital gain yield, which measures the ratio of the end-of-year stock price to the beginning-of-year stock price.

Predictability A stock return r_{t+1} is said to be predictable by some variable x_t if the expected return conditional on x_t, $E[r_{t+1} \mid x_t]$, is different from the unconditional expected return, $E[r_{t+1}]$. No predictability means that the best predictor of tomorrow's return is the constant, unconditional average return, i. e., $E[r_{t+1} \mid x_t] = E[r_{t+1}]$. When stock returns are unpredictable, stock prices are said to follow a random walk.

Market model The market model links the return on any asset i, r_{it} to the return on the market portfolio (r_t). Under joint normality of returns, it holds:

$$r_{it} = \alpha_i + \beta_i r_t + \varepsilon_{it} , \qquad (1)$$

with $E[\varepsilon_{it}] = 0$ and $\mathrm{Var}[\varepsilon_{it}] = \sigma_{\varepsilon_i}^2$, see [16]. The typical assumption in the literature until the 1980s has been that $E[r]$ is constant.

Definition of the Subject

The efficient market hypothesis, due to [21,22] and [23], states that financial markets are efficient with respect to a particular information set when prices aggregate all available information. Testing the efficient market hypothesis requires a "market model" which specifies how information is incorporated into asset prices. Efficiency of markets is then synonymous with the inability of investors to make economic, i. e., risk-adjusted, profits based on this information set [36]. The question of market efficiency and return predictability is of tremendous importance for investors and academics alike. For investors, the presence of return predictability would lead to different optimal asset allocation rules. Failing to make portfolios conditional on this information may lead to substantial welfare losses. For academics, return predictability or the lack thereof has substantial implications for general equilibrium models that are able to accurately describe the risks and returns in financial markets.

Introduction

Until the 1980s, the standard market model assumed constant expected returns. The first empirical evidence, which showed evidence that returns were predictable to some extent, was therefore interpreted as a sign of market inefficiency [25,54]. [56] proposed the alternative explanation of time-varying expected returns. This prompted the question of why aggregate stock market returns would be time varying in equilibrium. [23] provides a summary of this debate.

Recently developed general equilibrium models show that expected returns can indeed be time varying, even if

markets are efficient. Time-variation in expected returns can result from time-varying risk aversion [11], long-run consumption risk [5], or time-variation in risk-sharing opportunities, captured by variation in housing collateral [44]. Predictability of stock returns is now, by-and-large, interpreted as evidence of time-varying expected returns rather than market inefficiency.

Motivating Predictive Regressions

Define the gross return on an equity investment between period t and period $t + 1$ as

$$R_{t+1} = \frac{P_{t+1} + D_{t+1}}{P_t} ,$$

where P denotes the stock price and D denotes the dividend. [9] log-linearizes the definition of a return to obtain:

$$r_{t+1} = k + \Delta d_{t+1} + \rho dp_{t+1} - dp_t . \qquad (2)$$

All lower-case letters denote variables in logs; d_t stands for dividends, p_t stands for the price, $dp_t \equiv d_t - p_t$ is the log dividend–price ratio, and r_t stands for the return. The constants k and $\rho = (1 + \exp(\overline{dp}))^{-1}$ are related to the long-run average log dividend–price ratio $\overline{dp}$. By iterating forward on Eq. (2) and by imposing a transversality condition (i. e., we rule out rational bubbles), one obtains

$$dp_t = \overline{dp} + E_t \sum_{j=1}^{\infty} \rho^{j-1}\left[(r_{t+j} - \overline{r}) - (\Delta d_{t+j} - \overline{d})\right]. \quad (3)$$

Since this equation holds both ex-post and ex-ante, an expectation operator can be added on the right-hand side. This equation is one of the central tenets of the return predictability literature, the so-called Campbell and Shiller [12,13] equation. It says that, as long as the expected returns and expected dividend growth are stationary, deviations of the dividend–price ratio (dp_t) from its long-term mean ($\overline{dp}$) ought to forecast either future returns, or future dividend growth rates, or both.

This accounting identity has motivated some of the earliest empirical work in return predictability, which regressed returns on the lagged dividend–price ratio, as in Eq. (4):

$$(r_{t+1} - \overline{r}) = \kappa_r(dp_t - \overline{dp}) + \tau_{t+1}^r , \qquad (4)$$

$$(\Delta d_{t+1} - \overline{d}) = \kappa_d(dp_t - \overline{dp}) + \tau_{t+1}^d , \qquad (5)$$

$$(dp_{t+1} - \overline{dp}) = \phi(dp_t - \overline{dp}) + \tau_{t+1}^{dp} , \qquad (6)$$

where $\overline{r}$ is the long-run mean return and τ^r is a mean-zero innovation. The logic of (3) suggests that the dividend–

price ratio could predict future dividend growth rates instead of, or in addition to, future returns. Testing for dividend growth predictability would lead one to estimate Eq. (5), where $\overline{d}$ denotes the long-run mean log dividend growth.

The empirical return predictability literature started out by estimating Eq. (4) with the dividend–price ratio on the right-hand side; see [12,17,24,29,34,53] and [42], among others. It found evidence for return predictability, i. e., $\kappa_r > 0$. This finding was initially interpreted as evidence against the efficient market hypothesis.

Around the same time, [25] and [52] document a negative autocorrelation in long-horizon returns. Good past returns forecast bad future returns. [16] and [18] summarize the evidence based on long-horizon autocorrelations and variance ratios, and conclude that the statistical evidence in favor of mean reversion in long-horizon returns is weak, possibly due to small sample problems. This motivates [4] to use a large cross-section of countries and use a panel approach instead. They in turn document strong evidence in favor of mean-reversion of long-horizon returns with an estimated half-life of 3–3.5 years.

Second, other financial ratios, such as the earnings-price ratio or the book-to-market ratio, or macro-economic variables such as the consumption-wealth ratio, the labor income-to-consumption ratio, or the housing collateral ratio, as well as corporate decisions, and the cross-sectional price of risk have subsequently been shown to predict returns as well; see [3,38,39,43,45,50] and [51], among others.

Third, long-horizon returns are typically found to be more predictable than one-period ahead returns. The coefficient $\kappa_r(H)$ in the H-period regression

$$\sum_{j=1}^{H} r_{t+j} = \kappa_r(H)\, dp_t + \tau_{t,t+H}^r \qquad (7)$$

exceeds the coefficient κ_r in the one-period regression. This finding is interpreted as evidence for the fact that the time-varying component in expected returns is quite persistent.

Fourth, these studies conclude that growth rates of fundamentals, such as dividends or earnings, are much less forecastable than returns using financial ratios. This suggests that most of the variation of financial ratios is due to variation in expected returns.

Fifth, predictability of stock returns does not only arise for the US. Studies by [10,26,33], and [2] analyze a large cross-section of countries and find evidence in favor of predictability by financial ratios in some countries, even

though the evidence is mixed. More robust results are documented for the predictive ability of term structure variables.

These conclusions regarding predictability of stock returns are controversial because the forecasting relationship of financial ratios and future stock returns exhibits three disconcerting statistical features. First, correct inference is problematic because financial ratios are extremely persistent. The empirical literature typically augments Eq. (4) with an auto-regressive specification for the predictor variable, as in Eq. (6), where $\overline{dp}$ is the long-run mean of the dividend–price ratio. The estimated autoregressive parameter ϕ is near unity and standard tests leave the possibility of a unit root open (i. e., $\phi = 1$). [2,27,46,55] and [58] conclude that the statistical evidence of forecastability is weaker once tests are adjusted for high persistence. [1,2,15,42,57] and [20] derive asymptotic distributions for predictability coefficients under the assumption that the forecasting variable follows a local-to-unit root, yet stationary, process.

Second, financial ratios have poor out-of-sample forecasting power, as shown in [7,31], and [32], but see [35] and [14] for different interpretations of the out-of-sample tests and evidence.

Third, the forecasting relationship of returns and financial ratios exhibits significant instability over time. Figure 1 shows that in rolling 30-year regressions of annual log CRSP value-weighted returns on lagged log dividend–price ratios, the ordinary least squares (OLS) regression coefficient varies between zero and 0.5 and the associated R^2 ranges from close to zero to 30% depending on the subsample.

The figure plots estimation results for the equation $r_{t+1} - \bar{r} = \kappa_r(dp_t - \overline{dp}) + \tau_{t+1}^r$. It shows the estimates for κ_r using 30-year rolling windows. The dashed line in the left panels denote the point estimate plus or minus one standard deviation. The standard errors are asymptotic. The parameters $\bar{r}$ and $\overline{dp}$ are the sample means of log returns r and the log dividend–price ratio dp. The data are annual for 1927–2004.

[60] and [49] report evidence in favor of breaks in the OLS coefficient in the forecasting regression of returns on the lagged dividend–price ratio, while [41] report evidence for structural shifts in $\overline{dp}$. [47] use Bayesian methods to estimate structural breaks in the equity premium.

Empirical Evidence Revisited

Table 1 reviews the empirical evidence using annual value-weighted CRSP log return, dividend growth, and dividend–price ratio data for 1927–2004. In Panel A, the sys-

30–Year Rolling Estimates for Return Coefficient
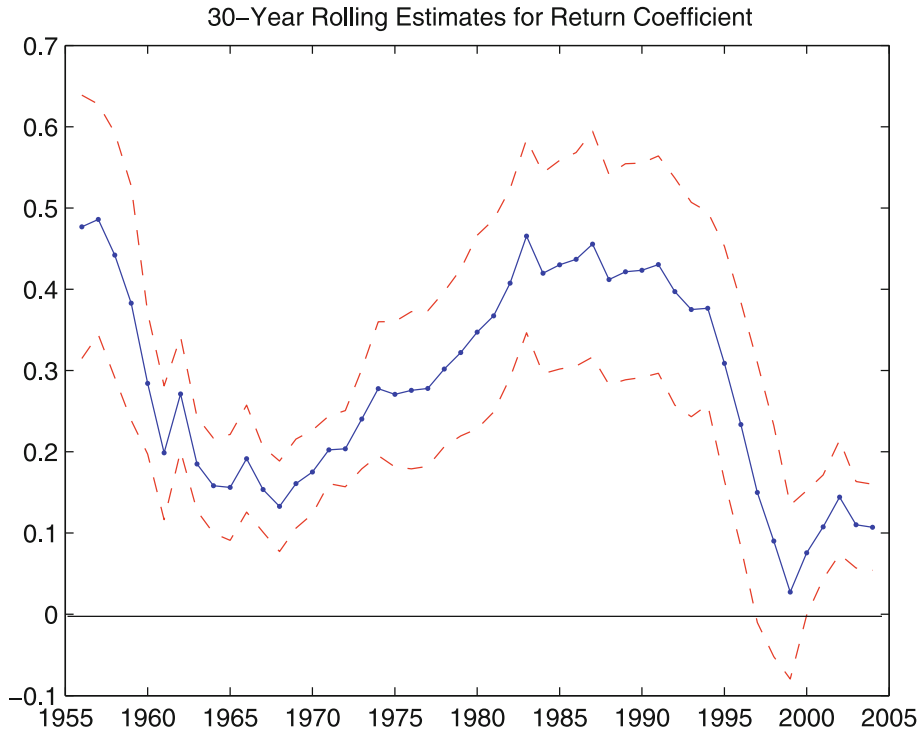
Financial Economics, Return Predictability and Market Efficiency, Figure 1
Parameter Instability in Return Predictability Coefficient

tem of Eqs. (4) and (5) is estimated by GMM. The first row indicates that a higher dividend–price ratio leads to a higher return ($\kappa_r = .094$ in Column 2) and a higher dividend growth rate ($\kappa_d = .005$ in Column 1). The latter coefficient has the wrong sign, but the coefficient is statistically indistinguishable from zero. The asymptotic standard error on the estimate for κ_r is .046. The corresponding asymptotic p-value is 3.6% so that κ_r is statistically different from zero at conventional levels. In other words, the dividend–price ratio seems to predict stock returns, but not dividend growth. A similar result holds if returns in excess of a risk-free rate are used, or real returns instead of nominal returns.

[41] conduct an extensive Monte Carlo analysis to investigate the small-sample properties of estimates for κ_r and κ_d. Consistent with [55], the estimate for κ_r displays an upward small-sample bias. In addition, the standard error on κ_r is understated by the asymptotic standard error. As a result, one can no longer reject the null hypothesis that κ_r is zero. Based on this evidence, one is tempted to conclude that neither returns nor dividend growth are forecastable.

The second and third rows implement the suggestion of [41] to correct the long-run mean dividend–price ratio,

$\overline{dp}$, for structural breaks. The data strongly suggest either one break in 1991, or two breaks in 1954 and 1994 in favor of either no breaks or three breaks. This break-adjusted dividend–price ratio is less persistent and less volatile. Its

Financial Economics, Return Predictability and Market Efficiency, Table 1
Return and Dividend Growth Predictability in the Data

	κ_d	κ_r	ϕ	PV violation
Panel A: No Long-Horizon Moments $H = \{1\}$				
No Break	.005	.094	.945	−.046
	(.037)	(.046)	(.052)	
1 Break ('91)	.019	.235	.813	.004
	(.047)	(.055)	(.052)	
2 Breaks ('54, '94)	.124	.455	.694	−.001
	(.073)	(.079)	(.070)	
Panel B: Long-Horizon Moments $H = \{1, 3, 5\}$				
No Break	.021	.068	.990	.189
	(.018)	(.038)	(.032)	
1 Break ('91)	.012	.210	.834	.076
	(.019)	(.043)	(.042)	
2 Breaks ('54, '94)	.080	.409	.697	.100
	(.065)	(.078)	(.060)	

lower persistence alleviates the econometric issues mentioned above.

The second row of Table 1 uses the one-break adjusted dividend–price ratio as a regressor in the return and dividend growth predictability equations. The evidence in favor of return predictability is substantially strengthened. The point estimate for κ_r more than doubles to .235, and is highly significant. In the two-break case in the third row, the point estimate further doubles to 0.455. The small-sample bias in κ_r is negligible relative to the size of the coefficient. The R^2 of the return equation is 10% in the one-break case and even 23% in the two-break case. This compares to 3.8% in the no-break case. Furthermore, rolling regression estimates of κ_r indicate that it is much more stable over time when the break-adjusted dp series is used as a regressor. The dividend growth coefficient κ_d remains statistically indistinguishable from zero. This evidence strengthens the view that returns are predictable and dividend growth is not, and that these findings are not an artefact of statistical issues.

This table reports GMM estimates for the parameters $(\kappa_d, \kappa_r, \phi)$ and their asymptotic standard errors (in parentheses). The results in panel A are for the system with one-year ahead equations for dividend growth and returns ($H = 1$, $N = 0$). The results in panel B are for the system with one-year, three-year and five-year ahead equations for dividend growth and returns ($H = \{1, 3, 5\}$, $N = 2$). The first-stage GMM weighting matrix is the identity matrix. The asymptotic standard errors and p-values are computed using the Newey–West HAC procedure (second stage weighting matrix) with four lags in panel A and $H = 5$ lags in panel B. The last column denotes the present-value constraint violation of the univariate OLS slope estimators: $(1 - \rho\phi^{\text{ols}})^{-1}(\kappa_r^{\text{ols}} - \kappa_d^{\text{ols}})$. It is expressed in the same units as κ_d and κ_r. In panel B this number is the average violation of the three constraints, one constraint at each horizon. The dividend–price ratio in rows 1 and 4 is the unadjusted one. In rows 2 and 5, the dividend–price ratio is adjusted for one break in 1991, and in rows 3 and 6, it is the series adjusted for two breaks in 1954 and 1994. All estimation results are for the annual sample 1927–2004.

Structural Model

What are researchers estimating when they run the return predictability regression (4)? How are the return and dividend growth predictability regressions in (4) and (5) related? To answer these important questions, we set up a simple structural model with time-varying expected returns and expected dividend growth rates. This structural model has the system of Eqs. (4)–(6) as its reduced-form. The main purpose of this model is to show that (i) the dividend–price ratio is a contaminated predictor of returns and dividend growth rates, (ii) that the parameters in (4)–(6) have to satisfy a cross-equation restriction, which we call the *present-value constraint*, and (iii) this restriction enables decomposing the dividend–price ratio into expected returns and expected dividend growth. Similar models can be derived for financial ratios other than the dividend–price ratio (e. g., [61]). [6] show how stock returns and book-to-market ratios are related in a general equilibrium model.

A Present-Value Model

We assume that expected dividend growth, z, and expected returns, x, follow an AR(1) process with autoregressive coefficient ϕ:

$$\Delta d_{t+1} - \bar{d} = z_t + \epsilon_{t+1}, \qquad z_{t+1} = \phi z_t + \zeta_{t+1}, \quad (8)$$

$$r_{t+1} - \bar{r} = x_t + \eta_{t+1}, \qquad x_{t+1} = \phi x_t + \xi_{t+1}. \quad (9)$$

The model has three fundamental shocks: an innovation in unexpected dividends ϵ_{t+1}, an innovation in expected dividends ζ_{t+1}, and an innovation in expected returns ξ_{t+1}. We assume that all three errors are serially uncorrelated and have zero cross-covariance at all leads and lags: $\text{Cov}(\epsilon_{t+1}, \zeta_{t+j}) = 0, \forall j \neq 1, \text{Cov}(\xi_{t+1}, \zeta_{t+j}) = 0, \forall j \neq 1$, and $\text{Cov}(\epsilon_{t+1}, \xi_{t+j}) = 0, \forall j$, except for a contemporaneous correlation between expected return and expected dividend growth innovations $\text{Cov}(\zeta_t, \xi_t) = \chi$, and a correlation between expected and unexpected dividend growth innovations $\text{Cov}(\zeta_t, \epsilon_t) = \lambda$. We discuss innovations to unexpected returns η below.

In steady-state, the log dividend–price ratio is a function of the long-run mean return and dividend growth rate $\overline{dp} = \log\left((\bar{r} - \bar{d})/(1 + \bar{d})\right)$. The log dividend–price ratio in (3) can then be written as:

$$dp_t - \overline{dp} = \frac{x_t - z_t}{1 - \rho\phi}. \quad (10)$$

The dividend–price ratio is the difference of two AR(1) processes with the same root ϕ, which is again an AR(1) process. I.e., we recover Eq. (6).

The return decomposition in [9] implies that the innovation to unexpected returns follows from the three fundamental shocks (i. e., combine (2) with (8)–(10)):

$$\eta_{t+1} = \frac{-\rho}{1 - \rho\phi}\xi_{t+1} + \frac{\rho}{1 - \rho\phi}\zeta_{t+1} + \epsilon_{t+1}. \quad (11)$$

Since both ρ and ϕ are positive and $\rho\phi < 1$, a positive shock to expected returns leads, ceteris paribus, to a neg-

ative contemporaneous return. Likewise, a shock to expected or unexpected dividend growth induces a positive contemporaneous return.

Contaminated Predictor

The first main insight from the structural model is that the demeaned dividend–price ratio in (10) is an imperfect forecaster of both returns and dividend growth. Returns are predicted by x_t (see Eq. (9)), but variation in the dividend–price ratio is not only due to variation in x, but also in expected dividend growth z_t. The same argument applies to dividend growth which is predicted by z_t (see Eq. (8)). This implies that the regressions in the reduced-form model in (4) and (5) suffer from an errors-in-variables problem [24,30,37].

To illustrate the bias, we can link the regression coefficients κ_r and κ_d explicitly to the underlying structural parameters:

$$\kappa_r = \frac{\mathrm{Cov}(r_{t+1}, dp_t)}{\mathrm{Var}(dp_t)} = \frac{(1 - \rho\phi)(\sigma_\xi^2 - \chi)}{\sigma_\xi^2 + \sigma_\zeta^2 - 2\chi}, \qquad (12)$$

$$\kappa_d = \frac{\mathrm{Cov}(\Delta d_{t+1}, dp_t)}{\mathrm{Var}(dp_t)} = \frac{-(1 - \rho\phi)(\sigma_\zeta^2 - \chi)}{\sigma_\xi^2 + \sigma_\zeta^2 - 2\chi}. \qquad (13)$$

If growth rates are constant, i.e., $\chi = 0$ and $\sigma_\zeta = 0$, then the dividend–price ratio is a perfect predictor of returns and $\kappa_r^\star = 1 - \rho\phi$. In all other cases, there is a bias in the return predictability coefficient:

$$\kappa_r^\star - \kappa_r = \frac{(1 - \rho\phi)(\sigma_\zeta^2 - \chi)}{\sigma_\xi^2 + \sigma_\zeta^2 - 2\chi}. \qquad (14)$$

[24] argue that κ_r is downward biased ($\kappa_r^\star - \kappa_r > 0$). In fact, the structural parameters that are implied by the reduced-form model parameters indicate an upward bias. This occurs because the correlation between expected dividend growth and expected returns is sufficiently high.

A similar argument applies to κ_d. [40] construct a variable based on the co-integrating relationship between consumption, dividends from asset wealth, and dividends from human wealth. They show that this variable has strong predictive power for dividend growth, and they show that expected returns and expected growth rates are highly positively correlated. This implies that expected growth rates and expected returns have an offsetting effect on financial ratios, which makes it hard to reliably detect time-varying growth rates using such financial ratios.

Present-Value Constraint

The second main insight from the structural model is that there is a cross-equation restriction on the three innova-

tions $\tau = (\tau^d, \tau^r, \tau^{dp})$ of the reduced-form model (4)–(6). Expressed in terms of the structural parameters, these innovations are:

$$\tau_{t+1}^d = \epsilon_{t+1} + x_t \left(\frac{-\kappa_d}{1 - \rho\phi} \right) + z_t \left(\frac{\kappa_r}{1 - \rho\phi} \right) \qquad (15)$$

$$\tau_{t+1}^r = \epsilon_{t+1} + x_t \left(\frac{-\kappa_d}{1 - \rho\phi} \right) + z_t \left(\frac{\kappa_r}{1 - \rho\phi} \right) \\ - \rho \left(\frac{\xi_{t+1} - \zeta_{t+1}}{1 - \rho\phi} \right) \qquad (16)$$

$$\tau_{t+1}^{dp} = \frac{\xi_{t+1} - \zeta_{t+1}}{1 - \rho\phi}. \qquad (17)$$

They imply the present value restriction:

$$\rho\tau_{t+1}^{dp} = \tau_{t+1}^d - \tau_{t+1}^r \iff \kappa_r - \kappa_d = 1 - \rho\phi. \qquad (18)$$

Another way to write this restriction is as a restriction on a weighted sum of κ_r and κ_d: Any two equations from the system (4)–(6) implies the third. Evidence that dividend growth is not forecastable is evidence that returns are forecastable: if $\kappa_d = 0$ in Eq. (18), then $\kappa_r > 0$ because $\rho\phi < 1$. If estimating (5) uncovers that a high dividend–price ratio forecasts a higher future dividend growth ($\kappa_d > 0$), as we showed it does, then this strengthens the evidence for return predictability. [19] makes an important and closely related point: That it is important to impose the present-value relationship when testing the null hypothesis of no return predictability. That null ($\kappa_r = 0$) is truly a joint hypothesis, because it implies a negative coefficient in the dividend growth equation ($\kappa_d < 0$). [19], too, finds strong evidence for return predictability.

Returning to Panel A of Table 1, Column 3 backs out the AR(1) coefficient ϕ from the estimated κ_d and κ_r, and from the present-value constraint (18).[1] In the first row, $\phi = .945$, and is statistically undistinguishable from a unit root. This high persistence is a familiar result in the literature. The last column reports the left-hand side and the right-hand side of Eq. (18) for *univariate* OLS regressions of (4)–(6). It shows the violation of the present-value constraint. In the first row, the violation is half as large as the actual point estimate κ_r. The standard OLS point estimates do not satisfy the present-value constraint, which can lead to faulty inference.

However, when we use the break-adjusted dividend–price ratio series in rows 2 and 3, we find that (1) the persistence of the break-adjusted dp ratio is much lower

---

[1]The linearization parameter ρ is tied to the average dividend–price ratio, and is held fixed at 0.9635.

than the unadjusted series (.81 and .69 versus .95), and (2) the present-value constraint is satisfied by the OLS coefficients.

A similar present-value constraint can be derived for long-horizon return and dividend growth regressions:

$$\kappa_r(H) = \kappa_r \left(\frac{1 - \phi^H}{1 - \phi} \right)$$

$$\kappa_d(H) = \kappa_d \left(\frac{1 - \phi^H}{1 - \phi} \right) \ .$$

Not only are the coefficients on the long-horizon return predictability regressions for all horizons linked to each other (see [8]), all long-horizon regression coefficients in the return equations are also linked to those from the dividend growth equations. I. e., there is one present-value constraint for each horizon H. Imposing these restrictions in a joint estimation procedure improves efficiency.

Panel B of Table 1 shows the results from a joint estimation of 1-year, 3-year, and 5-year cumulative returns and dividend growth rates on the lagged dividend–price ratio. Because of the restrictions, there are only two parameters to be estimated from these six equations. The results are close to those from the one-year system in Panel A, confirming the main message of [8]. The main conclusion remains that returns are strongly predictable, and dividend growth rates are not.

Exploiting Correlation in Innovations

The present-value model implies a restriction on the innovations in returns and the dividend–price ratio (see Eq. (18)). A third main insight from the structural model is that this correlation contains useful information for estimating the structural parameters, and hence for how much return predictability and dividend growth predictability there truly is. [48] show that exploiting the correlation between expected and unexpected stock returns can lead to substantially more accurate estimates. The information in correlations is incorporated by specifying a prior belief about the correlation between expected and unexpected returns, and updating that prior in a Bayesian fashion using observed data. Their method ignores the present-value constraint. The structural parameters in Panel B of Table 1, which impose the present-value constraint, imply that two-thirds of the variability in the price-dividend ratio is due to expected future returns and one-third is due to expected future dividend growth rates.

Likewise, [59] write down a model like (8)–(9) where expected returns and growth rates of dividends are autoregressive, exploiting the present-value constraint. Because the price-dividend ratio is linear in expected returns x and expected dividend growth z (see Eq. (10)), its innovations in (17) can be attributed to either innovations in expected returns or expected growth rates. The present-value constraint enables one to disentangle the information in price-dividend ratios about both expected returns and growth rates, and therefore to undo the contamination coming from correlated innovations. With this decomposition in hand, it is then possible to recover the full time-series of expected returns, x, and expected growth rates, z. [59] show that the resulting processes are strong predictors of realized returns and realized dividend growth rates, respectively. This underscores the importance of specifying a present-value model to address return predictability.

Geometric or Arithmetic Returns

As a final comment, most predictive regressions are estimated using geometric, i. e. log returns, instead of arithmetic, i. e. simple returns. This choice is predominantly motivated by the [12] log-linearization discussed before. Since investors are ultimately interested in arithmetic instead of log returns, [59] specify a process for expected simple returns instead. This is made possible by applying the techniques of linearity-inducing models, recently introduced by [28].

Future Directions

The efficient market hypothesis, which states that markets efficiently aggregate all information, was first interpreted to mean that returns are not predictable. Early evidence of predictability of stock returns by the lagged dividend–price ratio seemed to be evidence against the efficient market hypothesis. However, return predictability and efficient markets are not incompatible because return predictability arises naturally in a world with time-varying expected returns. In the last 15 years, the empirical literature has raised a set of statistical objections to return predictability findings. Meanwhile, the theoretical literature has progressed, seemingly independently, in its pursuit of new ways to build models with time-varying expected returns. Only very recently has it become clear that theory is necessary to understand the empirical facts.

In this entry, we have set up a simple present-value model with time-varying expected returns that generates the regression that is the focus of the empirical literature. The model also features time-varying expected dividend growth. It shows that the dividend–price ratio contains information about both expected returns and expected dividend growth. A regression of returns on the dividend–price ratio may therefore be a poor indicator of the true ex-

tent of return predictability. At the same time, the present-value model provides a solution to this problem: It disentangles the two pieces of information in the price-dividend ratio. This allows us to interpret the standard predictability regressions in a meaningful way. Combining data with the present-value model, we conclude that there is strong evidence for return predictability. We interpret this as evidence for the presence of time-varying expected returns, not evidence against the efficient market hypothesis. The main challenge for the future is to better understand the underlying reasons for this time-variation.

Bibliography

Primary Literature

1. Amihud Y, Hurvich CM (2004) Predictive regressions: A reduced-bias estimation method. Financial Quant Anal 39:813–841
2. Ang A, Bekaert G (2007) Stock return predictability: Is it there? Rev Financial Stud 20(3):651–707
3. Baker M, Wurgler J (2000) The equity share in new issues and aggregate stock returns. J Finance 55:2219–2258
4. Balvers R, Wu Y, Gilliland E (2000) Mean reversion across national stock markets and parametric contrarian investment strategies. J Finance 55:745–772
5. Bansal R, Yaron A (2004) Risks for the long-run: A potential resolution of asset pricing puzzles. J Finance 59(4):1481–1509
6. Berk JB, Green RC, Naik V (1999) Optimal investment, growth options and security returns. J Finance 54:1153–1607
7. Bossaerts P, Hillion P (1999) Implementing statistical criteria to select return forecasting models: What do we learn? Rev Financial Stud 12:405–428
8. Boudoukh J, Richardson M, Whitelaw RF (2007) The myth of long-horizon predictability. Rev Financial Stud (forthcoming)
9. Campbell JY (1991) A variance decomposition for stock returns. Econ J 101:157–179
10. Campbell JY (2003) Consumption-based asset pricing. In: Constantinides G, Harris M, Stulz R (eds) Handbook of the Economics of Finance. North-Holland, Amsterdam (forthcoming)
11. Campbell JY, Cochrane JH (1999) By force of habit: A consumption-based explanation of aggregate stock market behavior. J Political Econ 107:205–251
12. Campbell JY, Shiller RJ (1988) The dividend–price ratio and expectations of future dividends and discount factors. Rev Financial Stud 1:195–227
13. Campbell JY, Shiller RJ (1991) Yield spreads and interest rates: A bird's eye view. Rev Econ Stud 58:495–514
14. Campbell JY, Thompson S (2007) Predicting excess stock returns out of sample: Can anything beat the historical average? Rev Financial Stud (forthcoming)
15. Campbell JY, Yogo M (2002) Efficient tests of stock return predictability. Harvard University (unpublished paper)
16. Campbell JY, Lo AW, MacKinlay C (1997) The Econometrics of Financial Markets. Princeton University Press, Princeton
17. Cochrane JH (1991) Explaining the variance of price-dividend ratios. Rev Financial Stud 5(2):243–280
18. Cochrane JH (2001) Asset Pricing. Princeton University Press, Princeton
19. Cochrane JH (2006) The dog that did not bark: A defense of return predictability. University of Chicago Graduate School of Business (unpublished paper)
20. Eliasz P (2005) Optimal median unbiased estimation of coefficients on highly persistent regressors. Department of Economics, Princeton University (unpublished paper)
21. Fama EF (1965) The behavior of stock market prices. J Bus 38:34–101
22. Fama EF (1970) Efficient capital markets: A review of theory and empirical work. J Finance 25:383–417
23. Fama EF (1991) Efficient markets: II. J Finance 46(5):1575–1618
24. Fama EF, French KR (1988) Dividend yields and expected stock returns. J Financial Econ 22:3–27
25. Fama EF, French KR (1988) Permanent and temporary components of stock prices. J Political Econ 96(2):246–273
26. Ferson WE, Harvey CR (1993) The risk and predictability of international equity returns. Rev Financial Stud 6:527–566
27. Ferson WE, Sarkissian S, Simin TT (2003) Spurious regressions in financial economics? J Finance 58(4):1393–1413
28. Gabaix X (2007) Linearity-generating processes: A modelling tool yielding closed forms for asset prices. MIT (working paper)
29. Goetzman WN, Jorion P (1993) Testing the predictive power of dividend yields. J Finance 48:663–679
30. Goetzman WN, Jorion P (1995) A longer look at dividend yields. J Bus 68:483–508
31. Goyal A, Welch I (2003) Predicting the equity premium with dividend ratios. Manag Sci 49(5):639–654
32. Goyal A, Welch I (2006) A comprehensive look at the empirical performance of the equity premium prediction. Rev Financial Stud (forthcoming)
33. Hjalmarsson E (2004) On the predictability of global stock returns, Yale University (unpublished paper)
34. Hodrick R (1992) Dividend yields and expected stock returns: Alternative procedures for inference and measurement. Rev Financial Stud 5:357–386
35. Inoue A, Kilian L (2004) In-sample or out-of-sample tests of predictability: Which one should we use? Econom Rev 23:371–402
36. Jensen MC (1978) Some anomalous evidence regarding market efficiency. J Financial Econ 6:95–101
37. Kothari S, Shanken J (1992) Stock return variation and expected dividends: A time-series and cross-sectional analysis. J Financial Econ 31:177–210
38. Lamont O (1998) Earnings and expected returns. J Finance 53:1563–87
39. Lettau M, Ludvigson SC (2001) Consumption, aggregate wealth and expected stock returns. J Finance 56(3):815–849
40. Lettau M, Ludvigson SC (2005) Expected returns and expected dividend growth. J Financial Econ 76:583–626
41. Lettau M, Van Nieuwerburgh S (2006) Reconciling the return predictability evidence. Rev Financial Stud (forthcoming)
42. Lewellen JW (2004) Predicting returns with financial ratios. J Financial Econ 74(2):209–235
43. Lustig H, Van Nieuwerburgh S (2005) Housing collateral, consumption insurance and risk premia: An empirical perspective. J Finance 60(3):1167–1219
44. Lustig H, Van Nieuwerburgh S (2006) Can housing collateral explain long-run swings in asset returns? University of California at Los Angeles and New York University (unpublished manuscript)
45. Menzly L, Santos T, Veronesi P (2004) Understanding predictability. J Political Econ 112(1):1–47

46. Nelson CC, Kim MJ (1993) Predictable stock returns: The role of small sample bias. J Finance 43:641–661
47. Pastor L, Stambaugh RF (2001) The equity premium and structural breaks. J Finance 56(4):1207–1239
48. Pastor L, Stambaugh RF (2006) Predictive systems: Living with imperfect predictors, graduate School of Business. University of Chicago Journal of Finance (forthcoming)
49. Paye BS, Timmermann A (2006) Instability of return prediction models. J Empir Finance 13(3):274–315
50. Piazzesi M, Schneider M, Tuzel S (2007) Housing, consumption, and asset pricing. J Financial Econ 83(March):531–569
51. Polk C, Thompson S, Vuolteenaho T (2006) Cross-sectional forecasts of the equity risk premium. J Financial Econ 81: 101–141
52. Poterba JM, Summers LH (1988) Mean reversion in stock returns: Evidence and implications. J Financial Econ 22:27–60
53. Rozeff MS (1984) Dividend yields are equity risk premia. J Portfolio Manag 49:141–160
54. Shiller RJ (1984) Stock prices and social dynamics. Brook Pap Econ Act 2:457–498
55. Stambaugh RF (1999) Predictive regressions. J Financial Econ 54:375–421
56. Summers LH (1986) Does the stock market rationally reflect fundamental values? J Finance 41:591–601
57. Torous W, Volkanov R, Yan S (2004) On predicting returns with nearly integrated explanatory variables. J Bus 77:937–966
58. Valkanov R (2003) Long-horizon regressions: Theoretical results and applications. J Financial Econ 68:201–232
59. van Binsbergen J, Koijen RS (2007) Predictive regressions: A present-value approach. Duke University (working paper)
60. Viceira L (1996) Testing for structural change in the predictability of asset returns. Harvard University (unpublished manuscript)
61. Vuolteenaho T (2000) Understanding the aggregate book-market ratio and its implications to current equity-premium expectations. Harvard University (unpublished paper)

Books and Reviews

Campbell JY, Lo AW, MacKinlay C (1997) The Econometrics of Financial Markets. Princeton University Press, Princeton
Cochrane JH (2005) Asset Pricing. Princeton University Press, Princeton, NJ
Malkiel BG (2004) A Random Walk Down Wall Street. W.W. Norton, New York

Financial Economics, Time Variation in the Market Return

Mark J. Kamstra[1], Lisa A. Kramer[2]
[1] Schulich School of Business, York University, Toronto, Canada
[2] Rotman School of Management, University of Toronto, Toronto, Canada

Article Outline

Glossary

AR(k) An autoregressive process of order k; a time series model allowing for first order dependence; for instance, an AR(1) model is written as $y_t = \alpha + \rho_1 y_{t-1} + \epsilon_t$ where α and ρ are parameters, ρ is typically assumed to be less than 1 in absolute value, and ϵ_t is an innovation term, often assumed to be Gaussian, independent, and identically distributed over t.

ARCH(q) A special case of the GARCH(p, q) model (see below) where $p = 0$.

Basis point A hundredth of one percent.

Bootstrap A computer intensive resampling procedure, where random draws with replacement from an original sample are used, for instance to perform inference.

Discount rate The rate of return used to discount future cashflows, typically calculated as a risk-free rate (e. g. the 90-day US T-bill rate) plus an equity risk premium.

Equity premium puzzle The empirical observation that the ex post equity premium (see entry below) is higher than is indicated by financial theory.

Ex ante equity premium The extra return investors *expect* they will receive for holding risky assets, over and above the return they would receive for holding a risk-free asset like a Treasury bill. "Ex ante" refers to the fact that the expectation is formed in advance.

Ex post equity premium The extra return investors received *after* having held a risky asset for some period of time. The ex post equity premium often differs from the ex ante equity premium due to random events that impact a risky asset's return.

Free cash flows Cash flows that could be withdrawn from a firm without lowering the firm's current rate of growth. Free cash flows are substantially different from accounting earnings and even accounting measures of the cash flow of a firm.

Fundamental valuation The practice of determining a stock's intrinsic value by discounting cash flows to their present value using the required rate of return.

GARCH(p, q) Generalized autoregressive conditional heteroskedasticity of order (p, q), where p is the order of the lagged variance terms and q is the order of the lagged squared error terms; a time series model

allowing for dependence in the conditional variance of a random variable, *y*. A GARCH(1,1) model is specified as:

$$y_t = \alpha + \epsilon_t; \quad \epsilon_t \sim \left(0, h_t^2\right)$$
$$h_t^2 = \theta + \beta h_{t-1}^2 + \gamma \epsilon_{t-1}^2 ,$$

where α, θ, β, and γ are parameters and ϵ_t is an innovation term.

Market anomalies Empirical regularities in financial market prices or returns that are difficult to reconcile with conventional theories and/or valuation methods.

Markov model A model of a probabilistic process where the random variable can only take on a finite number of different values, typically called states.

Method of moments A technique for estimating parameters (like parameters of the conditional mean and conditional variance) by matching sample moments, then solving the equations for the parameters to be estimated.

SAD Seasonal Affective Disorder, a medical condition by which reduced daylight in the fall and winter leads to seasonal depression for roughly ten percent of the world's population.

Sensation seeking A measure used by psychologists to capture an individual's degree of risk tolerance. High sensation-seeking tendency correlates with low risk tolerance, including tolerance for risk of a financial nature.

Simulated method of moments A modified version of the method of moments (see entry above) that is based on Monte Carlo simulation, used in situations when the computation of analytic solutions is infeasible.

Definition of the Subject

The realized return to any given asset varies over time, occasionally in a dramatic fashion. The value of an asset, its *expected* return, and its volatility, are of great interest to investors and to policy makers. An asset's expected return in excess of the return to a riskless asset (such as a short-term US Treasury bill) is termed the equity premium. The value of the equity premium is central to the valuation of risky assets, and hence a much effort has been devoted to determining the value of the equity premium, whether it varies, and if it varies, how predictable it is. Any evidence of predictable returns is either evidence of a predictably varying equity premium (say, because risk varies predictably) or a challenge to the rationality of markets and the efficient allocation of our society's scarce resources.

In this article, we start by considering the topic of valuation, with emphasis on simulation-based techniques. We consider the valuation of income-generating assets in the context of a constant equity premium, and we also explore the consequences of allowing some time-variation and predictability in the equity premium. Next we consider the equity premium puzzle, discussing a simulation-based technique which allows for precise estimation of the value of the equity premium, and which suggests some constraints on the types of models that should be used for specifying the equity premium process. Finally, we focus on evidence of seasonally varying expected returns in financial markets. We consider evidence that as a whole either presents some challenges to traditional hypotheses of efficient markets, or suggests agents' risk tolerance may vary over time.

Introduction

The pricing of a firm is conceptually straightforward. One approach to valuing a firm is to use historical dividend payments and discount rate data to forecast future payments and discount rates. Restrictions on the dividend and discount rate processes are typically imposed to produce an analytic solution to the fundamental valuation equation (an equation that involves calculating the expectation of an infinite sum of discounted dividends).

Common among many of the available valuation techniques is some form of consideration of multiple scenarios, including good and bad growth and discount rate evolutions, with valuation based on a weighted average of prices from the various scenarios. The valuation technique we focus some attention on, the Donaldson and Kamstra [14] (henceforth DK) methodology, is similar to pricing path-dependent options, as it utilizes Monte Carlo simulation techniques and numerical integration of the possible paths followed by the joint processes of dividend growth and discount rates, explicitly allowing path-dependence of the evolutions. The DK method is very similar in spirit to other approaches in the valuation literature which consider multiple scenarios. One distinguishing feature of the DK methodology we consider is the technique it employs for modeling the discount rate.

Cochrane [9] highlights three interesting approaches for modeling the discount rate: a constant discount rate, a consumption-based discount rate, and a discount rate equal to some variable reference return plus a risk premium. Virtually the entire valuation literature limits its attention to the constant discount rate case, as constant discount rates lead to closed-form solutions to many valuation formulas. DK explore all three methods for modeling the discount rate and find they lead to qualitatively similar results. However, their quantitative results indicate an

overall better fit to the price and return data when using a reference return plus a risk premium. Given DK's findings, we use a discount rate equal to some variable reference return plus a risk premium. In implementing this approach for modeling the discount rate used in valuation, it is simplest to assume a *constant* equity premium is added to the reference rate, in particular since the reference rate is permitted to vary (since it is typically proxied using a variable rate like the three-month US T-bill rate). We do not, however, restrict ourselves to the constant equity premium case.

Using the simulation-based valuation methodology of DK and the method of simulated moments, we explore the evidence for a time-varying equity premium and its implications for a long-standing puzzle in financial economics, the equity premium puzzle of Mehra and Prescott [51]. Over the past century the average annual return to investing in the US stock market has been roughly 6% higher than the return to investing in risk-free US T-bills. Making use of consumption-based asset-pricing models, Mehra and Prescott argue that consumption within the US has not been sufficiently volatile to warrant such a large premium on risky stocks relative to riskless bonds, leading them to describe this large premium as the "equity premium puzzle."

Utilizing simulations of the distribution from which ex post equity premia are drawn, conditional on various possible values for investors' ex ante equity premium and calibrated to S&P 500 dividends and US interest rates, we present statistical tests that show a true ex ante equity premium as low as 2% could easily produce ex post premia of 6%. This result is consistent with the well-known observation that ex post equity premia are observed with error, and a large range of realized equity premia are consistent with any given value of the ex ante equity premium. Examining the marginal and joint distributions of financial statistics like price-dividend ratios and return volatility that arise in the simulations versus actual realizations from the US economy, we argue that the range of ex ante equity premia most consistent with the US market data is very close to 3.5%, and the ex ante equity premium process is very unlikely to be constant over time.

A natural question to ask is why might the equity premium fluctuate over time? There are only two likely explanations: changing risk or changing risk aversion. Evidence from the asset-pricing literature, including [20,37,49], and many others shows that priced risk varies over time. We explore some evidence that risk aversion itself may vary over time, as revealed in what is often termed market anomalies. Market anomalies are variations in expected returns which appear to be incongruous with variations in discount rates or risk. The most stark anomalies have to do with deterministic asset return seasonalities, including seasonalities at the weekly frequency such as the weekend effect (below-average equity returns on Mondays), annual effects like the above-average equity returns typically witnessed in the month of January, and other effects like the lower-than-average equity returns often witnessed following daylight saving time-change weekends, and opposing cyclicality in bond versus equity returns correlated to the length of day (known as the SAD effect). We briefly review some of these outstanding puzzles, focusing our attention on the SAD effect and the daylight saving effect.

Valuation

Overview

We begin our discussion of valuation with a broad survey of the literature, including dividend-based valuation, relative valuation, and accounting-based methods. We introduce dividend-based valuation first.

Fundamental valuation techniques that utilize dividends in a *discrete* time framework include Gordon [25], Hawkins [30], Michaud and Davis [53], Farrell [22], Sorensen and Williamson [73], Rappaport [63], Barsky and DeLong [2], Hurley and Johnson [33], [34], Donaldson and Kamstra [14], and Yao [78]. Invariably these approaches are partial equilibrium solutions to the valuation exercise. Papers that use *continuous* time tools to evaluate the fundamental present value equation include Campbell and Kyle [6], Chiang, Davidson, and Okuney [8], Dong and Hirshleifer [17], and Bakshi and Chen [3]. The Dong and Hirshleifer [17] and Bakshi and Chen [3] papers conduct valuation by assuming dividends are proportional to earnings and then modeling earnings. Continuous time papers in this literature typically start with the representative agent/complete markets economic paradigm. Models are derived from primitive assumptions on markets and preferences, such as the equilibrium condition that there exist no arbitrage opportunities, dividend (cash flow) growth rates follow an Ornstein–Uhlenbeck mean-reverting process, and preferences over consumption are represented by the log utility function. Time-varying stochastic discount rates (i.e. the pricing kernel) fall out of the marginal rate of utility of consumption in these models, and the solution to the fundamental valuation problem is derived with the same tools used to price financial derivatives. A critique of dividend-discounting methods is that dividends are typically smoothed and are set low enough so that the dividend payments can be maintained through economic downturns. Authors such as Hackel and Livnat (see p. 9 in [27]) argue that these sorts of considerations

imply that historical records of dividend payments may therefore be poor indicators of future cash payments to investors.

A distinct valuation approach, popular amongst practitioners, determines the value of inactively traded firms by finding an actively traded firm that has similar risk, profitability, and investment-opportunity characteristics and then multiplying the actively traded firm's price-earnings (P/E) ratio by the inactively traded firm's earnings. This approach to valuation is often referred to as the relative value method or the constant P/E model. References to this sort of approach can be found in textbooks like [4], and journal articles such as [60,62].

There are also several valuation approaches that are based on the book value of equity, abnormal earnings, and free-cash flows. These approaches are linked to dividends and hence to formal fundamental valuation by well-established accounting relationships. They produce price estimates by valuing firm assets and income streams. The most popular of this class of techniques include the residual income and free-cash-flow methods. See [23,57,61] for further information. All of these valuation methods implicitly or explicitly take the present value of the stream of firm-issued dividends to the investor. The motivation for considering accounting relationships is that these accounting measures are not easily manipulated by firms and so should reflect more accurately the ability of firms to generate cashflows and hence allow more accurate assessments of the fundamental value of a firm than techniques based on dividends.

Fundamental Valuation Methods in Detail

Now that we have surveyed the valuation literature in general, we turn to a formal derivation of several *fundamental* valuation techniques. Investor rationality requires that the current market price P_t of a stock which will pay a per share dividend (cash payment) D_{t+1} one period from now and then sell for P_{t+1}, discounting payments received during period t (i. e., from the beginning of period t to the beginning of period $t + 1$) at rate r_t, must satisfy Eq. (1):

$$P_t = \mathcal{E}_t \left\{ \frac{P_{t+1} + D_{t+1}}{1 + r_t} \right\} . \tag{1}$$

$\mathcal{E}_t$ is the expectations operator conditional on information available up to the end of period t. Solving Eq. (1) forward under the transversality condition that the expected present value of P_{t+k} goes to zero as k goes to infinity (a "no-bubble" assumption) produces the familiar result that the market price equals the expected present value of future dividends (cash payments); i. e.,

$$P_t = \sum_{k=0}^{\infty} \mathcal{E}_t \left\{ \left(\prod_{i=0}^{k} \left[\frac{1}{1 + r_{t+i}} \right] \right) D_{t+k+1} \right\} . \tag{2}$$

Defining the growth rate of dividends from the beginning of period t to the beginning of period $t + 1$ as $g_t^{\mathrm{d}} \equiv (D_{t+1} - D_t)/D_t$ it follows that

$$P_t = D_t \mathcal{E}_t \left\{ \sum_{k=1}^{\infty} \left(\prod_{i=0}^{k} \left[\frac{1 + g_{t+i}^{\mathrm{d}}}{1 + r_{t+i}} \right] \right) \right\} . \tag{3}$$

Equation (3) is the fundamental valuation equation, which is not controversial and can be derived under the law of one price and non-satiation alone, as by Rubinstein [69] and others. Notice that the cash payments D_{t+k} include all cash disbursements from the firm, including cash dividends and share repurchases. Fundamental valuation methods based directly on Eq. (3) are typically called dividend discount models.

Perhaps the most famous valuation estimate based on Eq. (3) comes from the Gordon [25] Growth Model. If dividend growth rates and discount rates are constant, then it is straightforward to derive the Gordon fundamental price estimate from Eq. (3):

$$P_t^{\mathrm{G}} = D_t \left[\frac{1 + g^{\mathrm{d}}}{r - g^{\mathrm{d}}} \right] , \tag{4}$$

where r is the constant discount rate value and g^{d} is the (conditionally) constant growth rate of dividends. To produce the Gordon Growth Model valuation estimate, all we need are estimates of the dividend growth rate and discount rate, which can be obtained in a variety of ways, including the use of historically observed dividends and returns.

Extensions of the Gordon Growth Model exploit the fundamental valuation equation, imposing less stringent assumptions. The simple Gordon Growth Model imposes a constant growth rate on dividends (dividends are expected to grow at the same rate every period) while Hurley and Johnson [33] and [34] and Yao [78] develop Markov models (models that presume a fixed probability of, say, maintaining the dividend payment at current levels, and a probability of raising it, thus incorporating more realistic dividend growth processes). Two examples of these models found in Yao [78] are the Additive Markov Gordon model (Eq. (1) of Yao [78]) and the Geometric Markov Gordon model (Eq. (2) of Yao [78]). These models can be interpreted as considering different scenarios for dividend growth for a particular asset, estimating the appropriate

price for the asset under each scenario, and then averaging the prices using as weights the probability of given scenarios being observed.

The Additive Markov Gordon Growth Model is:

$$P_t^{\text{ADD}} = D_t/r + \left[1/r + (1/r)^2\right]\left(q^{\text{u}} - q^{\text{d}}\right)\Delta, \qquad (5)$$

where r is the average discount rate, q^{u} is the proportion of the time the dividend increases, q^{d} is the proportion of the time the dividend decreases, and $\Delta = \sum_{t=2}^{T}|D_t - D_{t-1}|/(T-1)$ is the average absolute value of the level change in the dividend payment.

The Geometric Markov Gordon Growth Model is:

$$P_t^{\text{GEO}} = D_t \left[\frac{1 + (q^{\text{u}} - q^{\text{d}})\Delta^\%}{r - (q^{\text{u}} - q^{\text{d}})\Delta^\%}\right], \qquad (6)$$

where $\Delta^\% = \sum_{t=2}^{T}|(D_t - D_{t-1})/D_{t-1}|/(T-1)$ is the average absolute value of the percentage rate of change in the dividend payment.

The method of DK is also an extension of the Gordon Growth Model, taking the discounted dividend growth model of Eq. (3) and re-writing it as

$$P_t = D_t \sum_{k=0}^{\infty} \mathcal{E}_t \left\{\prod_{i=0}^{k} y_{t+i}\right\}, \qquad (7)$$

where $y_{t+i} = (1 + g_{t+i}^{\text{d}})/(1 + r_{t+i})$ is the discounted dividend growth rate. Under the DK method, the fundamental price is calculated by forecasting the range of possible evolutions of y_{t+i} up to some distant point in the future, period $t + I$, calculating $PV = D_t \sum_{k=0}^{I}(\prod_{i=0}^{k} y_{t+i})$ for each possible evolution of y_{t+i}, and averaging these values of PV across all the possible evolutions. (The value of I is chosen to produce a very small truncation error. Values of $I = 400$ to 500 for annual data have been found by DK to suffice). In this way, the DK approach mirrors other extensions of the Gordon Growth Model. It is primarily distinguished from other approaches that extend the Gordon Growth Model in two regards. First, more sophisticated time series models, estimated with historical data, are used to generate the different outcomes (scenarios) by application of Monte Carlo simulation. Second, in contrast to typical modeling in which only dividend growth rates vary, the joint evolution of cashflow growth rates and discount rates are explicitly modeled as time-varying.

Among the attractive features of the free-cash-flow and residual income valuation methods is that they avoid the problem of forecasting dividends, by exploiting relationships between accounting data and dividends. It is the practical problem of forecasting dividends to infinity that have led many researchers to explore methods based on accounting data. See, for instance, Penman and Sougiannis [61].

Assume a flat term structure (i. e., a constant discount rate $r_t = r$ for all t) and write

$$P_t = \sum_{k=1}^{\infty} \frac{\mathcal{E}_t\{D_{t+k}\}}{(1+r)^k}. \qquad (8)$$

The clean-surplus relationship relating dividends to earnings is invoked in order to derive the residual income model:

$$B_{t+k} = B_{t+k-1} + E_{t+k} - D_{t+k}, \qquad (9)$$

where B_{t+k} is book value and E_{t+k} is earnings per share. Solving for D_{t+k} in Eq. (9) and substituting into Eq. (8) yields

$$P_t = \sum_{k=1}^{\infty} \frac{\mathcal{E}_t\{B_{t+k-1} + E_{t+k} - B_{t+k}\}}{(1+r)^k},$$

or

$$\begin{aligned} P_t &= B_t + \sum_{k=1}^{\infty} \frac{\mathcal{E}_t\{E_{t+k} - r \cdot B_{t+k-1}\}}{(1+r)^k} - \frac{\mathcal{E}_t\{B_{t+\infty}\}}{(1+r)^\infty} \\ &= B_t + \sum_{k=1}^{\infty} \frac{\mathcal{E}_t\{E_{t+k} - r \cdot B_{t+k-1}\}}{(1+r)^k}, \end{aligned} \qquad (10)$$

where $B_{t+\infty}/(1+r)^\infty$ is assumed to equal zero. $E_{t+k} - r \cdot B_{t+k-1}$ is termed abnormal earnings.

To derive the free cash flow valuation model, we relate dividends to cash flows with a financial assets relation in place of the clean surplus relation:

$$fa_{t+k} = fa_{t+k-1} + i_{t+k} + c_{t+k} - D_{t+k}, \qquad (11)$$

where fa_{t+k} is financial assets net of financial obligations, i_{t+k} is interest revenues net of interest expenses, and c_{t+k} is cash flows realized from operating activities net of investments in operating activities, all of which can be positive or negative. A net interest relation is often assumed,

$$i_{t+k} = rfa_{t+k-1}. \qquad (12)$$

See Fetham and Ohlson [23] for further discussion. Solving for D_{t+k} in Eq. (11) and substituting into Eq. (8), utilizing Eq. (12) and assuming the discounted present value of financial assets fa_{t+k} goes to zero as k increases, yields the free-cash-flow valuation equation:

$$P_t = fa_t + \sum_{k=1}^{\infty} \frac{\mathcal{E}_t\{c_{t+k}\}}{(1+r)^k}. \qquad (13)$$

More on the Fundamental Valuation Method of Donaldson and Kamstra

A number of approaches can be taken to conduct valuation using the DK model shown in Eq. (7). By imposing a very simple structure for the conditional expectation of discounted dividend growth rate (y_t in Eq. (7)), the expression can be solved analytically, for instance by assuming that the discounted dividend growth rate is a constant. As shown by DK, however, analytic solutions become complex for even simple ARMA models, and with sufficient non-linearity, the analytics can be intractable. For this reason, we present a general solution algorithm based on the DK method of Monte Carlo simulation.

This method simulates y_t into the future and performs a numerical (Monte Carlo) integration to estimate the terms $\{\prod_{k=0}^{i} y_{t+k}\}$ where $y_{t+k} = (1 + g_{t+k}^{\mathrm{d}})/(1 + r_{t+k})$ in the classic case of a dividend-paying firm. A general heuristic is as follows:

Step I: Model y_t, $t = 1, \ldots, T$, as conditionally time-varying, for instance as an AR(k)-GARCH(p, q) process, and use the estimated model to make conditional mean forecasts $\hat{y}_t$, $t = 1, \ldots, T$, and variance forecasts, conditional on data observed only before period t. Ensure that this model is consistent with theory, for instance that the mean level of y is less than one. This mean value can be calibrated to available data, such as the mean annual y value of 0.94 observed in the last 50 years of S&P 500 data. Recall that although analytic solutions are available for simple processes, the algorithm presented here is applicable to virtually arbitrarily non-linear conditional processes for the discounted cash payment rate y.

Step IIa: Simulate discounted cash payment growth rates. That is, produce ys that might be observed in period t given what is known at period $t - 1$. To do this for a given period t, simulate a population of J independent possible shocks (say draws from a normal distribution with mean zero and appropriate variance, or bootstrapped from the data) $\epsilon_{t,j}$, $j = 1, \ldots, J$, and add these shocks separately to the conditional mean forecast $\hat{y}_t$ from Step I, producing $y_{t,j} = \hat{y}_t + \epsilon_{t,j}$, $j = 1, \ldots, J$. The result is a simulated cross-section of J possible realizations of y_t standing at time $t - 1$, i.e. different paths the economy may take next period.

Step IIb: Use the estimated model from Step I to make the conditional mean forecast $\hat{y}_{t+1,j}$, conditional on only the jth realization for period t (i.e., $y_{t,j}$ and $\epsilon_{t,j}$) and the data known at period $t - 1$, to form $y_{t+1,j}$.

Step IIc: Repeat Step IIb to form $y_{t+2,j}, y_{t+3,j}, \ldots, y_{t+I,j}$ for each of the J economies, where I is the number of periods into the future at which the simulation is truncated. Form the perfect foresight present value ($P_{t,j}^*$) for each of the J possible economies:

$$P_{t,j}^* = A_t \Big(y_{t,j} + y_{t,j} y_{t+1,j} + y_{t,j} y_{t+1,j} y_{t+2,j} + \cdots + \prod_{i=0}^{I} y_{t+i,j} \Big); \quad j = 1, \ldots, J.$$

Provided I is chosen to be large enough, the truncated terms $\prod_{i=0}^{K} y_{t+i,j}$, $K = I + 1, \ldots, \infty$ will be negligible.

Step III: Calculate the DK fundamental price for each $t = 1, \ldots, T$:

$$P_t^{\mathrm{DK}} = \sum_{j=1}^{J} P_{t,j}^*/J . \tag{14}$$

These fundamental price estimates P_t^{DK} can be compared to the actual price (if market prices exist) at the beginning of period t to test for bubbles as demonstrated by DK, or if period t is the future, P_t^{DK} is the fundamental price forecast. This procedure is represented diagrammatically in Exhibit 1.

To illustrate the sort of forecasts that can be produced using this technique, we illustrate graphically the S&P 500 index over the past 100 years together with predicted values based on the Gordon Growth Model and the DK method. The free-cash-flow and residual income methods are not easily adapted to forecasting index prices like the S&P 500, and so are omitted here. The type of data depicted in the following figure is described in some detail by Kamstra [39].

Figure 1 has four panels. In the panels, we plot the level of the S&P 500 index (marked with bullets and a solid line) alongside price forecasts from each of the valuation techniques. In Panel A we plot the index together with the basic Gordon Growth Model price forecasts (marked with stars), in Panels B and C we plot the index together with the Additive and Geometric Gordon Growth Models' forecasts (with squares and triangles respectively), and in Panel D we plot the index alongside the DK method's forecasts (marked with diamonds). In each panel the price scale is logarithmic.

We see in Panels A, B, and C that the use of the any of the Gordon models for forming annual forecasts of the S&P 500 index level produces excessively smooth price forecasts. (If we had plotted return volatility, then the market returns would appear excessively volatile in comparison to to forecasted returns). Evidence of periods of inflated market prices relative to the forecasted prices, i. e.,
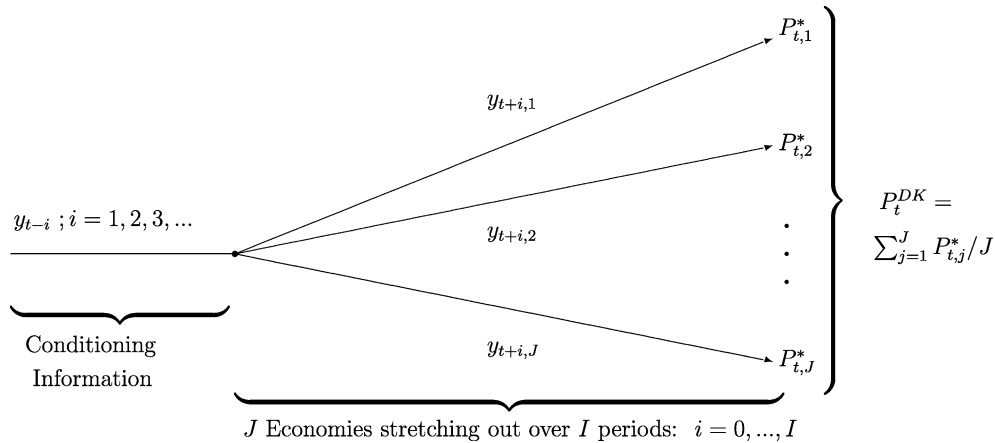
Financial Economics, Time Variation in the Market Return, Exhibit 1
Diagram of DK Monte Carlo integration

evidence of price bubbles, is apparent in the periods covering the 1920s, the 1960s, the last half of the 1980s, and the 1990s. However, if the Gordon models are too simple (since each Gordon-based model ignores the forecastable nature of discount rates and dividend growth rates), then this evidence may be misleading.

In Panel D, we see that the DK model is better able to capture the volatility of the market, including the boom of the 1920s, the 1960s and the 1980s. The relatively better performance of the DK price estimate highlights the importance of accounting for the slow fade rate of dividend growth rates and discount rates, i. e., the autocorrelation of these series. The failure of the DK method to capture the height of the 1990s boom leaves evidence of surprisingly high prices during the late 1990s. If the equity premium fell in the 1990s, as some researchers have speculated (see for instance Pástor and Stambaugh [59]), then all four sets of the plotted fundamental valuation forecasts would be expected to produce forecasts that undershoot actual prices in the 1990s, as all these methods incorporate a constant equity premium. If this premium were set too high, future cashflows would be discounted too aggressively, biasing the valuation methods downward.

The Equity Premium Puzzle

The fact that all four fundamental valuation methods we consider spectacularly fail to capture the price boom of the 1990s, possibly as a result of not allowing a time-varying equity premium, sets the stage to investigate the equity premium puzzle of Mehra and Prescott [51]. The equity premium is the extra return, or premium, that investors demand in order to be compelled to purchase risky stock

instead of risk-free debt. We call this premium the ex ante equity premium (denoted π_e), and it is formally defined as the difference between the expected return on risky assets, $\mathcal{E}\{R\}$, and the expected risk-free rate, $\mathcal{E}\{r_f\}$:

$$\pi_e \equiv \mathcal{E}\{R\} - \mathcal{E}\{r_f\} \; . \tag{15}$$

The ex post equity premium is typically estimated using historical equity returns and risk-free rates, as we do not observe the ex ante premium. Define $\overline{R}$ as the average historical annual return on the S&P 500 and $\overline{r}_f$ as the average historical return on US T-bills. A standard approach to calculate ex post equity premium, $\hat{\pi}_e$, is:

$$\hat{\pi}_e \equiv \overline{R} - \overline{r}_f \; . \tag{16}$$

Of course it is unlikely that the stock return we estimate ex post equals investors' anticipated ex ante return. Thus a 6% ex post equity premium in the US data may not be a challenge to economic theory. The question we ask is therefore: if investors' true ex ante premium is X%, what is the probability that the US economy could randomly produce an ex post premium of at least 6%? We can then argue whether or not the 6% ex post premium observed in the US data is consistent with various ex ante premium values, X%, with which standard economic theory may be more compatible. We can also consider key financial statistics and yields from the US economy to investigate if an X% ex ante equity premium could likely be consistent with the combinations that have been observed, such as high Sharpe ratio and low dividend yields, low interest rates and high ex post equity premia, and so on.

Authors have investigated the extent to which ex ante considerations may impact the realized equity premium.
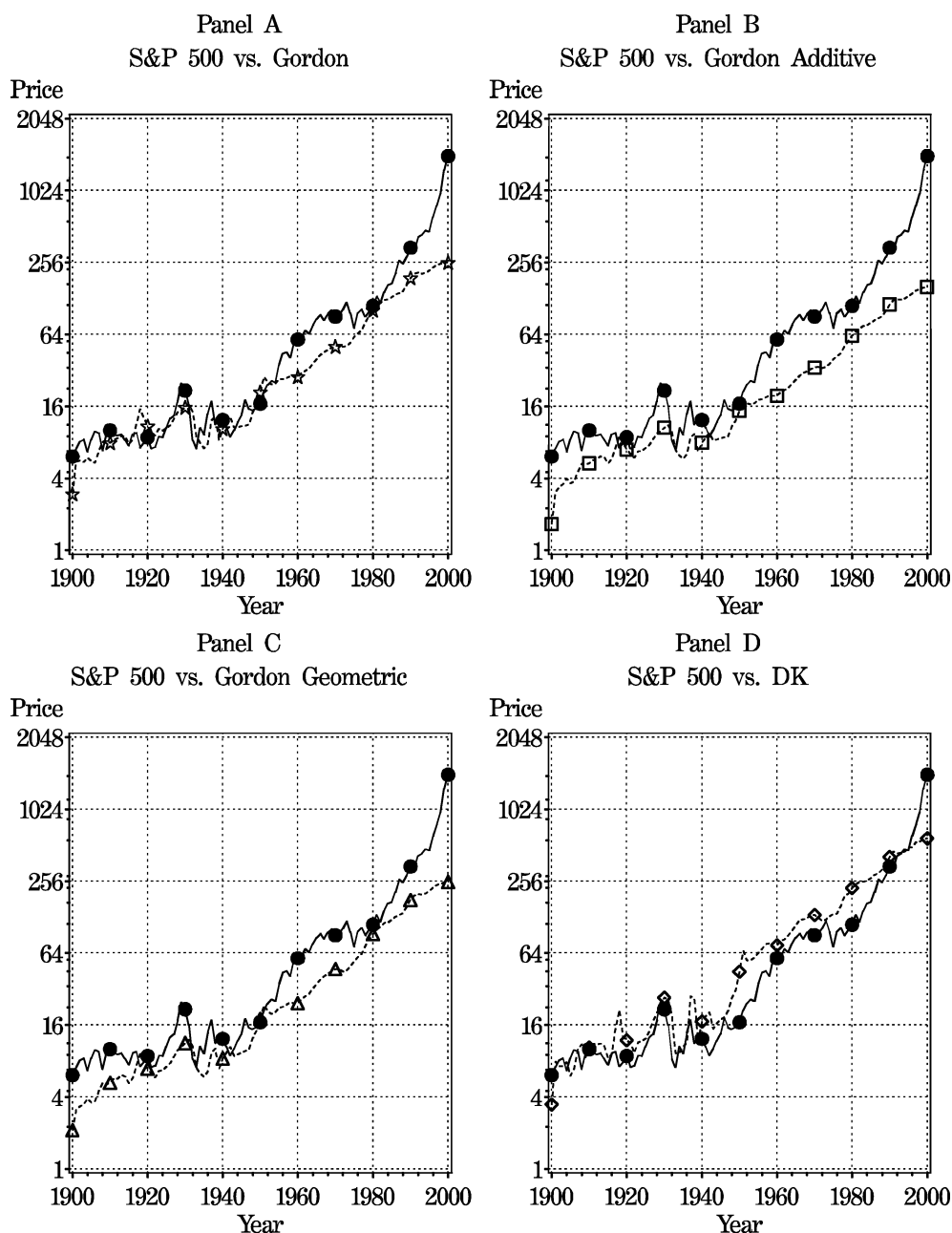
Financial Economics, Time Variation in the Market Return, Figure 1
S&P 500 **index level versus price forecasts from four models. S&P 500 index:** ●, **Gordon Growth price:** ⋆, **Additive Gordon Growth price:** □, **Geometric Gordon Growth price:** △, **DK price** ⋄

For example, Rietz [65] investigated the effect that the fear of a serious, but never realized, depression would have on equilibrium asset prices and equity premia. Jorion and Goetzmann [38] take the approach of comparing the US stock market's performance with stock market experiences in many other countries. They find that, while some mar-

kets such as the US and Canada have done very well over the past century, other countries have not been so fortunate; average stock market returns from 1921 to 1996 in France, Belgium, and Italy, for example, are all close to zero, while countries such as Spain, Greece, and Romania have experienced negative returns. It is difficult, how-

ever, to conduct statistical tests because, first, the stock indices Jorion and Goetzmann consider are largely contemporaneous and returns from the various indices are not independent. Statistical tests would have to take into account the panel nature of the data and explicitly model covariances across countries. Second, many countries in the comparison pool are difficult to compare directly to the United States in terms of economic history and underlying data generating processes. (Economies like Egypt and Romania, for example may have equity premia generated from data generating processes that differ substantially from that of the US).

There are some recent papers that make use of fundamental information in examining the equity premium. One such paper, Fama and French [21], uses historical dividend yields and other fundamental information to calculate estimates of the equity premium which are smaller than previous estimates. Fama and French obtain point estimates of the ex post equity premium ranging from 2.55% (based on dividend growth rate fundamentals) to 4.78% (based on bias-adjusted earnings growth rate fundamentals), however these estimates have large standard errors. For example, for their point estimate of 4.32% based on non-bias-adjusted earnings growth rates, a 99% confidence interval stretches from approximately −1% to about 9%. Mehra and Prescott's [51] initially troubling estimate of 6% is easily within this confidence interval and is in fact within one standard deviation of the Fama and French point estimate.

Calibrating to economy-wide dividends and discount rates, Donaldson, Kamstra, and Kramer [16] employ simulation methods similar to DK to simulate a distribution of possible price and return outcomes. Comparing these simulated distributions with moments of the actual data then permits them to test various models for the equity premium process. Could a realized equity premium of 6% be consistent with an ex ante equity premium of 2%? Could an ex ante equity premium of 2% have produced the low dividend yields, high ex post equity premia, and high Sharpe ratios observed in the US over the last half century?

A summary of the basic methodology implemented by Donaldson, Kamstra, and Kramer [16], is as follows:

(a) Assume a mean value for the equity premium that investors demand when they first purchase stock (e. g., 2%) and a time series process for the premium, say a deterministic drift downward in the premium of 5 basis points per year, asymptoting no lower than perhaps 1%. This assumed premium is added to the risk-free interest rate to determine the discount rate that an investor would rationally apply to a forecasted dividend stream in order to calculate the present value of dividend-paying stock.

(b) Estimate econometric models for the time-series processes driving dividends and interest rates in the US economy (and, if necessary, for the equity premium process), allowing for autocorrelation and covariation. Then use these models to Monte Carlo simulate a variety of potential paths for US dividends, interest rates, and equity premia. The simulated paths are of course different in each of these simulated economies because different sequences of random innovations are applied to the common stochastic processes in each case. However, the key drivers of the simulated economies themselves are all still identical to those of the US economy since all economies share common stochastic processes fitted to US data.

(c) Given the assumed process for the equity premium investors demand ex ante (which is the same for all simulated economies in a given experiment), use a discounted-dividend model to calculate the fundamental stock returns (and hence ex post equity premia) that arise in each simulated economy. All economies have the same ex ante equity premium process, and yet all economies have different ex post equity premia. Given the returns and ex post equity premia for each economy, as well as the means of the interest rates and dividend growth rates produced for each economy, it is feasible to calculate various other important characteristics, like Sharpe ratios and dividend yields.

(d) Examine the distribution of ex post equity premia, interest rates, dividend growth rates, Sharpe ratios, and dividend yields that arise conditional on various values of the ex ante equity premia. Comparing the performance of the US economy with intersections of the various univariate and multivariate distributions of these quantities and conducting joint hypothesis tests allows the determination of a narrow range of equity premia consistent with the US market data. Note that this is the method of simulated moments, which is well adapted to estimate the ex ante equity premium. The simulated method of moments was developed by McFadden [50] and Pakes and Pollard [58]. Duffie and Singleton [18] and Corradi and Swanson [11] employ simulated method of moments in an asset pricing context.

Further details on the simulation methodology are provided by Donaldson, Kamstra, and Kramer [16]. They make use of annual US stock and Treasury data observed from 1952 through 2004, with the starting year of 1952 motivated by the US Federal Reserve Board's adoption of

a modern monetary policy regime in 1951. The model that generated the data we use to illustrate this simulation methodology is Model 6 of Donaldson, Kamstra, and Kramer [16], a model that allows for trending, autocorrelated, and co-varying dividend growth rates, interest rates and equity premia, as well as for a structural break in the equity premium process. We show later that allowing for trends and structural breaks in the equity premium process is a crucial factor in the model's ability to capture the behavior of the observed US market data.

We focus on the intuition behind the Donaldson, Kamstra, and Kramer technique by looking at bivariate plots of simulated data, conditional on various values of the ex ante equity premium. In every case, the pair of statistics we plot are dependent on each other in some way, allowing us to make interesting conditional statements. Among the bivariate distributions we consider, we will see some that serve primarily to confirm the ability of our simulations to produce the character and diversity of results observed in US markets. Some sets of figures rule out ex ante equity premia below 2.5% while others rule out ex ante equity premia above 4.5%. Viewed collectively, the figures serve to confirm that the range of ex ante equity premia consistent with US market data is in the close vicinity of 3.5%.

Figure 2 contains joint distributions of mean returns and return standard deviations arising in our simulations based on four particular values of the ex ante equity premium (2.5% in Panel A, 3.5% in Panel B, 4.5% in Panel C, and 6% in Panel D). Each panel contains a scatter plot of two thousand points, with each point representing a pair of statistics (mean return and return standard deviation) arising in one of the simulated half-century economies. The combination based on the US realization is shown in each plot with a crosshair (a pair of solid straight lines with the intersection marked by a solid dot). The set of simulated pairs in each panel is surrounded by an ellipse which represents a 95% bivariate confidence bound, based on the asymptotic normality (or log-normality, where appropriate) of the plotted variables. (The 95% confidence ellipsoids are asymptotic approximations based on joint normality of the sample estimates of the moments of the simulated data. All of the sample moment estimates we consider are asymptotically normally distributed, as can be seen by appealing to the appropriate law of large numbers). The confidence ellipse for the 2.5% case is marked with diamonds, the 3.5% case with circles, the 4.5% case with squares, and the 6% case with circled crosses.

Notice that the sample mean for the US economy (the intersection of the crosshairs) lies loosely within cloud of points that depict the set of simulated economies for

each ex ante equity premium case. That is, our simulations produce mean returns and return volatility that roughly match the US observed moments of returns, *without our having calibrated to returns*. Notice also that the intersection of the crosshairs is outside (or very nearly outside) the 95% confidence ellipse in all cases except that of the 3.5% ex ante equity premium. (In unreported results that study a finer grid of ex ante equity premium values, we found that only those simulations based on values of the ex ante equity premium between about 2.5% and 4.5% lead to 95% confidence ellipses that encompass the US economy crosshairs. As the value of the ex ante equity premium falls below 2.5% or rises above 4.5%, the confidence ellipse drifts further away from the crosshairs). Based on this set of plots, we can conclude that ex ante equity premia much less than or much greater than 3.5% are inconsistent at the 5% confidence level with the observed mean return and return volatility of S&P 500 returns. χ^2 tests presented in Donaldson, Kamstra, and Kramer [16] confirm this result.

We can easily condense the information contained in these four individual plots into one plot, as shown in Panel A of Fig. 3. The scatterplot of points representing individual simulations are omitted in the condensed plot, but the confidence ellipses themselves (and the symbols used to distinguish between them) are retained. Panel A of Fig. 3 repeats the ellipses shown in Fig. 2, so that again we see that only the 3.5% ex ante equity premium case is well within the confidence ellipse at the 5% significance level. In presenting results for additional bivariate combinations, we follow the same practice, omitting the points that represent individual simulations and using the same set of symbols to distinguish between confidence ellipses based on ex ante equity premia of 2.5%, 3.5%, 4.5%, and 6%.

In Panel B of Fig. 3 we consider the four sets of confidence ellipses for mean return and mean dividend yield combinations. Notice that as we increase the ex ante equity premium, the confidence ellipses shift upward and to the right. Notice also that with higher values of the ex ante equity premium we tend to have more variable dividend yields. That is, the confidence ellipse covers a larger range of dividend yields when the value of the ex ante equity premium is larger. The observed combination of S&P 500 mean return and mean dividend yield, represented by the intersecting crosshairs, lies within the confidence ellipse for the 2.5% and 3.5% cases, very close to the ellipse for the 4.5% case, and far outside the ellipse for the 6% case.

Panel C of Fig. 3 plots confidence ellipses for mean interest rates versus mean ex post equity premia. The intersection of the crosshairs is within all four of the shown confidence ellipses. As we calibrate our model to the US interest rate, and as the ex post equity premium has a large

Panel A
Ex Ante Equity Premium of 2.50%

Panel B
Ex Ante Equity Premium of 3.50%

Panel C
Ex Ante Equity Premium of 4.50%

Panel D
Ex Ante Equity Premium of 6.00%
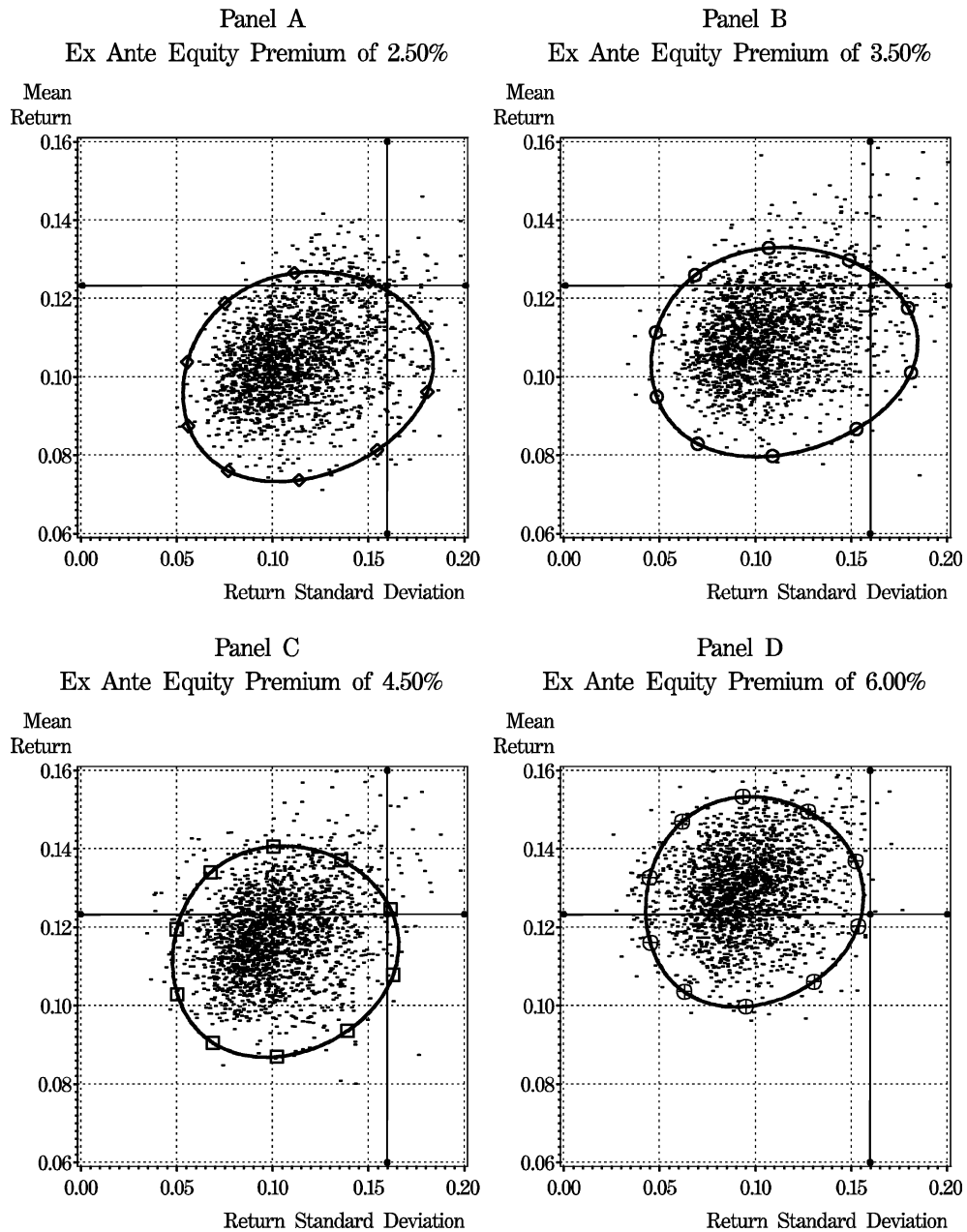
Financial Economics, Time Variation in the Market Return, Figure 2
Bivariate scatterplots of simulated data for a model allowing for trends and structural breaks. The model upon which these scatterplots are based allows for trends and structural breaks in the equity premium process, as well as autocorrelated and co-varying dividend growth rates, interest rates, and equity premia. Observed market data are indicated with crosshairs, and confidence ellipses are marked as follows. Ex ante equity premium of 2.5%: ⋄, Ex ante equity premium of 3.5%: ○, Ex ante equity premium of 4.5%: □, Ex ante equity premium of 6%: ⊕

variance, it is not surprising that the US experience is consistent with the simulated data from the entire range of ex ante equity premia considered here. This result is merely telling us that the ex post equity premium is not, by itself, particularly helpful in narrowing the possible range for the

ex ante equity premium (consistent with the empirical imprecision in measuring the ex post equity premium which has been extensively documented in the literature). Notice as well that the confidence ellipses in Panel C are all negatively sloped: we see high mean interest rates with low eq-
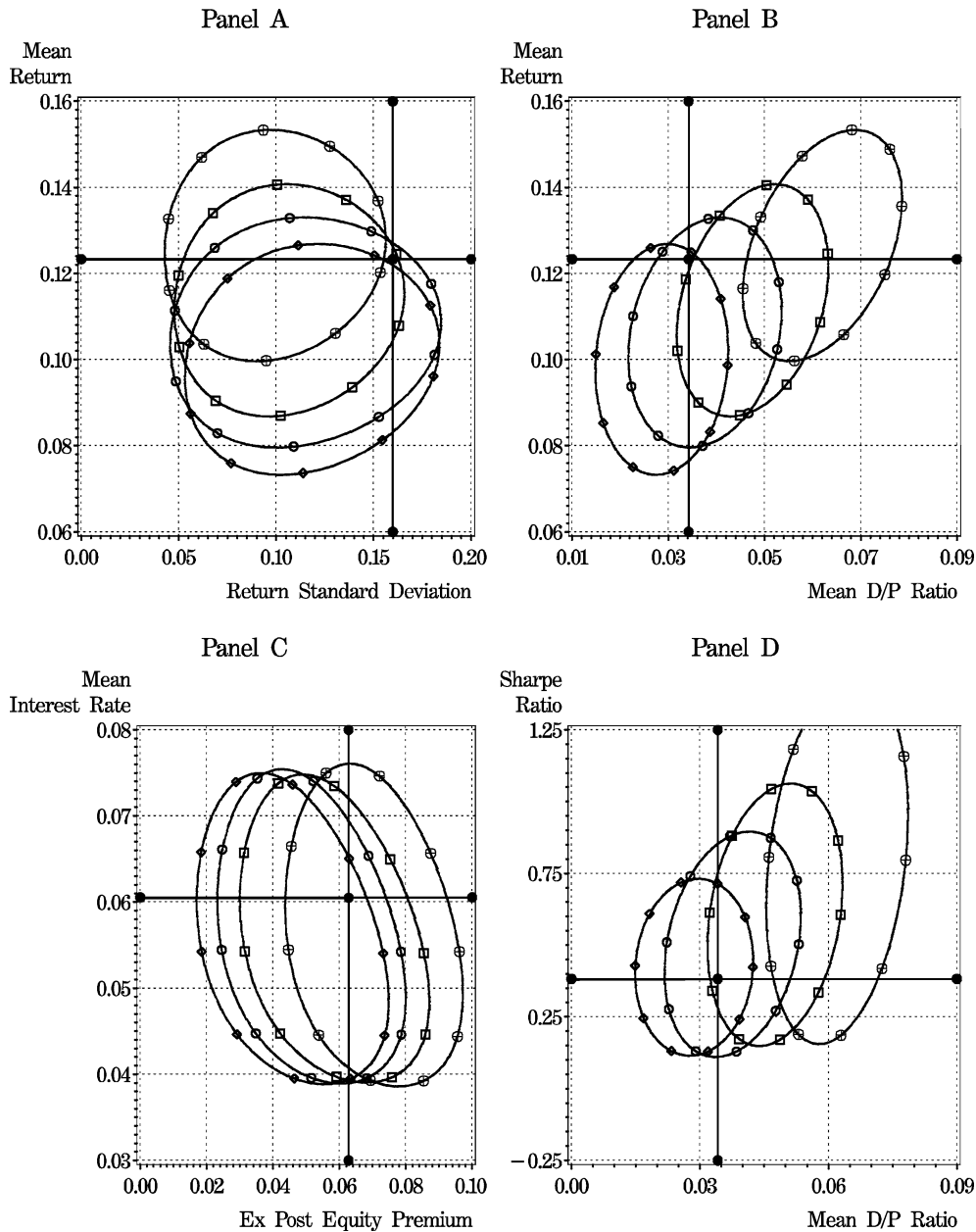
Financial Economics, Time Variation in the Market Return, Figure 3
Confidence ellipses based on simulated data for a model allowing for trends and structural breaks. The model upon which these scatterplots are based allows for trends and structural breaks in the equity premium process, as well as autocorrelated and co-varying dividend growth rates, interest rates, and equity premia. Observed market data are indicated with crosshairs, and confidence ellipses are marked as follows. 2.5% ex post equity premium: ◇, 3.5% ex post equity premium: ○, 4.5% ex post equity premium: □, 6% ex post equity premium: ⊕

uity premia and low mean interest rates with high equity premia. Many researchers, including Weil [74], have commented that the flip side of the high equity premium puzzle is the low risk-free rate puzzle. Here we confirm that the dual puzzle arises in our simulated economies as well.

It appears that this puzzle is a mechanical artifact coming out of the calculation of the premium. As the ex post equity premium equals the mean return minus the mean interest rate, a decrease in the interest rate, all else held constant, must lead to a higher ex post equity equity premium.

Panel D of Fig. 3 contains the confidence ellipses for the Sharpe ratio (or reward-to-risk ratio, calculated as the average annual difference between the arithmetic return and the risk-free rate divided by the standard deviation of the annual differences) and the mean dividend yield. As the ex ante equity premium is increased from 2.5%, the confidence ellipses shift from being centered on the crosshairs to far to the right of the crosshairs. The US expe-
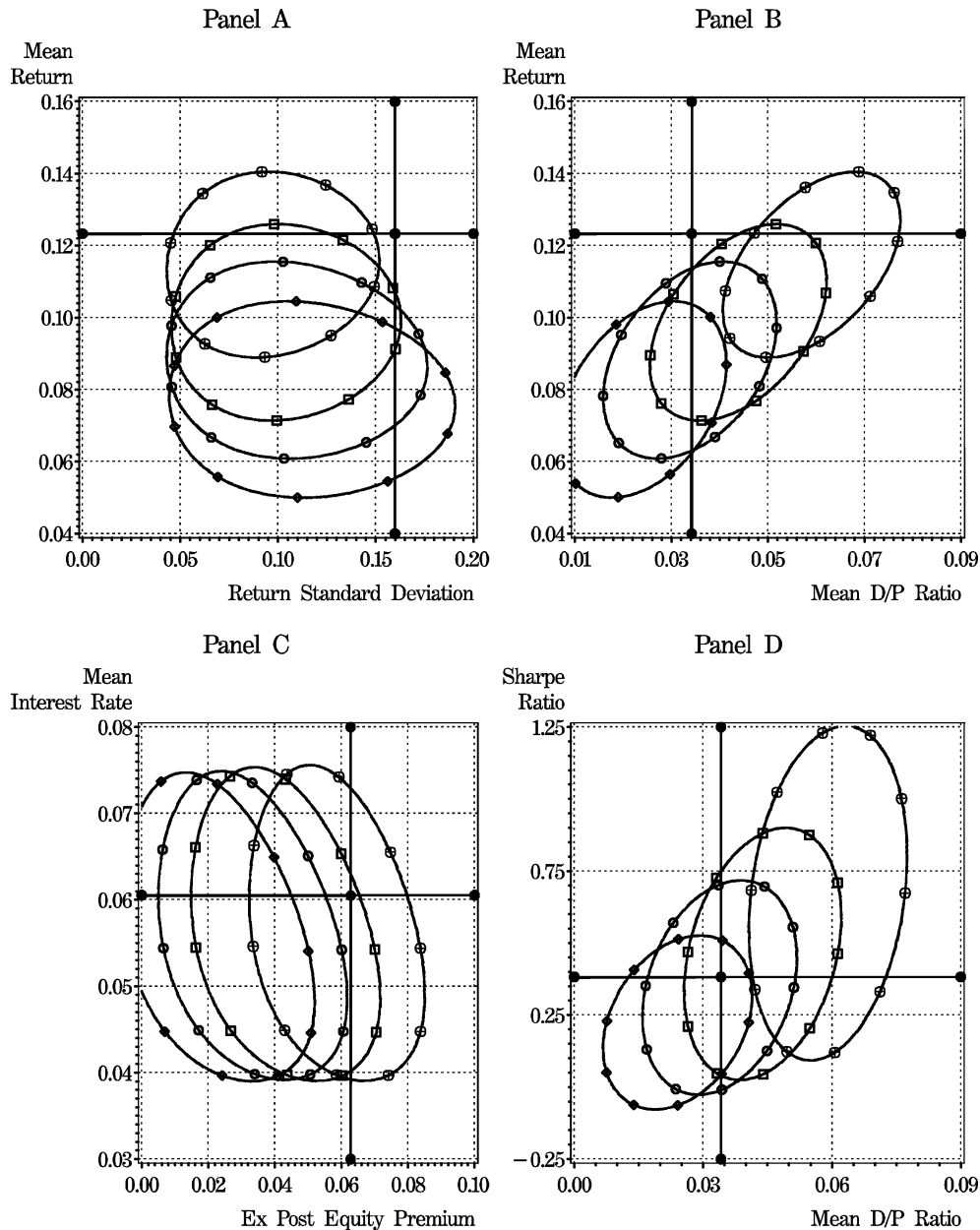
Financial Economics, Time Variation in the Market Return, Figure 4
Confidence ellipses based on simulated data for a restricted model that does not allow for trends and structural breaks. The model upon which these scatterplots are based does not allow for trends or structural breaks in the equity premium process, but does allow for autocorrelated and co-varying dividend growth rates, interest rates, and equity premia. Observed market data are indicated with crosshairs, and confidence ellipses are marked as follows. 2.5% ex post equity premium: ◇, 3.5% ex post equity premium: ○, 4.5% ex post equity premium: □, 6% ex post equity premium: ⊕

rience, indicated by the crosshairs at a Sharpe ratio of approximately 0.4 and a mean dividend yield of about 3.5%, is well outside the 95% confidence ellipse for the 6% ex ante equity premium case, suggesting a 6% ex ante equity premium is inconsistent with the jointly observed S&P 500 Sharpe ratio and mean dividend yield. Indeed Fama and French [21] and Jagannathan, McGrattan, and Scherbina [35] make reference to dividend yields to argue that the equity premium may be much smaller than 6%; our analysis gives us a glimpse of just how much smaller it might be.

Overall in Fig. 3, the joint realization of key characteristics of the US market data suggests that the true ex ante equity premium is no lower than 2.5%, no higher than 4.5%, and is most likely near 3.5%. Multivariate χ^2 tests performed by Donaldson, Kamstra, and Kramer [16] indicate a 95% confidence interval of plus-or-minus 50 basis points around 3.5%.

Consider now Fig. 4, which details simulated data from a restricted model that has a time-varying equity premium but no trends or structural breaks. Donaldson, Kamstra, and Kramer [16] study this simplified model and find that it performs poorly relative to the model we consider in Figs. 2 and 3 in terms of its ability to capture the behavior of US market data. Figure 4 shows that with the restricted model, no values of the ex ante equity premium are consistent with the observed US mean return, standard deviation, and dividend yield. That is, the simulation-based mean return and dividend yield ellipses do not contain the US data crosshairs for any value of the ex ante equity premium considered. (χ^2 tests presented in Donaldson, Kamstra, and Kramer [16] strongly support this conclusion). The implication is that it is essential to model trends and structural breaks in the equity premium process in order to accurately capture the dynamics of observed US data. Donaldson, Kamstra, and Kramer show that model failure becomes even more stark if the equity premium is constrained to be constant.

Overall, the evidence in Figs. 3 and 4 does not itself resolve the equity premium puzzle, but evidence in Fig. 3 (based on the model that allows for trends and structural breaks in the equity premium process) does provide a narrow target range of plausible equity premia that economic models should be able to explain. Additionally, the evidence in Figs. 3 and 4 points to a secondary issue ignored in the literature prior to the work of Donaldson, Kamstra, and Kramer [16], that it is crucial to model the equity premium as both time-varying and as having trends and structural breaks. We saw in Fig. 4 that high return volatility, high ex post equity premia, and low dividend yields cannot be explained easily by constant equity premium models. This result has clear implications for valuation: simple techniques that restrict the discount rate to a constant are remarkably inconsistent with the US experience of time-varying equity premia, and serious attention should be paid to modeling a time-varying rate for use in discounting future expected cash flows.

Time-Varying Equity Premia: Possible Biological Origins

To the extent that the simulation techniques considered in the previous section suggest that the equity premium varies over time, it is interesting to consider some empirical evidence of time-varying equity premia. We first survey some examples of high-frequency variations in the equity premium, and then we explore in detail two examples which may arise due to reasons that relate to human biology and/or psychology.

There is a wide range of evidence of high-frequency movement in the equity premium. At the highest frequency, we observe roughly 'U-shaped' intra-day returns (see [29,36,77]), with returns being perhaps somewhat higher during the morning trading period than in the afternoon (see [46]). At the weekly frequency, returns from Friday's close until Monday's close are low and even negative on average, as first identified by Cross [12]. Rogalski [66] found prices rose during Mondays, thus identifying the negative average realizations that followed Fridays as a weekend effect and not a Monday effect. Turning to the monthly domain, Ogden [56] documented a turn of the month effect where returns in the first half of the month are higher than returns in the second half of the month. At the annual frequency, there is the well-known turn-of-the-year effect, first shown by Rozeff and Kinney [68]. Keim [45] showed that half of the year's excess returns for small firms arose in January, and half of the January returns took place in the first five days of the month. Further, Reinganum [64] showed that January returns are higher for small firms whose price performed poorly in the previous year. All of this is consistent with the tax-loss-selling hypothesis whereby investors realize losses at the end of the tax year, leading to higher returns in January after the tax-loss selling ends.

Next we turn our attention to two cases of time-varying equity premia that may arise for reasons related to human physiology. One is Seasonal Affective Disorder (SAD), and another is daylight saving time changes.

Seasonal Affective Disorder

Past research suggests there are seasonal patterns in the equity premium which may arise due to cyclical changes in

the risk tolerance of individual investors over the course of the year related to SAD. The medical condition of SAD, according to Rosenthal [67], is a recurrent depression associated with diminished daylight in the fall, affecting many millions of Americans, as well as peoples from around the world, even those located near the equator. (In a study of 303 patients attending a primary care facility in Vancouver, Schlager, Froom, and Jaffe [70] found that 9% were clinically diagnosed with SAD and another 29% had significant winter depressive symptoms without meeting conditions for major depression. Other studies have found similar magnitudes, though some research has found that prevalence varies by latitude, with more extreme latitudes having a larger proportion of SAD-sufferers.) SAD is classified as a major depressive disorder. The symptoms of SAD include anxiety, periods of sadness, chronic fatigue, difficulty concentrating, lethargy, sleep disturbance, sugar and carbohydrate craving and associated weight gain, loss of interest in sex, and of course, clinical depression. Psychologists have shown that depressed people have less tolerance for risk in general. (See [7,32,82,83]). Psychologists refer to risk tolerance in terms of "sensation seeking" tendency, measured using a scale developed by Zuckerman [80], [81]. Those who tolerate (or seek) high levels of risk tend to score high on the sensation-seeking scale. Differences in sensation-seeking tendencies have been linked to gender (see [5] for example), race (see [31] for instance), age (see, for example, [84]), and other personal characteristics.

Economists and psychologists working together have shown that sensation-seeking tendency translates into tolerance for risk of a specifically financial or economic nature. For instance, Wong and Carducci [76] find that individuals who score low on tests of sensation seeking display greater risk aversion in making financial decisions, including the decision to purchase stocks, bonds, and insurance. Harlow and Brown [28] document the link between sensation seeking and financial risk tolerance by building on results from psychiatry which show that high blood levels of a particular enzyme are associated with depression and a lack of sensation seeking while low levels of the enzyme are associated with a high degree of sensation seeking. Harlow and Brown write, "Individuals with neurochemical activity characterized by lower levels of [the enzyme] and with a higher degree of sensation-seeking are *more willing to accept economic risk* . . . Conversely, high levels of this enzyme and a low level of sensation seeking appear to be associated with risk-averse behavior." (pp. 50–51, emphasis added). These findings suggest an individual's level of sensation seeking is indicative of his or her tolerance for financial risk.

Given these relationships, Kamstra, Kramer, and Levi [42] conjecture that during the fall and winter seasons, when a fraction of the population suffers from SAD, the proportion of risk-averse investors rises. Risk-averse investors shun risky stocks in the fall, they argue, which has a negative influence on stock prices and returns. As winter progresses and daylight becomes more plentiful, investors start to recover from their depression and become more willing to hold risky assets, at which time stock prices and returns should be positively influenced.

If the extent or severity of SAD is greater at more extreme latitudes, then the SAD effect on stock returns should be greater in stock markets at high latitudes and less in markets close to the equator. Also, the pattern of returns in the Southern Hemisphere should be the opposite of that in the Northern Hemisphere as are the seasons. Thus, Kamstra, Kramer and Levi [42] study stock market indices for the US, Sweden, Britain, Germany, Canada, New Zealand, Japan, Australia, and South Africa. They regress each country's daily stock returns on a variety of standard control variables plus two variables intended to capture the impact of SAD on returns. The first of these two variables, SAD_t, is a simple function of the length of night at the latitude of the respective market for the fall and winter months for which SAD has been documented to be most severe. The second of these variables, a fall dummy variable denoted $Fall_t$, is included because the SAD hypothesis implies the expected effect on returns is different before versus after winter solstice. Specifically, when agents initially become more risk averse, they should shun risky assets which should cause prices to be lower than would otherwise be observed, and when agents revert to normal as daylight becomes more plentiful, prices should rebound. The result should be lower returns in the autumn, higher returns in the winter, and thus a high equity premium for investors who hold through the autumn and winter periods. The $Fall_t$ dummy variable is used to capture the lower autumn returns. Both SAD_t and $Fall_t$ are appropriately defined for the Southern Hemisphere countries, accounting for the six month difference in seasons relative to the Northern Hemisphere markets.

Table 1 summarizes the average annual effect due to each of the SAD variables, SAD_t and $Fall_t$, for each of the international indices Kamstra, Kramer, and Levi [42] study. For comparison, the unconditional average annual return for each index is also provided. Observe that the annualized return due to SAD_t is positive in every country, varying from 5.7 to 17.5 percent. The SAD effect is generally larger the further are the markets from the equator. The negative annualized returns due to $Fall_t$ demonstrate the fact that SAD typically causes returns to be

Financial Economics, Time Variation in the Market Return, Table 1
Average annual percentage return due to SAD variables

Country (Index)	Annual return due to SAD_t	Annual return due to $fall_t$	Unconditional annual return
US (S&P 500)	9.2***	−3.6**	6.3***
Sweden (Veckans Affärar)	13.5**	−6.9**	17.1***
Britain (FTSE 100)	10.3**	−2.3	9.6***
Germany (DAX 30)	8.2*	−4.3**	6.5**
Canada (TSX 300)	13.2***	−4.3**	6.1***
New Zealand (Capital 40)	10.5**	−6.6**	3.3
Japan (NIKKEI 225)	6.9*	−3.7**	9.7***
Australia (All ordinaries)	5.7	0.5	8.8***
South Africa (Datastream global index)	17.5*	−2.1	14.6***

One, two, and three asterisks denote significantly different from zero at the ten, five, and one percent level respectively, based on one-sided tests. Source: Table 3 in [42].

shifted from the fall to the winter. Garrett, Kamstra, and Kramer [24] study seasonally-varying risk aversion in the context of an equilibrium asset pricing model, allowing the price of risk to vary with length of night through the fall and winter seasons. They find the risk premium on equity varies through the seasons in a manner consistent with investors being more risk averse due to SAD in the fall and winter.

Kamstra, Kramer, and Levi [43] show that there is an opposite seasonal pattern in Treasury returns relative to stock returns, consistent with time-varying risk aversion being the underlying force behind the seasonal pattern previously shown to exist in stock returns. If SAD-affected investors are shunning risky stocks in the fall as they become more risk averse, then they should be favoring safe assets at that time, which should lead to an opposite pattern in Treasury returns relative to stock returns. The seasonal cycle in the Treasury market is striking, with a variation of more than 80 basis points between the highest and lowest average monthly returns. The highest Treasury returns are observed when equity returns are lowest, and *vice versa*, which is a previously unknown pattern in Treasury returns.

Kamstra, Kramer, and Levi [43] define a new measure which is linked directly to the clinical incidence of SAD. The new measure uses data on the weekly or monthly onset of and recovery from SAD, obtained from studies of SAD patients in Vancouver and Chicago conducted by medical researchers. Young, Meaden, Fogg, Cherin, and Eastman [79] and Lam [47] document the clinical *onset* of SAD symptoms and *recovery* from SAD symptoms among North Americans known to be affected by SAD. Young et al. study 190 SAD-sufferers in Chicago and find that 74 percent of them are first diagnosed with SAD in the

weeks between mid-September and early November. Lam studies 454 SAD patients in Vancouver on a monthly basis and finds, that the peak timing of diagnosis is during the early fall. Lam [47] also studies the timing of clinical remission of SAD and finds it peaks in April, with almost half of all SAD-sufferers first experiencing complete remission in that month. March is the second most common month for subjects to first experience full remission, corresponding to almost 30 percent of subjects. For most SAD patients, the initial onset and full recovery are separated by several months over the fall and winter.

Direct use of Kamstra, Kramer, and Levi's [43] variable (which is an estimate of population-wide SAD onset/recovery based on specific samples of individuals) could impart an error-in-variables problem (see [48]), thus they utilize an instrumented version detailed in the paper, which they call Onset/Recovery, denoted $\hat{OR}_t$. The instrumented SAD measure $\hat{OR}_t$ reflects the change in the proportion of SAD-affected individuals actively suffering from SAD. The measure is defined year-round (unlike the original Kamstra, Kramer, and Levi [42], SAD_t variable, which is defined for only the fall and winter months), taking on positive values in the summer and fall and negative values in the winter and spring. Its value peaks near the fall equinox and reaches a trough near the spring equinox. (The exact monthly values of $\hat{OR}_t$ are reported by Kamstra, Kramer, and Levi [43].) The opposite signs on $\hat{OR}_t$ across the fall and winter seasons should, in principle, permit it to capture the opposite impact on equity or Treasury returns across the seasons, without use of a dummy variable. Kamstra, Kramer, and Levi [43] find that use of $\hat{OR}_t$ as a regressor to explain seasonal patterns in Treasury and equity returns renders the SAD_t and $Fall_t$ (used by Kamstra, Kramer, and Levi [42]) as economically and statisti-

cally insignificant, suggesting the Onset/Recovery variable does a far better job of explaining seasonal variation in returns than the original proxies which are not directly related to the incidence of SAD.

Kamstra, Kramer, and Levi [43] show that the seasonal Treasury and equity return patterns are unlikely to arise from macroeconomic seasonalities, seasonal variation in risk, cross-hedging between equity and Treasury markets, investor sentiment, seasonalities in the Treasury market auction schedule, seasonalities in the Treasury debt supply, seasonalities in the Federal Reserve Board's interest-rate-setting cycle, or peculiarities of the sample period considered. They find that the seasonal cycles in equity and Treasury returns become more pronounced during periods of high market volatility, consistent with time-varying risk aversion among market participants. Furthermore, they apply the White [75] reality test and find that the correlation between returns and the clinical incidence of seasonal depression cannot be easily dismissed as the simple result of data snooping.

DeGennaro, Kamstra, and Kramer [13] and Kamstra, Kramer, and Levi [13] provide further corroborating evidence for the hypothesis that SAD leads to time variation in financial markets by considering (respectively) bid-ask spreads for stocks and the flow of funds in and out of risky and safe mutual funds. In both papers they find strong support for the link between seasonal depression and time-varying risk aversion.

Daylight Saving Time Changes

The second potential biological source of time-varying equity premia we consider arises on the two dates of the year when most of the developed world shifts clocks forward or backward an hour in the name of daylight saving. Psychologists have found that changes in sleep patterns (due to shift work, jet lag, or daylight saving time changes, for example) are associated with increased anxiety, which is suggestive of a link between changes in sleep habits and time-varying risk tolerance. See [26,52], and citations found in [10] and [72] for more details on the link between sleep disruptions and anxiety. In addition to causing heightened anxiety, changes in sleep patterns also inhibit rational decision-making, lower one's information-processing ability, affect judgment, slow reaction time, and reduce problem-solving capabilities. Even a change of one hour can significantly affect behavior.

Kamstra, Kramer, and Levi [40] explore the financial market ramifications of a link between daylight saving time-change-induced disruptions in sleep patterns and individuals' tolerance for risk. They find, consistent with

psychology studies that show a gain or loss of an hour's sleep leads to increased anxiety, investors seem to shun risky stock on the trading day following a daylight saving time change. They consider stock market indexes from four countries where the time changes happen on non-overlapping dates, the US, Canada, Britain, and Germany. Based on stock market behavior over the past three decades, the authors find that the magnitude of the average return on spring daylight saving weekends is typically between two to five times that of ordinary weekends, and the effect is even stronger in the fall. Kamstra, Kramer, and Levi [41] show that the effect is not driven by a few extremely negative observations, but rather the entire distribution of returns shifts to the left following daylight saving time changes, consistent with anxious investors selling risky stock.

Future Directions

We divide our discussion in this section into three parts, one for each major topic discussed in the article.

Regarding fundamental valuation, a promising future path is to compare estimates emerging from sophisticated valuation methods to market prices, using the comparison to highlight inconsistencies in the modeling assumptions (such as restrictions on the equity premium used by the model, restrictions on the growth rate imposed for expected cash flows, and the implied values of those quantities that can be inferred from market prices). Even if one believes that markets are efficient and investors are rational, there is still much to be learned from calculating fundamentals using models and examining discrepancies relative to observed market prices.

Regarding the simulation techniques for estimating the equity premium, a promising direction for future research is to exploit these tools to forecast the volatility of stock prices. This may lead to new alternatives to existing option-implied volatility calculations and time-series techniques such as ARCH (for an overview of these methods see [15]). Another fruitful future direction would be to apply the simulation techniques to the valuation of individual companies' stock (as opposed to valuing, say, stock market indexes).

Regarding the topic of time-varying equity premia that may arise for biological reasons, a common feature of both of the examples explored in Sect. "Time-Varying Equity Premia: Possible Biological Origins", SAD and daylight-saving-time-change-induced fluctuations in the risk premium, is that in both cases the empirical evidence is based on aggregate financial market data. There is a recent trend in finance toward documenting phenomena at

the individual level, using data such as individuals' financial asset holdings and trades in their brokerage accounts. (See [1,54,55] for instance). A natural course forward is to build upon the existing aggregate market support for the prevalence of time-varying risk aversion by testing at the individual level whether risk aversion varies through the course of the year due to seasonal depression and during shorter intervals due to changes in sleep patterns. An additional potentially fruitful direction for future research is to integrate into classical asset pricing models the notion that biological factors might impact asset returns through changes in agents' degree of risk aversion. That is, human traits such as seasonal depression may lead to regularities in financial markets that are not mere anomalies; rather they may be perfectly consistent with rational agents making sensible decisions given their changing tolerance for risk. This new line of research would be similar in spirit to the work of Shefrin [71] who considers the way behavioral biases like overconfidence can be incorporated into the pricing kernel in standard asset pricing models. While the behavioral biases Shefrin considers typically involve humans making errors, the biological factors described here might be considered rational due to their involvement of time-varying risk aversion.

Bibliography

Primary Literature

1. Barber B, Odean T (2001) Boys will be boys: Gender, overconfidence, and common stock investment. Q J Econ 116: 261–292
2. Barsky RB, DeLong JB (1993) Why does the stock market fluctuate? Q J Econ 108:291–311
3. Bakshi G, Chen Z (2005) Stock valuation in dynamic economies. J Financ Market 8:115–151
4. Brealey RA, Myers SC, Allen F (2006) Principles of corporate finance, 8th edn. McGraw-Hill Irwin, New York
5. Byrnes JP, Miller DC, Schafer WD (1999) Gender differences in risk taking: A meta-analysis. Psychol Bull 125:367–383
6. Campbell JY, Kyle AS (1993) Smart money, noise trading and stock price behavior. Rev Econ Stud 60:1–34
7. Carton S, Jouvent R, Bungener C, Widlöcher D (1992) Sensation seeking and depressive mood. Pers Individ Differ 13: 843–849
8. Chiang R, Davidson I, Okuney J (1997) Some theoretical and empirical implications regarding the relationship between earnings, dividends and stock prices. J Bank Financ 21:17–35
9. Cochrane JH (2001) Asset pricing. Princeton University Press, Princeton
10. Coren S (1996) Sleep Thieves. Free Press, New York
11. Corradi V, Swanson NR (2005) Bootstrap specification tests for diffusion processes. J Econom 124:117–148
12. Cross F (1973) The behavior of stock prices on Fridays and Mondays. Financ Anal J 29:67–69
13. DeGennaro R, Kamstra MJ, Kramer LK (2005) Seasonal variation in bid-ask spreads. University of Toronto (unpublished manuscript)
14. Donaldson RG, Kamstra MJ (1996) A new dividend forecasting procedure that rejects bubbles in asset prices. Rev Financ Stud 9:333–383
15. Donaldson RG, Kamstra MJ (2005) Volatility forecasts, trading volume, and the ARCH versus option-implied volatility trade-off. J Financ Res 28:519–538
16. Donaldson RG, Kamstra MJ, Kramer LA (2007) Estimating the ex ante equity premium. University of Toronto Manuscript
17. Dong M, Hirshleifer DA (2005) A generalized earnings-based stock valuation model. Manchester School 73:1–31
18. Duffie D, Singleton KJ (1993) Simulated moments estimation of Markov models of asset prices. Econometrica 61:929–952
19. Engle RF (1982) Autoregressive conditional heteroskedasticity with estimates of the variance of United Kingdom inflation. Econometrica 50:987–1007
20. Fama EF, French KR (1993) Common risk factors in the returns on stocks and bonds. J Financ Econ 33:3–56
21. Fama EF, French KR (2002) The equity premium. J Financ 57:637–659
22. Farrell JL (1985) The dividend discount model: A primer. Financ Anal J 41:16–25
23. Feltham GA, Ohlson JA (1995) Valuation and clean surplus accounting for operating and financial activities. Contemp Account Res 11:689–731
24. Garrett I, Kamstra MJ, Kramer LK (2005) Winter blues and time variation in the price of risk. J Empir Financ 12:291–316
25. Gordon M (1962) The Investment, Financing and Valuation of the Corporation. Irwin, Homewood
26. Gordon NP, Cleary PD, Parker CE, Czeisler CA (1986) The prevalence and health impact of shiftwork. Am J Public Health 76:1225–1228
27. Hackel KS, Livnat J (1996) Cash flow and security analysis. Irwin, Chicago
28. Harlow WV, Brown KC (1990) Understanding and assessing financial risk tolerance: A biological perspective. Financ Anal J 6:50–80
29. Harris L (1986) A transaction data study of weekly and intradaily patterns in stock returns. J Financ Econ 16:99–117
30. Hawkins DF (1977) Toward an old theory of equity valuation. Financ Anal J 33:48–53
31. Hersch J (1996) Smoking, seat belts and other risky consumer decisions: differences by gender and race. Managerial Decis Econ 17:471–481
32. Horvath P, Zuckerman M (1993) Sensation seeking, risk appraisal, and risky behavior. Personal Individ Diff 14:41–52
33. Hurley WJ, Johnson LD (1994) A realistic dividend valuation model. Financ Anal J 50:50–54
34. Hurley WJ, Johnson LD (1998) Generalized Markov dividend discount models. J Portf Manag 24:27–31
35. Jagannathan R, McGrattan ER, Scherbina A (2000) The declining US equity premium. Fed Reserve Bank Minneap Q Rev 24:3–19
36. Jain PC, Joh G (1988) The dependence between hourly prices and trading volume. J Financ Quant Analysis 23:269–283
37. Jegadeesh N, Titman S (1993) Returns to buying winners and selling losers: Implications for stock market efficiency. J Financ 48:65–91

38. Jorion P, Goetzmann WN (1999) Global stock markets in the twentieth century. J Financ 54:953–980

39. Kamstra MJ (2003) Pricing firms on the basis of fundamentals. Fed Reserve Bank Atlanta Econ Rev First Quarter:49–70

40. Kamstra MJ, Kramer LA, Levi MD (2000) Losing sleep at the market: the daylight saving anomaly. Am Econ Rev 90:1005–1011

41. Kamstra MJ, Kramer LA, Levi MD (2002) Losing sleep at the market: The daylight saving anomaly: Reply. Am Econ Rev 92:1257–1263

42. Kamstra MJ, Kramer LA, Levi MD (2003) Winter blues: A SAD stock market cycle. Am Econ Rev 93:324–343

43. Kamstra MJ, Kramer LA, Levi MD (2007) Opposing Seasonalities in Treasury versus Equity Returns. University of Toronto Manuscript

44. Kamstra MJ, Kramer LA, Levi MD, Wermers R (2008) Seasonal asset allocation: evidence from mutual fund flows. University of Toronto Manuscript

45. Keim DB (1983) Size-related anomalies and stock return seasonality: Further Empirical Evidence. J Financ Econ 12:13–32

46. Kramer LA (2002) Intraday stock returns, time-varying risk premia, and diurnal mood variation. University of Toronto Manuscript

47. Lam RW (1998) Seasonal Affective Disorder: Diagnosis and management. Prim Care Psychiatry 4:63–74

48. Levi MD (1973) Errors in the variables bias in the presence of correctly measured variables. Econometrica 41:985–986

49. Liu W (2006) A liquidity-augmented capital asset pricing model. J Financ Econ 82:631–671

50. McFadden D (1989) A method of simulated moments for estimation of discrete response models without numerical integration. Econometrica 47:995–1026

51. Mehra R, Prescott EC (1985) The equity premium: A puzzle. J Monet Econ 15:145–161

52. Mellinger GD, Balter MB, Uhlenhuth EH (1985) Insomnia and its treatment: prevalence and correlates. Arch Gen Psychiatry 42:225–232

53. Michaud RO, Davis PL (1982) Valuation model bias and the scale structure of dividend discount returns. J Financ 37:563–576

54. Odean T (1998) Are investors reluctant to realize their losses? J Financ 53:1775–1798

55. Odean T (1999) Do investors trade too much? Am Econ Rev 89:1279–1298

56. Ogden JP (1990) Turn-of-month evaluations of liquid profits and stock returns: A common explanation for the monthly and January effects. J Financ 45:1259–1272

57. Ohlson JA (1995) Earnings, book values, and dividends in equity valuation. Contemp Acc Res 11:661–687

58. Pakes A, Pollard D (1989) Simulation and the asymptotics of optimization estimators. Econometrica 57:1027–1057

59. Pástor Ľ, Stambaugh R (2001) The equity premium and structural breaks. J Financ 56:1207–1239

60. Penman SH (1998) Combining earnings and book value in equity valuation. Contemp Acc Res 15:291–324

61. Penman SH, Sougiannis T (1998) A comparison of dividend, cash flow and earnings approaches to equity valuation. Contemp Acc Res 15:343–383

62. Peters DJ (1991) Using PE/Growth ratios to develop a contrarian approach to growth stocks. J Portf Manag 17:49–51

63. Rappaport A (1986) The affordable dividend approach to equity valuation. Financ Anal J 42:52–58

64. Reinganum MR (1983) The anomalous stock market behavior of small firms in January. J Financ Econ 12:89–104

65. Rietz TA (1988) The equity risk premium: A solution. J Monet Econ 22:117–31

66. Rogalski RJ (1984) New findings regarding day-of-the-week returns: A note. J Financ 35:1603–1614

67. Rosenthal NE (1998) Winter Blues: Seasonal Affective Disorder: What is It and How to Overcome It, 2nd edn. Guilford Press, New York

68. Rozeff MS, Kinney WR (1976) Capital market seasonality: The case of stock returns. J Financ Econ 3:379–402

69. Rubinstein M (1976) The valuation of uncertain income streams and the pricing of options. Bell J Econ 7:407–425

70. Schlager D, Froom J, Jaffe A (1995) Winter depression and functional impairment among ambulatory primary care patients. Compr Psychiatry 36:18–24

71. Shefrin H (2005) A Behavioral Approach to Asset Pricing. Academic Press, Oxford

72. Spira AP, Friedman L, Flint A, Sheikh J (2005) Interaction of sleep disturbances and anxiety in later life: perspectives and recommendations for future research. J Geriatr Psychiatry Neurol 18:109–115

73. Sorensen EH, Williamson DA (1985) Some evidence on the value of dividend discount models. Financ Anal J 41:60–69

74. Weil P (1989) The equity premium puzzle and the risk-free rate puzzle. J Monet Econ 24:401–421

75. White H (2000) A reality check for data snooping. Econometrica 68:1097–1126

76. Wong A, Carducci B (1991) Sensation seeking and financial risk taking in everyday money matters. J Bus Psychol 5:525–530

77. Wood RA, McInish TH, Ord JK (1985) An investigation of transactions data for NYSE stocks. J Financ 40:723–741

78. Yao Y (1997) A trinomial dividend valuation model. J Portf Manag 23:99–103

79. Young MA, Meaden PM, Fogg LF, Cherin EA, Eastman CI (1997) Which environmental variables are related to the onset of Seasonal Affective Disorder? J Abnorm Psychol 106:554–562

80. Zuckerman M (1976) Sensation seeking and anxiety, traits and states, as determinants of behavior in novel situations. In: Sarason IG, Spielberger CD (eds) Stress and Anxiety, vol 3. Hemisphere, Washington DC

81. Zuckerman M (1983) Biological Bases of Sensation Seeking, Impulsivity and Anxiety. Lawrence Erlbaum Associates, Hillsdale

82. Zuckerman M (1984) Sensation seeking: A comparative approach to a human trait. Behav Brain Sci 7:413–471

83. Zuckerman M, Buchsbaum MS, Murphy DL (1980) Sensation seeking and its biological correlates. Psychol Bull 88:187–214

84. Zuckerman M, Eysenck S, Eysenck HJ (1978) Sensation seeking in England and America: Cross-cultural, age, and sex comparisons. J Consult Clin Psychol 46:139–149

Books and Reviews

Dimson E (1988) Stock Market Anomalies. Cambridge University Press, Cambridge

Kocherlakota NR (1996) The equity premium: It's still a puzzle. J Econ Lit 34:42–71

Mehra R (2003) The equity premium: Why is it a puzzle? Financ Anal J 59:54–69

Mehra R, Prescott EC (2003) The equity premium in retrospect. In: Constantinides GM, Harris M, Stulz RM (eds) Handbook of the Economics of Finance: Financial Markets and Asset Pricing, vol 1B. North Holland, Amsterdam, pp 889–938

Penman S (2003) Financial Statement Analysis and Security Valuation, 2nd edn. McGraw-Hill/Irwin, New York

Siegel JJ, Thaler RH (1997) Anomalies: The equity premium puzzle. J Econ Perspect 11:191–200

Thaler RH (2003) The Winner's Curse: Paradoxes and Anomalies of Economic Life. Princeton University Press, Princeton

Financial Forecasting, Non-linear Time Series in

GLORIA GONZÁLEZ-RIVERA, TAE-HWY LEE
Department of Economics, University of California, Riverside, USA

Article Outline

Glossary

Arbitrage pricing theory (APT) the expected return of an asset is a linear function of a set of factors.

Artificial neural network is a nonlinear flexible functional form, connecting inputs to outputs, being capable of approximating a measurable function to any desired level of accuracy provided that sufficient complexity (in terms of number of hidden units) is permitted.

Autoregressive conditional heteroskedasticity (ARCH) the variance of an asset returns is a linear function of the past squared surprises to the asset.

Bagging short for *bootstrap aggregating*. Bagging is a method of smoothing the predictors' instability by averaging the predictors over bootstrap predictors and thus lowering the sensitivity of the predictors to training samples. A predictor is said to be unstable if perturbing the training sample can cause significant changes in the predictor.

Capital asset pricing model (CAPM) the expected return of an asset is a linear function of the covariance of the asset return with the return of the market portfolio.

Factor model a linear factor model summarizes the dimension of a large system of variables by a set of factors that are linear combinations of the original variables.

Financial forecasting prediction of prices, returns, direction, density or any other characteristic of financial assets such as stocks, bonds, options, interest rates, exchange rates, etc.

Functional coefficient model a model with time-varying and state-dependent coefficients. The number of states can be infinite.

Linearity in mean the process $\{y_t\}$ is linear in mean conditional on X_t if

$$\Pr\left[\mathbb{E}(y_t|X_t) = X_t'\theta^*\right] = 1 \quad \text{for some } \theta^* \in \mathbb{R}^k.$$

Loss (cost) function When a forecast $f_{t,h}$ of a variable Y_{t+h} is made at time t for h periods ahead, the loss (or cost) will arise if a forecast turns out to be different from the actual value. The loss function of the forecast error $e_{t+h} = Y_{t+h} - f_{t,h}$ is denoted as $c_{t+h}(Y_{t+h}, f_{t,h})$, and the function $c_{t+h}(\cdot)$ can change over t and the forecast horizon h.

Markov-switching model features parameters changing in different regimes, but in contrast with the threshold models the change is dictated by a non-observable state variable that is modelled as a hidden Markov chain.

Martingale property tomorrow's asset price is expected to be equal to today's price given some information set

$$\mathbb{E}(p_{t+1}|\mathcal{F}_t) = p_t .$$

Nonparametric regression is a data driven technique where a conditional moment of a random variable is specified as an unknown function of the data and estimated by means of a kernel or any other weighting scheme on the data.

Random field a scalar random field is defined as a function $m(\omega, x) : \Omega \times A \to R$ such that $m(\omega, x)$ is a random variable for each $x \in A$ where $A \subseteq R^k$.

Sieves the sieves or approximating spaces are approximations to an unknown function, that are dense in the original function space. Sieves can be constructed us-

ing linear spans of power series, e. g., Fourier series, splines, or many other basis functions such as artificial neural network (ANN), and various polynomials (Hermite, Laguerre, etc.).

Smooth transition models threshold model with the indicator function replaced by a smooth monotonically increasing differentiable function such as a probability distribution function.

Threshold model a nonlinear model with time-varying coefficients specified by using an indicator which takes a non-zero value when a state variable falls on a specified partition of a set of states, and zero otherwise. The number of partitions is finite.

Varying cross-sectional rank (VCR) of asset i is the proportion of assets that have a return less than or equal to the return of firm i at time t

$$z_{i,t} \equiv M^{-1} \sum_{j=1}^{M} \mathbf{1}(y_{j,t} \le y_{i,t})$$

Volatility Volatility in financial economics is often measured by the conditional variance (e. g., ARCH) or the conditional range. It is important for any decision making under uncertainty such as portfolio allocation, option pricing, risk management.

Definition of the Subject

Financial Forecasting

Financial forecasting is concerned with the prediction of prices of financial assets such as stocks, bonds, options, interest rates, exchange rates, etc. Though many agents in the economy, i. e. investors, money managers, investment banks, hedge funds, etc. are interested in the forecasting of financial prices per se, the importance of financial forecasting derives primarily from the role of financial markets within the macro economy. The development of financial instruments and financial institutions contribute to the growth and stability of the overall economy. Because of this interconnection between financial markets and the real economy, financial forecasting is also intimately linked to macroeconomic forecasting, which is concerned with the prediction of macroeconomic aggregates such as growth of the gross domestic product, consumption growth, inflation rates, commodities prices, etc. Financial forecasting and macroeconomic forecasting share many of the techniques and statistical models that will be explained in detail in this article.

In financial forecasting a major object of study is the return to a financial asset, mostly calculated as the continuously compounded return, i. e., $y_t = \log p_t - \log p_{t-1}$

where p_t is the price of the asset at time t. Nowadays financial forecasters use sophisticated techniques that combine the advances in modern finance theory, pioneered by Markowitz [113], with the advances in time series econometrics, in particular the development of nonlinear models for conditional moments and conditional quantiles of asset returns.

The aim of finance theory is to provide models for expected returns taking into account the uncertainty of the future asset payoffs. In general, financial models are concerned with investors' decisions under uncertainty. For instance the portfolio allocation problem deals with the allocation of wealth among different assets that carry different levels of risk. The implementation of these theories relies on econometric techniques that aim to estimate financial models and testing them against the data. Financial econometrics is the branch of econometrics that provides model-based statistical inference for financial variables, and therefore financial forecasting will provide their corresponding model-based predictions. However there are also econometric developments that inform the construction of ad hoc time series models that are valuable on describing the stylized facts of financial data.

Since returns $\{y_t\}$ are random variables, the aim of financial forecasting is to forecast conditional moments, quantiles, and eventually the conditional distribution of these variables. Most of the time our interest will be centered on expected returns and volatility as these two moments are crucial components on portfolio allocation problems, option valuation, and risk management, but it is also possible to forecast quantiles of a random variable, and therefore to forecast the expected probability density function. Density forecasting is the most complete forecast as it embeds all the information on the financial variable of interest. Financial forecasting is also concerned with other financial variables like durations between trades and directions of price changes. In these cases, it is also possible to construct conditional duration models and conditional probit models that are the basis for forecasting durations and market timing.

Critical to the understanding of the methodological development in financial forecasting is the statistical concept of *martingale*, which historically has its roots in the games of chance also associated with the beginnings of probability theory in the XVI century. Borrowing from the concept of fair game, financial prices are said to enjoy the *martingale property* if tomorrow's price is expected to be equal to today's price given some information set; in other words tomorrow's price has an equal chance to either move up or move down, and thus the best forecast must be the current price. The martingale property is writ-

ten as

$$\mathbb{E}(p_{t+1}|\mathcal{F}_t) = p_t$$

where $\mathbb{E}$ is the expectation operator and the information set $\mathcal{F}_t \equiv \{p_t, p_{t-1}, p_{t-2}, \dots\}$ is the collection of past and current prices, though it may also include other variables known at time t such as volume. From a forecasting point of view, the martingale model implies that changes in financial prices $(p_{t+1} - p_t)$ are not predictable.

The most restrictive form of the martingale property, proposed by Bachelier [6] in his theory of speculation is the model (in logarithms)

$$\log p_{t+1} = \mu_t + \log p_t + \varepsilon_{t+1} \, ,$$

where $\mu_t = \mu$ is a constant drift and ε_{t+1} is an identically and independently distributed (i.i.d.) error that is assumed to be normally distributed with zero mean and constant variance σ^2. This model is also known as a random walk model. Since the return is the percentage change in prices, i. e. $y_t = \log p_t - \log p_{t-1}$, an equivalent model for asset returns is

$$y_{t+1} = \mu_t + \varepsilon_{t+1} \, .$$

Then, taking conditional expectations, we find that $\mathbb{E}(y_{t+1}|\mathcal{F}_t) = \mu_t$. If the conditional mean return is not time-varying, $\mu_t = \mu$, then the returns are not forecastable based on past price information. In addition and given the assumptions on the error term, returns are independent and identically distributed random variables. These two properties, a constant drift and an i.i.d error term, are too restrictive and they rule out the possibility of any predictability in asset returns. A less restrictive and more plausible version is obtained when the i.i.d assumption is relaxed. The error term may be heteroscedastic so that returns have different (unconditional or conditional) variances and consequently they are not identically distributed, and/or the error term, though uncorrelated, may exhibit dependence in higher moments and in this case the returns are not independent random variables.

The advent of modern finance theory brings the notion of systematic risk, associated with return variances and covariances, into asset pricing. Though these theories were developed to explain the cross-sectional variability of financial returns, they also helped many years later with the construction of time series models for financial returns. Arguably, the two most important asset pricing models in modern finance theory are the Capital Asset Pricing Model (CAPM) proposed by Sharpe [137] and Lintner [103] and the Arbitrage Pricing Theory (APT)

proposed by Ross [131]. Both models claim that the expected return to an asset is a linear function of risk; in CAPM risk is related to the covariance of the asset return with the return to the market portfolio, and in APT risk is measured as exposure to a set of factors, which may include the market portfolio among others. The original version of CAPM, based on the assumption of normally distributed returns, is written as

$$\mathbb{E}(y_i) = y_f + \beta_{im} \left[\mathbb{E}(y_m) - y_f \right] \, ,$$

where y_f is the risk-free rate, y_m is the return to the market portfolio, and β_{im} is the risk of asset i defined as

$$\beta_{im} = \frac{\text{cov}(y_i, y_m)}{\text{var}(y_m)} = \frac{\sigma_{im}}{\sigma_m^2} \, .$$

This model has a time series version known as the conditional CAPM [17] that it may be useful for forecasting purposes. For asset i and given an information set as $\mathcal{F}_t = \{y_{i,t}, y_{i,t-1}, \dots; y_{m,t}, y_{m,t-1}, \dots\}$, the expected return is a linear function of a time-varying beta

$$\mathbb{E}(y_{i,t+1}|\mathcal{F}_t) = y_f + \beta_{im,t} \left[\mathbb{E}(y_{m,t+1}|\mathcal{F}_t) - y_f \right]$$

where $\beta_{im,t} = \frac{\text{cov}(y_{i,t+1}, y_{m,t+1}|\mathcal{F}_t)}{\text{var}(y_{m,t+1}|\mathcal{F}_t)} = \frac{\sigma_{im,t}}{\sigma_{m,t}^2}$. From this type of models is evident that we need to model the conditional second moments of returns jointly with the conditional mean. A general finding of this type of models is that when there is high volatility, expected returns are high, and hence forecasting volatility becomes important for the forecasting of expected returns. In the same spirit, the APT models have also conditional versions that exploit the information contained in past returns. A K-factor APT model is written as

$$y_t = c + B' f_t + \varepsilon_t \, ,$$

where f_t is a $K \times 1$ vector of factors and B is a $K \times 1$ vector of sensitivities to the factors. If the factors have time-varying second moments, it is possible to specify an APT model with a factor structure in the time-varying covariance matrix of asset returns [48], which in turn can be exploited for forecasting purposes.

The conditional CAPM and conditional APT models are fine examples on how finance theory provides a base to specify time-series models for financial returns. However there are other time series specifications, more *ad hoc* in nature, that claim that financial prices are nonlinear functions – not necessarily related to time-varying second moments – of the information set and by that, they impose some departures from the martingale property. In this

case it is possible to observe some predictability in asset prices. This is the subject of nonlinear financial forecasting. We begin with a precise definition of linearity versus nonlinearity.

Linearity and Nonlinearity

Lee, White, and Granger [99] are the first who precisely define the concept of "linearity". Let $\{Z_t\}$ be a stochastic process, and partition Z_t as $Z_t = (y_t\ X_t')'$, where (for simplicity) y_t is a scalar and X_t is a $k \times 1$ vector. X_t may (but need not necessarily) contain a constant and lagged values of y_t. LWG define that the process $\{y_t\}$ is *linear in mean conditional on* X_t if

$$\Pr\left[\mathbb{E}(y_t|X_t) = X_t'\theta^*\right] = 1 \quad \text{for some } \theta^* \in \mathbb{R}^k .$$

In the context of forecasting, Granger and Lee [71] define linearity as follows. Define $\mu_{t+h} = \mathbb{E}(y_{t+h}|\mathcal{F}_t)$ being the optimum least squares h-step forecast of y_{t+h} made at time t. μ_{t+h} will generally be a nonlinear function of the contents of $\mathcal{F}_t$. Denote m_{t+h} the optimum *linear* forecast of y_{t+h} made at time t be the best forecast that is constrained to be a linear combination of the contents of $X_t \in \mathcal{F}_t$. Granger and Lee [71] define that $\{y_t\}$ is said to be *linear in conditional mean* if μ_{t+h} is linear in X_t, i. e., $\Pr\left[\mu_{t+h} = m_{t+h}\right] = 1$ for all t and for all h. Under this definition the focus is the conditional mean and thus a process exhibiting autoregressive conditional heteroskedasticity (ARCH) [44] may nevertheless exhibit linearity of this sort because ARCH does not refer to the conditional mean. This is appropriate whenever we are concerned with the adequacy of linear models for forecasting the conditional mean returns. See [161], Section 2, for a more rigorous treatment on the definitions of linearity and nonlinearity.

This definition may be extended with some caution to the concept of linearity in higher moments and quantiles, but the definition may depend on the focus or interest of the researcher. Let $\varepsilon_{t+h} = y_{t+h} - \mu_{t+h}$ and $\sigma_{t+h}^2 = \mathbb{E}(\varepsilon_{t+h}^2|\mathcal{F}_t)$. If we consider the ARCH and GARCH as linear models, we say $\{\sigma_{t+h}^2\}$ is linear in conditional variance if σ_{t+h}^2 is a linear function of lagged ε_{t-j}^2 and σ_{t-j}^2 for some h or for all h. Alternatively, $\sigma_{t+h}^2 = \mathbb{E}(\varepsilon_{t+h}^2|\mathcal{F}_t)$ is said to be linear in conditional variance if σ_{t+h}^2 is a linear function of $x_t \in \mathcal{F}_t$ for some h or for all h. Similarly, we may consider linearity in conditional quantiles. The issue of linearity versus nonlinearity is most relevant for the conditional mean. It is more relevant whether a certain specification is correct or incorrect (rather than linear or nonlinear) for higher order conditional moments or quantiles.

Introduction

There exists a nontrivial gap between martingale difference and serial uncorrelatedness. The former implies the latter, but not vice versa. Consider a stationary time series $\{y_t\}$. Often, serial dependence of $\{y_t\}$ is described by its autocorrelation function $\rho(j)$, or by its standardized spectral density

$$h(\omega) = \frac{1}{2\pi} \sum_{j=-\infty}^{\infty} \rho(j)e^{-ij\omega}, \quad \omega \in [-\pi, \pi] .$$

Both $h(\omega)$ and $\rho(j)$ are the Fourier transform of each other, containing the same information of serial correlations of $\{y_t\}$. A problem with using $h(\omega)$ and $\rho(j)$ is that they cannot capture nonlinear time series that have zero autocorrelation but are not serially independent. Nonlinear MA and Bilinear series are good examples:

$$\text{Nonlinear MA}: \quad Y_t = be_{t-1}e_{t-2} + e_t ,$$
$$\text{Bilinear}: \quad Y_t = be_{t-1}Y_{t-2} + e_t .$$

These processes are serially uncorrelated, but they are predictable using the past information. Hong and Lee [86] note that the autocorrelation function, the variance ratios, and the power spectrum can easily miss these processes. Misleading conclusions in favor of the martingale hypothesis could be reached when these test statistics are insignificant. It is therefore important and interesting to explore whether there exists a gap between serial uncorrelatedness and martingale difference behavior for financial forecasting, and if so, whether the neglected nonlinearity in conditional mean can be explored to forecast financial asset returns.

In the forthcoming sections, we will present, without being exhaustive, nonlinear time series models for financial returns, which are the basis for nonlinear forecasting. In Sect. "Nonlinear Forecasting Models for the Conditional Mean", we review nonlinear models for the conditional mean of returns. A general representation is $y_{t+1} = \mu(y_t, y_{t-1}, \dots) + \varepsilon_{t+1}$ with $\mu(\cdot)$ a nonlinear function of the information set. If $\mathbb{E}(y_{t+1}|y_t, y_{t-1}, \dots) = \mu(y_t, y_{t-1}, \dots)$, then there is a departure from the martingale hypothesis, and past price information will be relevant to predict tomorrow's return. In Sect. "Nonlinear Forecasting Models for the Conditional Variance", we review models for the conditional variance of returns. For instance, a model like $y_{t+1} = \mu + u_{t+1}\sigma_{t+1}$ with time-varying conditional variance $\sigma_{t+1}^2 = \mathbb{E}((y_{t+1} - \mu)^2|\mathcal{F}_t)$ and i.i.d. error u_{t+1}, is still a martingale-difference for returns but it represents a departure from the

independence assumption. The conditional mean return may not be predictable but the conditional variance of the return will be. In addition, as we have seen modeling time-varying variances and covariances will be very useful for the implementation of conditional CAPM and APT models.

Nonlinear Forecasting Models for the Conditional Mean

We consider models to forecast the expected price changes of financial assets and we restrict the loss function of the forecast error to be the mean squared forecast error (MSFE). Under this loss, the optimal forecast is $\mu_{t+h} = \mathbb{E}(y_{t+h}|\mathcal{F}_t)$. Other loss functions may also be used but it will be necessary to forecast other aspects of the forecast density. For example, under a mean absolute error loss function the optimal forecast is the conditional median.

There is evidence for μ_{t+h} being time-varying. Simple linear autoregressive polynomials in lagged price changes are not sufficient to model μ_{t+h} and nonlinear specifications are needed. These can be classified into parametric and nonparametric. Examples of parametric models are autoregressive bilinear and threshold models. Examples of nonparametric models are artificial neural network, kernel and nearest neighbor regression models.

It will be impossible to have an exhaustive review of the many nonlinear specifications. However, as discussed in White [161] and Chen [25], some nonlinear models are universal approximators. For example, the sieves or approximating spaces are proven to approximate very well unknown functions and they can be constructed using linear spans of power series, Fourier series, splines, or many other basis functions such as artificial neural network (ANN), Hermite polynomials as used in e. g., [56] for modelling semi-nonparametric density, and Laguerre polynomials used in [119] for modelling the yield curve. Diebold and Li [36] and Huang, Lee, and Li [89] use the Nelson–Siegel model in forecasting yields and inflation.

We review parametric nonlinear models like threshold model, smooth transition model, Markov switching model, and random fields model; nonparametric models like local linear, local polynomial, local exponential, and functional coefficient models; and nonlinear models based on sieves like ANN and various polynomials approximations. For other nonlinear specifications we recommend some books on nonlinear time series models such as Fan and Yao [52], Gao [57], and Tsay [153]. We begin with a very simple nonlinear model.

A Simple Nonlinear Model with Dummy Variables

Goyal and Welch [66] forecast the equity premium on the S&P 500 index – index return minus T-bill rate – using many predictors such as stock-related variables (e. g., dividend-yield, earning-price ratio, book-to-market ratio, corporate issuing activity, etc.), interest-rate-related variables (e. g., treasury bills, long-term yield, corporate bond returns, inflation, investment to capital ratio), and ex ante consumption, wealth, income ratio (modified from [101]). They find that these predictors have better performance in bad times, such as the Great Depression (1930–33), the oil-shock period (1973–75), and the tech bubble-crash period (1999–2001). Also, they argue that it is reasonable to impose a lower bound (e. g., zero or 2%) on the equity premium because no investor is interested in (say) a negative premium.

Campbell and Thompson [23], inspired by the out-of-sample forecasting of Goyal and Welch [66], argue that if we impose some restrictions on the signs of the predictors' coefficients and excess return forecasts, some predictors can beat the historical average equity premium. Similarly to Goyal and Welch [66], they also use a rich set of forecasting variables – valuation ratios (e. g., dividend price ratio, earning price ratio, and book to market ratio), real return on equity, nominal interest rates and inflation, and equity share of new issues and consumption-wealth ratio. They impose two restrictions – the first one is to restrict the predictors' coefficients to have the theoretically expected sign and to set wrong-signed coefficients to zero, and the second one is to rule out a negative equity premium forecast. They show that the effectiveness of these theoretically-inspired restrictions almost always improve the out-of sample performance of the predictive regressions. This is an example where "shrinkage" works, that is to reduce the forecast error variance at the cost of a higher forecast bias but with an overall smaller mean squared forecast error (the sum of error variance and the forecast squared bias).

The results from Goyal and Welch [66] and Campbell and Thompson [23] support a simple form of nonlinearity that can be generalized to threshold models or time-varying coefficient models, which we consider next.

Threshold Models

Many financial and macroeconomic time series exhibit different characteristics over time depending upon the state of the economy. For instance, we observe bull and bear stock markets, high volatility versus low volatility periods, recessions versus expansions, credit crunch versus excess liquidity, etc. If these different regimes are present

in economic time series data, econometric specifications should go beyond linear models as these assume that there is only a single structure or regime over time. Nonlinear time series specifications that allow for the possibility of different regimes, also known as state-dependent models, include several types of models: threshold, smooth transition, and regime-switching models.

Threshold autoregressive (TAR) models [148,149] assume that the dynamics of the process is explained by an autoregression in each of the n regimes dictated by a conditioning or threshold variable. For a process $\{y_t\}$, a general specification of a TAR model is

$$y_t = \sum_{j=1}^{n} \left[\phi_o^{(j)} + \sum_{i=1}^{p_j} \phi_i^{(j)} y_{t-i} + \varepsilon_t^{(j)} \right] \mathbf{1}(r_{j-1} < x_t \leq r_j).$$

There are n regimes, in each one there is an autoregressive process of order p_j with different autoregressive parameters $\phi_i^{(j)}$, the threshold variable is x_t with r_j thresholds and $r_o = -\infty$ and $r_n = +\infty$, and the error term is assumed i.i.d. with zero mean and different variance across regimes $\varepsilon_t^{(j)} \sim$ i.i.d. $\left(0, \sigma_j^2\right)$, or more generally $\varepsilon_t^{(j)}$ is assumed to be a martingale difference. When the threshold variable is the lagged dependent variable itself y_{t-d}, the model is known as self-exciting threshold autoregressive (SETAR) model. The SETAR model has been applied to the modelling of exchange rates, industrial production indexes, and gross national product (GNP) growth, among other economic data sets. The most popular specifications within economic time series tend to find two, at most three regimes. For instance, Boero and Marrocu [18] compare a two and three-regime SETAR models with a linear AR with GARCH disturbances for the euro exchange rates. On the overall forecasting sample, the linear model performs better than the SETAR models but there is some improvement in the predictive performance of the SETAR model when conditioning on the regime.

Smooth Transition Models

In the SETAR specification, the number of regimes is discrete and finite. It is also possible to model a *continuum* of regimes as in the Smooth Transition Autoregressive (STAR) models [144]. A typical specification is

$$y_t = \phi_0 + \sum_{i=1}^{p} \phi_i y_{t-i} + \left(\theta_0 + \sum_{i=1}^{p} \theta_i y_{t-i} \right) F(y_{t-d}) + \varepsilon_t$$

where $F(y_{t-d})$ is the transition function that is continuous and in most cases is either a logistic function or an

exponential,

$$F(y_{t-d}) = \left[1 + \exp\left(-\gamma \left(y_{t-d} - r\right)\right) \right]^{-1}$$
$$F(y_{t-d}) = 1 - \left[\exp\left(-\gamma \left(y_{t-d} - r\right)^2\right) \right]$$

This model can be understood as many autoregressive regimes dictated by the values of the function $F(y_{t-d})$, or alternatively as an autoregression where the autoregressive parameters change smoothly over time. When $F(y_{t-d})$ is logistic and $\gamma \to \infty$, the STAR model collapses to a threshold model SETAR with two regimes. One important characteristic of these models, SETAR and STAR, is that the process can be stationary within some regimes and non-stationary within others moving between explosive and contractionary stages.

Since the estimation of these models can be demanding, the first question to solve is whether the nonlinearity is granted by the data. A test for linearity is imperative before engaging in the estimation of nonlinear specifications. An LM test that has power against the two alternatives specifications SETAR and STAR is proposed by Luukkonen et al. [110] and it consists of running two regressions: under the null hypothesis of linearity, a linear autoregression of order p is estimated in order to calculate the sum of squared residuals, SSE_0; the second is an auxiliary regression

$$y_t = \beta_0 + \sum_{i=1}^{P} \beta_i y_{t-i} + \sum_{i=1}^{P} \sum_{j=1}^{P} \psi_{ij} y_{t-i} y_{t-j}$$
$$+ \sum_{i=1}^{P} \sum_{j=1}^{P} \zeta_{ij} y_{t-i} y_{t-j}^2 + \sum_{i=1}^{P} \sum_{j=1}^{P} \xi_{ij} y_{t-i} y_{t-j}^3 + u_t$$

from which we calculate the sum of squared residuals, SSE_1. The test is constructed as $\chi^2 = T(SSE_0 - SSE_1)/SSE_0$ that under the null hypothesis of linearity is chi-squared distributed with $p(p + 1)/2 + 2p^2$ degrees of freedom. There are other tests in the literature, for instance Hansen [80] proposes a likelihood ratio test that has a non-standard distribution, which is approximated by implementing a bootstrap procedure. Tsay [151] proposes a test based on arranged regressions with respect to the increasing order of the threshold variable and by doing this the testing problem is transformed into a change-point problem.

If linearity is rejected, we proceed with the estimation of the nonlinear specification. In the case of the SETAR model, if we fix the values of the delay parameter d and the thresholds r_j, the model reduces to n linear regressions for which least squares estimation is straightforward.

Tsay [151] proposes a conditional least squares (CLS) estimator. For simplicity of exposition suppose that there are two regimes in the data and the model to estimate is

$$
y_t = \left[\phi_o^{(1)} + \sum_{i=1}^{p_1} \phi_i^{(1)} y_{t-i} \right] \mathbf{1}(y_{t-d} \leq r)
$$
$$
+ \left[\phi_o^{(2)} + \sum_{i=1}^{p_2} \phi_i^{(2)} y_{t-i} \right] \mathbf{1}(y_{t-d} > r) + \varepsilon_t
$$

Since r and d are fixed, we can apply least squares estimation to the model and to obtain the LS estimates for the parameters ϕ_i's. With the LS residual $\hat{\varepsilon}_t$, we obtain the total sum of squares $S(r, d) = \sum_t \hat{\varepsilon}_t^2$. The CLS estimates of r and d are obtained from $(\hat{r}, \hat{d}) = \arg \min S(r, d)$.

For the STAR model, it is also necessary to specify a priori the functional form of $F(y_{t-d})$. Teräsvirta [144] proposes a modeling cycle consisting of three stages: specification, estimation, and evaluation. In general, the specification stage consists of sequence of null hypothesis to be tested within a linearized version of the STAR model. Parameter estimation is carried out by nonlinear least squares or maximum likelihood. The evaluation stage mainly consists of testing for no error autocorrelation, no remaining nonlinearity, and parameter constancy, among other tests.

Teräsvirta and Anderson [146] find strong nonlinearity in the industrial production indexes of most of the OECD countries. The preferred model is the logistic STAR with two regimes, recessions and expansions. The dynamics in each regime are country dependent. For instance, in USA they find that the economy tends to move from recessions into expansions very aggressively but it will take a large negative shock to move rapidly from an expansion into a recession. Other references for applications of these models to financial series are found in [28,73,94].

For forecasting with STAR models, see Lundbergh and Teräsvirta [109]. It is easy to construct the one-step-ahead forecast but the multi-step-ahead forecast is a complex problem. For instance, for the 2-regime threshold model, the one-step-ahead forecast is constructed as the conditional mean of the process given some information set

$$
\mathbb{E}(y_{t+1} | \mathcal{F}_t; \theta)
$$
$$
= \left[\phi_o^{(1)} + \sum_{i=1}^{p_1} \phi_i^{(1)} y_{t+1-i} \right] \mathbf{1}(y_{t+1-d} \leq r)
$$
$$
+ \left[\phi_o^{(2)} + \sum_{i=1}^{p_2} \phi_i^{(2)} y_{t+1-i} \right] \mathbf{1}(y_{t+1-d} > r)
$$

provided that $y_{t+1-i}, y_{t+1-d} \in \mathcal{F}_t$. However, a multi-step-ahead forecast will be a function of variables that be-

ing dated at a future date do not belong to the information set; in this case the solution requires the use of numerical integration techniques or simulation/bootstrap procedures. See Granger and Teräsvirta [72], Chapter 9, and Teräsvirta [145] for more details on numerical methods for multi-step forecasts.

Markov-Switching Models

A Markov-switching (MS) model [76,77] also features changes in regime, but in contrast with the SETAR models the change is dictated by a non-observable state variable that is modelled as a Markov chain. For instance, a first order autoregressive Markov switching model is specified as

$$
y_t = c_{s_t} + \phi_{s_t} y_{t-1} + \varepsilon_t
$$

where $s_t = 1, 2, \ldots, N$ is the unobserved state variable that is modelled as an N-state Markov chain with transition probabilities $p_{ij} = P(s_t = j | s_{t-1} = i)$, and $\varepsilon_t \sim$ i.i.d. $N(0, \sigma^2)$ or more generally ε_t is a martingale difference. Conditioning in a given state and an information set $\mathcal{F}_t$, the process $\{y_t\}$ is linear but unconditionally the process is nonlinear. The conditional forecast is $\mathbb{E}(y_{t+1} | s_{t+1} = j, \mathcal{F}_t; \theta) = c_j + \phi_j y_t$ and the unconditional forecast based on observable variables is the sum of the conditional forecasts for each state weighted by the probability of being in that state,

$$
\mathbb{E}(y_{t+1} | \mathcal{F}_t; \theta)
$$
$$
= \sum_{j=1}^{N} P(s_{t+1} = j | \mathcal{F}_t; \theta) \mathbb{E}(y_{t+1} | s_{t+1} = j, \mathcal{F}_t; \theta) .
$$

The parameter vector $\theta = (c_1 \ldots c_N, \phi_1 \ldots \phi_N, \sigma^2)'$ as well as the transition probabilities p_{ij} can be estimated by maximum likelihood.

MS models have been applying to the modeling of foreign exchange rates with mixed success. Engel and Hamilton [43] fit a two-state MS for the Dollar and find that there are long swings and by that they reject the random walk behavior in the exchange rate. Marsh [114] estimates a two-state MS for the Deutschemark, the Pound Sterling, and the Japanese Yen. Though the model approximates the characteristics of the data well, the forecasting performance is poor when measured by the profit/losses generated by a set of trading rules based on the predictions of the MS model. On the contrary, Dueker and Neely [40] find that for the same exchange rate a MS model with three states variables – in the scale factor of the variance of a Student-t error, in the kurtosis of the error, and in

the expected return – produces out-of-sample excess returns that are slightly superior to those generated by common trading rules. For stock returns, there is evidence that MS models perform relatively well on describing two states in the mean (high/low returns) and two states in the variance (stable/volatile periods) of returns [111]. In addition, Perez-Quiros and Timmermann [124] propose that the error term should be modelled as a mixture of Gaussian and Student-t distributions to capture the outliers commonly found in stock returns. This model provides some gains in predictive accuracy mainly for small firms returns. For interest rates in USA, Germany, and United Kingdom, Ang and Bekaert [5] find that a two-state MS model that incorporates information on international short rate and on term spread is able to predict better than an univariate MS model. Additionally they find that in USA the classification of regimes correlates well with the business cycles.

SETAR, STAR, and MS models are successful specifications to approximate the characteristics of financial and macroeconomic data. However, good in-sample performance does not imply necessarily a good out-of-sample performance, mainly when compared to simple linear ARMA models. The success of nonlinear models depends on how prominent the nonlinearity is in the data. We should not expect a nonlinear model to perform better than a linear model when the contribution of the nonlinearity to the overall specification of the model is very small. As it is argued in Granger and Teräsvirta [72], the prediction errors generated by a nonlinear model will be smaller only when the nonlinear feature modelled in-sample is also present in the forecasting sample.

A State Dependent Mixture Model Based on Cross-sectional Ranks

In the previous section, we have dealt with nonlinear time series models that only incorporate time series information. González-Rivera, Lee, and Mishra [63] propose a nonlinear model that combines time series with cross sectional information. They propose the modelling of expected returns based on the joint dynamics of a sharp jump in the cross-sectional rank and the realized returns. They analyze the marginal probability distribution of a jump in the cross-sectional rank within the context of a duration model, and the probability of the asset return conditional on a jump specifying different dynamics depending on whether or not a jump has taken place. The resulting model for expected returns is a mixture of normal distributions weighted by the probability of jumping.

Let $y_{i,t}$ be the return of firm i at time t, and $\{y_{i,t}\}_{i=1}^M$ be the collection of asset returns of the M firms that constitute

the *market* at time t. For each time t, the asset returns are ordered from the smallest to the largest, and define $z_{i,t}$, the *Varying Cross-sectional Rank* (VCR) of firm i within the market, as the proportion of firms that have a return less than or equal to the return of firm i. We write

$$z_{i,t} \equiv M^{-1} \sum_{j=1}^M \mathbf{1}(y_{j,t} \le y_{i,t}) , \tag{1}$$

where $\mathbf{1}(\cdot)$ is the indicator function, and for M large, $z_{i,t} \in (0, 1]$. Since the rank is a highly dependent variable, it is assumed that small movements in the asset ranking will not contain significant information and that most likely large movements in ranking will be the result of news in the overall market and/or of news concerning a particular asset. Focusing on large rank movements, we define, at time t, a sharp jump as a binary variable that takes the value one when there is a minimum (upward or downward) movement of 0.5 in the ranking of asset i, and zero otherwise:

$$J_{i,t} \equiv \mathbf{1}(|z_{i,t} - z_{i,t-1}| \ge 0.5) . \tag{2}$$

A jump of this magnitude brings the asset return above or below the median of the cross-sectional distribution of returns. Note that this notion of jumps differs from the more traditional meaning of the word in the context of continuous-time modelling of the univariate return process. A jump in the cross-sectional rank implicitly depends on numerous univariate return processes.

The analytical problem now consists in modeling the joint distribution of the return $y_{i,t}$ and the jump $J_{i,t}$, i.e. $f(y_{i,t}, J_{i,t}|\mathcal{F}_{t-1})$ where $\mathcal{F}_{t-1}$ is the information set up to time $t-1$. Since $f(y_{i,t}, J_{i,t}|\mathcal{F}_{t-1}) = f_1(J_{i,t}|\mathcal{F}_{t-1})f_2(y_{i,t}|J_{i,t}, \mathcal{F}_{t-1})$, the analysis focuses first on the modelling of the marginal distribution of the jump, and subsequently on the modelling of the conditional distribution of the return.

Since $J_{i,t}$ is a Bernoulli variable, the marginal distribution of the jump is $f_1(J_{i,t}|\mathcal{F}_{t-1}) = p_{i,t}^{J_{i,t}}(1 - p_{i,t})^{(1-J_{i,t})}$ where $p_{i,t} \equiv \Pr(J_{i,t} = 1|\mathcal{F}_{t-1})$ is the conditional probability of a jump in the cross-sectional ranks. The modelling of $p_{i,t}$ is performed within the context of a dynamic duration model specified in calendar time as in Hamilton and Jordà [79]. The calendar time approach is necessary because asset returns are reported in calendar time (days, weeks, etc.) and it has the advantage of incorporating any other available information also reported in calendar time.

It is easy to see that the probability of jumping and duration must have an inverse relationship. If the probability

of jumping is high, the expected duration must be short, and vice versa. Let $\Psi_{N(t)}$ be the expected duration. The expected duration until the next jump in the cross-sectional rank is given by $\Psi_{N(t)} = \sum_{j=1}^{\infty} j(1-p_t)^{j-1} p_t = p_t^{-1}$. Note that $\sum_{j=0}^{\infty}(1-p_t)^j = p_t^{-1}$. Differentiating with respect to p_t yields $\sum_{j=0}^{\infty} -j(1-p_t)^{j-1} = -p_t^{-2}$. Multiplying by $-p_t$ gives $\sum_{j=0}^{\infty} j(1-p_t)^{j-1} p_t = p_t^{-1}$ and thus $\sum_{j=1}^{\infty} j(1-p_t)^{j-1} p_t = p_t^{-1}$. Consequently, to model $p_{i,t}$, it suffices to model the expected duration and compute its inverse. Following Hamilton and Jordà [79], an autoregressive conditional hazard (ACH) model is specified. The ACH model is a calendar-time version of the autoregressive conditional duration (ACD) of Engle and Russell [49]. In both ACD and ACH models, the expected duration is a linear function of lag durations. However as the ACD model is set up in event time, there are some difficulties on how to introduce information that arrives between events. This is not the case in the ACH model because the set-up is in calendar time. In the ACD model, the forecasting object is the expected time between events; in the ACH model, the objective is to forecast the probability that the event will happen tomorrow given the information known up to today. A general ACH model is specified as

$$\Psi_{N(t)} = \sum_{j=1}^{m} \alpha_j D_{N(t)-j} + \sum_{j=1}^{r} \beta_j \Psi_{N(t)-j} . \quad (3)$$

Since p_t is a probability, it must be bounded between zero and one. This implies that the conditional duration must have a lower bound of one. Furthermore, working in calendar time it is possible to incorporate information that becomes available between jumps and can affect the probability of a jump in future periods. The conditional hazard rate is specified as

$$p_t = [\Psi_{N(t-1)} + \delta' X_{t-1}]^{-1} , \quad (4)$$

where X_{t-1} is a vector of relevant calendar time variables such as past VCRs and past returns. This completes the marginal distribution of the jump $f_1(J_{i,t}|\mathcal{F}_{t-1}) = p_{i,t}^{J_{i,t}}(1-p_{i,t})^{(1-J_{i,t})}$.

On modelling $f_2(y_t|J_t, \mathcal{F}_{t-1}; \theta_2)$, it is assumed that the return to asset i may behave differently depending upon the occurrence of a jump. The modelling of two potential different states (whether a jump has occurred or not) will permit to differentiate whether the conditional expected return is driven by active or/and passive movements in the asset ranking in conjunction with its own return dynamics. A priori, different dynamics are possible in these two

states. A general specification is

$$f_2(y_t|J_t, \mathcal{F}_{t-1}; \theta_2) = \begin{cases} N(\mu_{1,t}, \sigma_{1,t}^2) & \text{if } J_t = 1 \\ N(\mu_{0,t}, \sigma_{0,t}^2) & \text{if } J_t = 0 \end{cases} , \quad (5)$$

where $\mu_{j,t}$ is the conditional mean and $\sigma_{j,t}^2$ the conditional variance in each state ($j = 1, 0$). Whether these two states are present in the data is an empirical question and it should be answered through statistical testing.

Combining the models for the marginal density of the jump and the conditional density of the returns, the estimation can be conducted with maximum likelihood techniques. For a sample $\{y_t, J_t\}_{t=1}^{T}$, the joint log-likelihood function is

$$\sum_{t=1}^{T} \ln f(y_t, J_t|\mathcal{F}_{t-1}; \theta)$$
$$= \sum_{t=1}^{T} \ln f_1(J_t|\mathcal{F}_{t-1}; \theta_1) + \sum_{t=1}^{T} \ln f_2(y_t|J_t, \mathcal{F}_{t-1}; \theta_2) .$$

Let us call $\mathcal{L}_1(\theta_1) = \sum_{t=1}^{T} \ln f_1(J_t|\mathcal{F}_{t-1}; \theta_1)$ and $\mathcal{L}_2(\theta_2) = \sum_{t=1}^{T} \ln f_2(y_t|J_t, \mathcal{F}_{t-1}; \theta_2)$. The maximization of the joint log-likelihood function can be achieved by maximizing $\mathcal{L}_1(\theta_1)$ and $\mathcal{L}_2(\theta_2)$ separately without loss of efficiency by assuming that the parameter vectors θ_1 and θ_2 are "variation free" in the sense of Engle et al. [45].

The log-likelihood function $\mathcal{L}_1(\theta_1) = \sum_{t=1}^{T} \ln f_1(J_t| \mathcal{F}_{t-1}; \theta_1)$ is

$$\mathcal{L}_1(\theta_1) = \sum_{t=1}^{T} \left[J_t \ln p_t(\theta_1) + (1 - J_t) \ln(1 - p_t(\theta_1)) \right], \quad (6)$$

where θ_1 includes all parameters in the conditional duration model.

The log-likelihood function $\mathcal{L}_2(\theta_2) = \sum_{t=1}^{T} \ln f_2(y_t| J_t, \mathcal{F}_{t-1}; \theta_2)$ is

$$\mathcal{L}_2(\theta_2) = \sum_{t=1}^{T} \ln \left[\frac{J_t}{\sqrt{2\pi\sigma_{1,t}^2}} \exp\left\{ -\frac{1}{2}\left(\frac{y_t - \mu_{1,t}}{\sigma_{1,t}} \right)^2 \right\} \right.$$
$$\left. + \frac{1 - J_t}{\sqrt{2\pi\sigma_{0,t}^2}} \exp\left\{ -\frac{1}{2}\left(\frac{y_t - \mu_{0,t}}{\sigma_{0,t}} \right)^2 \right\} \right],$$

where θ_2 includes all parameters in the conditional means and conditional variances under both regimes.

If the two proposed states are granted in the data, the marginal density function of the asset return must be

a mixture of two normal density functions where the mixture weights are given by the probability of jumping p_t:

$$
\begin{aligned}
g(y_t|\mathcal{F}_{t-1};\theta) &\equiv \sum_{J_t=0}^{1} f(y_t, J_t|\mathcal{F}_{t-1};\theta) \\
&= \sum_{J_t=0}^{1} f_1(J_t|\mathcal{F}_{t-1};\theta_1) f_2(y_t|J_t, \mathcal{F}_{t-1};\theta_2) \\
&= p_t \cdot f_2(y_t|J_t = 1, \mathcal{F}_{t-1};\theta_2) \\
&\quad + (1-p_t) \cdot f_2(y_t|J_t = 0, \mathcal{F}_{t-1};\theta_2),
\end{aligned}
\tag{7}
$$

as $f_1(J_t|\mathcal{F}_{t-1};\theta_1) = p_t^{J_t}(1-p_t)^{(1-J_t)}$. Therefore, the one-step ahead forecast of the return is

$$
\begin{aligned}
&\mathbb{E}(y_{t+1}|\mathcal{F}_t;\theta) \\
&= \int y_{t+1} \cdot g(y_{t+1}|\mathcal{F}_t;\theta)\mathrm{d}y_{t+1} \\
&= p_{t+1}(\theta_1) \cdot \mu_{1,t+1}(\theta_2) + (1-p_{t+1}(\theta_1)) \cdot \mu_{0,t+1}(\theta_2).
\end{aligned}
\tag{8}
$$

The expected return is a function of the probability of jumping p_t, which is a nonlinear function of the information set as shown in (4). Hence the expected returns are nonlinear functions of the information set, even in a simple case where $\mu_{1,t}$ and $\mu_{0,t}$ are linear.

This model was estimated for the returns of the constituents of the SP500 index from 1990 to 2000, and its performance was assessed in an out-of-sample exercise from 2001 to 2005 within the context of several trading strategies. Based on the one-step-ahead forecast of the mixture model, a proposed trading strategy called VCR-Mixture Trading Rule is shown to be a superior rule because of its ability to generate large risk-adjusted mean returns when compared to other technical and model-based trading rules. The VCR-Mixture Trading Rule is implemented by computing for each firm in the SP500 index the one-step ahead forecast of the return as in (8). Based on the forecasted returns $\{\hat{y}_{i,t+1}(\hat{\theta}_t)\}_{t=R}^{T-1}$, the investor predicts the VCR of all assets in relation to the overall market, that is,

$$
\hat{z}_{i,t+1} = M^{-1} \sum_{j=1}^{M} \mathbf{1}(\hat{y}_{j,t+1} \leq \hat{y}_{i,t+1}),
$$

$$
t = R, \ldots, T-1, \quad (9)
$$

and buys the top K performing assets if their forecasted return is above the risk-free rate. In every subsequent out-of-sample period ($t = R, \ldots, T-1$), the investor revises

her portfolio, selling the assets that fall out of the top performers and buying the ones that rise to the top, and she computes the one-period portfolio return

$$
\pi_{t+1} = K^{-1} \sum_{j=1}^{M} y_{j,t+1} \cdot \mathbf{1}\left(\hat{z}_{j,t+1} \geq z_{t+1}^{K}\right),
\tag{10}
$$

$$
t = R, \ldots, T-1,
$$

where z_{t+1}^{K} is the cutoff cross-sectional rank to select the K best performing stocks such that $\sum_{j=1}^{M} \mathbf{1}\left(\hat{z}_{j,t+1} \geq z_{t+1}^{K}\right) = K$. In the analysis of González-Rivera, Lee, and Mishra [63] a portfolio is formed with the top 1% ($K = 5$ stocks) performers in the SP500 index. Every asset in the portfolio is weighted equally. The evaluation criterion is to compute the "mean trading return" over the forecasting period

$$
MTR = P^{-1} \sum_{t=R}^{T-1} \pi_{t+1}.
$$

It is also possible to correct MTR according to the level of risk of the chosen portfolio. For instance, the traditional Sharpe ratio will provide the excess return per unit of risk measured by the standard deviation of the selected portfolio

$$
SR = P^{-1} \sum_{t=R}^{T-1} \frac{(\pi_{t+1} - r_{f,t+1})}{\sigma_{t+1}^{\pi}(\hat{\theta}_t)},
$$

where $r_{f,t+1}$ is the risk free rate. The VCR-Mixture Trading Rule produces a weekly MTR of 0.243% (63.295% cumulative return over 260 weeks), equivalent to a yearly compounded return of 13.45%, that is significantly more than the next most favorable rule, which is the Buy-and-Hold-the-Market Trading Rule with a weekly mean return of -0.019%, equivalent to a yearly return of -1.00%. To assess the return-risk trade off, we implement the Sharpe ratio. The largest SR (mean return per unit of standard deviation) is provided by the VCR-Mixture rule with a weekly return of 0.151% (8.11% yearly compounded return per unit of standard deviation), which is lower than the mean return provided by the same rule under the MTR criterion, but still a dominant return when compared to the mean returns provided by the Buy-and-Hold-the-Market Trading Rule.

Random Fields

Hamilton [78] proposed a flexible parametric regression model where the conditional mean has a linear parametric component and a potential nonlinear component

represented by an isotropic Gaussian random field. The model has a nonparametric flavor because no functional form is assumed but, nevertheless, the estimation is fully parametric.

A scalar random field is defined as a function $m(\omega, x)$: $\Omega \times A \rightarrow R$ such that $m(\omega, x)$ is a random variable for each $x \in A$ where $A \subseteq R^k$. A random field is also denoted as $m(x)$. If $m(x)$ is a system of random variables with finite dimensional Gaussian distributions, then the scalar random field is said to be Gaussian and it is completely determined by its mean function $\mu(x) = \mathbb{E}\left[m(x)\right]$ and its covariance function with typical element $C(x, z) = \mathbb{E}\left[(m(x) - \mu(x))(m(z) - \mu(z))\right]$ for any $x, z \in A$. The random field is said to be homogeneous or stationary if $\mu(x) = \mu$ and the covariance function depends only on the difference vector $x - z$ and we should write $C(x, z) = C(x - z)$. Furthermore, the random field is said to be isotropic if the covariance function depends on $d(x, z)$, where $d(\cdot)$ is a scalar measure of distance. In this situation we write $C(x, z) = C(d(x, z))$.

The specification suggested by Hamilton [78] can be represented as

$$y_t = \beta_0 + x_t'\beta_1 + \lambda m(g \odot x_t) + \epsilon_t , \qquad (11)$$

for $y_t \in R$ and $x_t \in R^k$, both stationary and ergodic processes. The conditional mean has a linear component given by $\beta_0 + x_t'\beta_1$ and a nonlinear component given by $\lambda m(g \odot x_t)$, where $m(z)$, for any choice of z, represents a realization of a Gaussian and homogenous random field with a moving average representation; x_t could be predetermined or exogenous and is independent of $m(\cdot)$, and ϵ_t is a sequence of independent and identically distributed $N(0, \sigma^2)$ variates independent of both $m(\cdot)$ and x_t as well as of lagged values of x_t. The scalar parameter λ represents the contribution of the nonlinear part to the conditional mean, the vector $g \in R_{0,+}^k$ drives the curvature of the conditional mean, and the symbol $\odot$ denotes element-by-element multiplication.

Let H_k be the covariance (correlation) function of the random field $m(\cdot)$ with typical element defined as $H_k(x, z) = \mathbb{E}\left[m(x)m(z)\right]$. Hamilton [78] proved that the covariance function depends solely upon the Euclidean distance between x and z, rendering the random field isotropic. For any x and $z \in R^k$, the correlation between $m(x)$ and $m(z)$ is given by the ratio of the volume of the overlap of k-dimensional unit spheroids centered at x and z to the volume of a single k-dimensional unit spheroid. If the Euclidean distance between x and z is greater than two, the correlation between $m(x)$ and $m(z)$ will be equal to zero. The general expression of the corre-

lation function is

$$H_k(h) = \begin{cases} G_{k-1}(h, 1)/G_{k-1}(0, 1) & \text{if } h \leq 1 \\ 0 & \text{if } h > 1 \end{cases}, \qquad (12)$$

$$G_k(h, r) = \int_h^r (r^2 - w^2)^{k/2} \mathrm{d}w ,$$

where $h \equiv \frac{1}{2}d_{L_2}(x, z)$, and $d_{L_2}(x, z) \equiv \left[(x-z)'(x-z)\right]^{1/2}$ is the Euclidean distance between x and z.

Within the specification (11), Dahl and González-Rivera [33] provided alternative representations of the random field that permit the construction of Lagrange multiplier tests for neglected nonlinearity, which circumvent the problem of unidentified nuisance parameters under the null of linearity and, at the same time, they are robust to the specification of the covariance function associated with the random field. They modified the Hamilton framework in two directions. First, the random field is specified in the L_1 norm instead of the L_2 norm, and secondly they considered random fields that may not have a simple moving average representation. The advantage of the L_1 norm, which is exploited in the testing problem, is that this distance measure is a linear function of the nuisance parameters, in contrast to the L_2 norm which is a nonlinear function. Logically, Dahl and González-Rivera proceeded in an opposite fashion to Hamilton. Whereas Hamilton first proposed a moving average representation of the random field, and secondly, he derived its corresponding covariance function, Dahl and González-Rivera first proposed a covariance function, and secondly they inquire whether there is a random field associated with it. The proposed covariance function is

$$C_k(h^*) = \begin{cases} (1 - h^*)^{2k} & \text{if } h^* \leq 1 \\ 0 & \text{if } h^* > 1 \end{cases}, \qquad (13)$$

where $h^* \equiv \frac{1}{2}d_{L_1}(x, z) = \frac{1}{2}|x - z|'1$. The function (13) is a permissible covariance, that is, it satisfies the positive semidefiniteness condition, which is $q'C_k q \geq 0$ for all $q \neq 0_T$. Furthermore, there is a random field associated with it according to the Khinchin's theorem (1934) and Bochner's theorem (1959). The basic argument is that the class of functions which are covariance functions of homogenous random fields coincides with the class of positive semidefinite functions. Hence, (13) being a positive semidefinite function must be the covariance function of a homogenous random field.

The estimation of these models is carried out by maximum likelihood. From model (11), we can write $y \sim N(X\beta, \lambda^2 C_k + \sigma^2 I_T)$ where $y = (y_1, y_2, \ldots, y_T)'$, $X_1 = (x_1', x_2', \ldots, x_T')'$, $X = (1 : X_1)$, $\beta = (\beta_0, \beta_1')'$, $\epsilon =$

$(\epsilon_1, \epsilon_2, \ldots, \epsilon_T)'$ and σ^2 is the variance of ϵ_t. C_k is a generic covariance function associated with the random field, which could be equal to the Hamilton spherical covariance function in (12), or to the covariance in (13). The log-likelihood function corresponding to this model is

$$\ell(\beta, \lambda^2, g, \sigma^2) = -\frac{T}{2}\log(2\pi) - \frac{1}{2}\log|\lambda^2 C_k + \sigma^2 I_T|$$
$$-\frac{1}{2}(y - X\beta)'(\lambda^2 C_k + \sigma^2 I_T)^{-1}(y - X\beta).$$
(14)

The flexible regression model has been applied successfully to detect nonlinearity in the quarterly growth rate of the US real GNP [34] and in the Industrial Production Index of sixteen OECD countries [33]. This technology is able to mimic the characteristics of the actual US business cycle. The cycle is dissected according to measures of duration, amplitude, cumulation and excess cumulation of the contraction and expansion phases. In contrast to Harding and Pagan [82] who find that nonlinear models are not uniformly superior to linear ones, the flexible regression model represents a clear improvement over linear models, and it seems to capture just the right shape of the expansion phase as opposed to Hamilton [76] and Durland and McCurdy [41] models, which tend to overestimate the cumulation measure in the expansion phase. It is found that the expansion phase must have at least two subphases: an aggressive early expansion after the trough, and a moderate/slow late expansion before the peak implying the existence of an inflexion point that we date approximately around one-third into the duration of the expansion phase. This shape lends support to parametric models of the growth rate that allow for three regimes [136], as opposed to models with just two regimes (contractions and expansions). For the Industrial Production Index, testing for nonlinearity within the flexible regression framework brings similar conclusions to those in Teräsvirta and Anderson [146], who propose parametric STAR models for industrial production data. However, the tests proposed in Dahl and González-Rivera [33], which have superior performance to detect smooth transition dynamics, seem to indicate that linearity cannot be rejected in the industrial production indexes of Japan, Austria, Belgium and Sweden as opposed to the findings of Teräsvirta and Anderson.

Nonlinear Factor Models

For the last ten years forecasting using a data-rich environment has been one of the most researched topic in economics and finance, see [140,141]. In this literature, factor models are used to reduce the dimension of the data but mostly they are linear models. Bai and Ng (BN) [7] introduce a nonlinear factor model with a quadratic principal component model as a special case. First consider a simple factor model

$$x_{it} = \lambda_i' F_t + e_{it}.$$
(15)

By the method of principal component, the elements of $\mathbf{f}_t$ are linear combinations of elements of $\mathbf{x}_t$. The factors are estimated by minimizing the sum of squared residuals of the linear model, $x_{it} = \lambda_i F_t + e_{it}$.

The factor model in (15) assumes a linear link function between the predictor $\mathbf{x}_t$ and the latent factors F_t. BN consider a more flexible approach by a nonlinear link function $g(\cdot)$ such that

$$g(x_{it}) = \phi_i' J_t + v_{it},$$

where J_t are the common factors, and ϕ_i is the vector of factor loadings. BN consider $g(x_{it})$ to be x_{it} augmented by some or all of the unique cross-products of the elements of $\{x_{it}\}_{i=1}^N$. The second-order factor model is then $x_{it}^* = \phi_i' J_t + v_{it}$ where x_{it}^* is an $N^* \times 1$ vector. Estimation of J_t then proceeds by the usual method of principal components. BN consider $x_{it}^* = \{x_{it}\, x_{it}^2\}_{i=1}^N$ with $N^* = 2N$, which they call the SPC (squared principal components).

Once the factors are estimated, the forecasting equation for y_{t+h} would be

$$y_{t+h} = (1\hat{F}_t')\boldsymbol{\gamma} + \varepsilon_t.$$

The forecasting equation remains linear whatever the link function g is. An alternative way of capturing nonlinearity is to augment the forecasting equation to include functions of the factors

$$y_{t+h} = (1\hat{F}_t')\boldsymbol{\gamma} + a(\hat{F}_t) + \varepsilon_t,$$

where $a(\cdot)$ is nonlinear. A simple case when $a(\cdot)$ is quadratic is referred to as PC2 (squared factors) in BN.

BN note that the PC2 is conceptually distinct from SPC. While the PC2 forecasting model allows the volatility of factors estimated by linear principal components to have predictive power for y, the SPC model allows the factors to be possibly nonlinear functions of the predictors while maintaining a linear relation between the factors and y. Ludvigson and Ng [108] found that the square of the first factor estimated from a set of financial factors (i. e., volatility of the first factor) is significant in the regression model for the mean excess returns. In contrast, factors estimated from the second moment of data (i. e., volatility factors) are much weaker predictors of excess returns.

Artificial Neural Network Models

Consider an augmented single hidden layer feedforward neural network model $f(x_t, \theta)$ in which the network output y_t is determined given input x_t as

$$y_t = f(x_t, \theta) + \varepsilon_t$$
$$= x_t \beta + \sum_{j=1}^{q} \delta_j \psi(x_t \gamma_j) + \varepsilon_t$$

where $\theta = (\beta' \gamma' \delta')'$, β is a conformable column vector of connection strength from the input layer to the output layer; γ_j is a conformable column vector of connection strength from the input layer to the hidden units, $j = 1, \ldots, q$; δ_j is a (scalar) connection strength from the hidden unit j to the output unit, $j = 1, \ldots, q$; and ψ is a squashing function (e. g., the logistic squasher) or a radial basis function. Input units x send signals to intermediate hidden units, then each of hidden unit produces an activation ψ that then sends signals toward the output unit. The integer q denotes the number of hidden units added to the affine (linear) network. When $q = 0$, we have a two layer *affine* network $y_t = x_t \beta + \varepsilon_t$. Hornick, Stinchcombe and White [88] show that neural network is a nonlinear flexible functional form being capable of approximating any Borel measurable function to any desired level of accuracy provided sufficiently many hidden units are available. Stinchcombe and White [138] show that this result holds for any $\psi(\cdot)$ belonging to the class of "generically comprehensively revealing" functions. These functions are "comprehensively revealing" in the sense that they can reveal arbitrary model misspecifications $\mathbb{E}(y_t|x_t) \neq f(x_t, \theta^*)$ with non-zero probability and they are "generic" in the sense that almost any choice for γ will reveal the misspecification.

We build an artificial neural network (ANN) model based on a test for neglected nonlinearity likely to have power against a range of alternatives. See White [158] and Lee, White, and Granger [99] on the neural network test and its comparison with other specification tests. The neural network test is based on a test function $h(x_t)$ chosen as the activations of 'phantom' hidden units $\psi(x_t \Gamma_j)$, $j = 1, \ldots, q$, where Γ_j are random column vectors independent of x_t. That is,

$$\mathbb{E}[\psi(x_t \Gamma_j) \varepsilon_t^* | \Gamma_j] = \mathbb{E}[\psi(x_t \Gamma_j) \varepsilon_t^*] = 0 \quad j = 1, \ldots, q,$$
(16)

under H_0, so that

$$\mathbb{E}(\Psi_t \varepsilon_t^*) = 0,$$
(17)

where $\Psi_t = (\psi(x_t \Gamma_1), \ldots, \psi(x_t \Gamma_q))'$ is a phantom hidden unit activation vector. Evidence of correlation of ε_t^* with Ψ_t is evidence against the null hypothesis that y_t is linear in mean. If correlation exists, augmenting the linear network by including an additional hidden unit with activations $\psi(x_t \Gamma_j)$ would permit an improvement in network performance. Thus the tests are based on sample correlation of affine network errors with phantom hidden unit activations,

$$n^{-1} \sum_{t=1}^{n} \Psi_t \hat{\varepsilon}_t = n^{-1} \sum_{t=1}^{n} \Psi_t (y_t - x_t \hat{\beta}).$$
(18)

Under suitable regularity conditions it follows from the central limit theorem that $n^{-1/2} \sum_{t=1}^{n} \Psi_t \hat{\varepsilon}_t \xrightarrow{d} N(0, W^*)$ as $n \to \infty$, and if one has a consistent estimator for its asymptotic covariance matrix, say $\hat{W}_n$, then an asymptotic chi-square statistic can be formed as

$$\left(n^{-1/2} \sum_{t=1}^{n} \Psi_t \hat{\varepsilon}_t \right)' \hat{W}_n^{-1} \left(n^{-1/2} \sum_{t=1}^{n} \Psi_t \hat{\varepsilon}_t \right) \xrightarrow{d} \chi^2(q).$$
(19)

Elements of Ψ_t tend to be collinear with X_t and with themselves. Thus LWG conduct a test on $q^* < q$ principal components of Ψ_t not collinear with x_t, denoted Ψ_t^*. This test is to determine whether or not there exists some advantage to be gained by adding hidden units to the affine network. We can estimate $\hat{W}_n$ robust to the conditional heteroskedasticity, or we may use with the empirical null distribution of the statistic computed by a bootstrap procedure that is robust to the conditional heteroskedasticity, e. g., wild bootstrap.

Estimation of an ANN model may be tedious and sometimes results in unreliable estimates. Recently, White [161] proposes a simple algorithm called Quick-Net, a form of "relaxed greedy algorithm" because Quick-Net searches for a single best additional hidden unit based on a sequence of OLS regressions, that may be analogous to the least angular regressions (LARS) of Efron, Hastie, Johnstone, and Tibshirani [42]. The simplicity of the QuickNet algorithm achieves the benefits of using a forecasting model that is nonlinear in the predictors while mitigating the other computational challenges to the use of nonlinear forecasting methods. See White [161], Section 5, for more details on QuickNet, and for other issues of controlling for overfit and the selection of the random parameter vectors Γ_j independent of x_t.

Campbell, Lo, and MacKinlay [22], Section 12.4, provide a review of these models. White [161] reviews

the research frontier in ANN models. Trippi and Turban [150] review the applications of ANNs to finance and investment.

Functional Coefficient Models

A functional coefficient model is introduced by Cai, Fan, and Yao [24] (CFY), with time-varying and state-dependent coefficients. It can be viewed as a special case of Priestley's [127] state-dependent model, but it includes the models of Tong [149], Chen and Tsay [26] and regime-switching models as special cases. Let $\{(y_t, s_t)'\}_{t=1}^n$ be a stationary process, where y_t and s_t are scalar variables. Also let $X_t \equiv (1, y_{t-1}, \ldots, y_{t-d})'$. We assume

$$\mathbb{E}(y_t | \mathcal{F}_{t-1}) = a_0(s_t) + \sum_{j=1}^d a_j(s_t) y_{t-j},$$

where the $\{a_j(s_t)\}$ are the autoregressive coefficients depending on s_t, which may be chosen as a function of X_t or something else. Intuitively, the functional coefficient model is an AR process with time-varying autoregressive coefficients. The coefficient functions $\{a_j(s_t)\}$ can be estimated by local linear regression. At each point s, we approximate $a_j(s_t)$ locally by a linear function $a_j(s_t) \approx a_j + b_j(s_t - s)$, $j = 0, 1, \ldots, d$, for s_t near s, where a_j and b_j are constants. The local linear estimator at point s is then given by $\hat{a}_j(s) = \hat{a}_j$, where $\{(\hat{a}_j, \hat{b}_j)\}_{j=0}^d$ minimizes the sum of local weighted squares $\sum_{t=1}^n [y_t - \mathbb{E}(y_t | \mathcal{F}_{t-1})]^2 K_h(s_t - s)$, with $K_h(\cdot) \equiv K(\cdot/h)/h$ for a given kernel function $K(\cdot)$ and bandwidth $h \equiv h_n \to 0$ as $n \to \infty$. CFY [24], p. 944, suggest to select h using a modified multi-fold "leave-one-out-type" cross-validation based on MSFE.

It is important to choose an appropriate smooth variable s_t. Knowledge on data or economic theory may be helpful. When no prior information is available, s_t may be chosen as a function of explanatory vector X_t or using such data-driven methods as AIC and cross-validation. See Fan, Yao and Cai [52] for further discussion on the choice of s_t. For exchange rate changes, Hong and Lee [85] choose s_t as the difference between the exchange rate at time $t - 1$ and the moving average of the most recent L periods of exchange rates at time $t - 1$. The moving average is a proxy for the local trend at time $t - 1$. Intuitively, this choice of s_t is expected to reveal useful information on the direction of changes.

To justify the use of the functional coefficient model, CFY [24] suggest a goodness-of-fit test for an AR(d) model against a functional coefficient model. The null hypothesis of AR(d) can be stated as

$$\mathbb{H}_0 : a_j(s_t) = \beta_j, \quad j = 0, 1, \ldots, d,$$

where β_j is the autoregressive coefficient in AR(d). Under $\mathbb{H}_0$, $\{y_t\}$ is linear in mean conditional on X_t. Under the alternative to $\mathbb{H}_0$, the autoregressive coefficients depend on s_t and the AR(d) model suffers from "neglected nonlinearity". To test $\mathbb{H}_0$, CFY compares the residual sum of squares (RSS) under $\mathbb{H}_0$

$$RSS_0 \equiv \sum_{t=1}^n \hat{\varepsilon}_t^2 = \sum_{t=1}^n \left[Y_t - \hat{\beta}_0 - \sum_{j=1}^d \hat{\beta}_j Y_{t-j} \right]^2$$

with the RSS under the alternative

$$RSS_1 \equiv \sum_{t=1}^n \tilde{\varepsilon}_t^2 = \sum_{t=1}^n \left[Y_t - \hat{a}_0(s_t) - \sum_{j=1}^d \hat{a}_j(s_t) Y_{t-j} \right]^2.$$

The test statistic is $T_n = (RSS_0 - RSS_1)/RSS_1$. We reject $\mathbb{H}_0$ for large values of T_n. CFY suggest the following bootstrap method to obtain the p-value of T_n: (i) generate the bootstrap residuals $\{\varepsilon_t^b\}_{t=1}^n$ from the centered residuals $\tilde{\varepsilon}_t - \bar{\varepsilon}$ where $\bar{\varepsilon} \equiv n^{-1} \sum_{t=1}^n \tilde{\varepsilon}_t$ and define $y_t^b \equiv X_t' \hat{\beta} + \varepsilon_t^b$, where $\hat{\beta}$ is the OLS estimator for AR(d); (ii) calculate the bootstrap statistic T_n^b using the bootstrap sample $\{y_t^b, X_t', s_t\}_{t=1}^n$; (iii) repeat steps (i) and (ii) B times ($b = 1, \ldots, B$) and approximate the bootstrap p-value of T_n by $B^{-1} \sum_{b=1}^B \mathbf{1}(T_n^b \geq T_n)$. See Hong and Lee [85] for empirical application of the functional coefficient model to forecasting foreign exchange rates.

Nonparametric Regression

Let $\{y_t, x_t\}$, $t = 1, \ldots, n$, be stochastic processes, where y_t is a scalar and $x_t = (x_{t1}, \ldots, x_{tk})$ is a $1 \times k$ vector which may contain the lagged values of y_t. Consider the regression model

$$y_t = m(x_t) + u_t$$

where $m(x_t) = \mathbb{E}(y_t | x_t)$ is the true but unknown regression function and u_t is the error term such that $\mathbb{E}(u_t | x_t) = 0$.

If $m(x_t) = g(x_t, \delta)$ is a correctly specified family of parametric regression functions then $y_t = g(x_t, \delta) + u_t$ is a correct model and, in this case, one can construct a consistent least squares (LS) estimator of $m(x_t)$ given by $g(x_t, \hat{\delta})$, where $\hat{\delta}$ is the LS estimator of the parameter δ.

In general, if the parametric regression $g(x_t, \delta)$ is incorrect or the form of $m(x_t)$ is unknown then $g(x_t, \hat{\delta})$ may not be a consistent estimator of $m(x_t)$. For this case, an alternative approach to estimate the unknown $m(x_t)$ is to use the consistent nonparametric kernel regression estimator which is essentially a local constant LS (LCLS) es-

timator. To obtain this estimator take a Taylor series expansion of $m(x_t)$ around x so that

$$y_t = m(x_t) + u_t$$
$$= m(x) + e_t$$

where $e_t = (x_t - x)m^{(1)}(x) + \frac{1}{2}(x_t - x)^2 m^{(2)}(x) + \cdots + u_t$ and $m^{(s)}(x)$ represents the sth derivative of $m(x)$ at $x_t = x$. The LCLS estimator can then be derived by minimizing

$$\sum_{t=1}^{n} e_t^2 K_{tx} = \sum_{t=1}^{n} (y_t - m(x))^2 K_{tx}$$

with respect to constant $m(x)$, where $K_{tx} = K\left(\frac{x_t - x}{h}\right)$ is a decreasing function of the distances of the regressor vector x_t from the point $x = (x_1, \ldots, x_k)$, and $h \to 0$ as $n \to \infty$ is the window width (smoothing parameter) which determines how rapidly the weights decrease as the distance of x_t from x increases. The LCLS estimator so estimated is

$$\hat{m}(x) = \frac{\sum_{t=1}^{n} y_t K_{tx}}{\sum_{t=1}^{n} K_{tx}} = (\mathbf{i}'\mathbf{K}(x)\mathbf{i})^{-1} \mathbf{i}'\mathbf{K}(x)\mathbf{y}$$

where $\mathbf{K}(x)$ is the $n \times n$ diagonal matrix with the diagonal elements K_{tx} $(t = 1, \ldots, n)$, $\mathbf{i}$ is an $n \times 1$ column vector of unit elements, and $\mathbf{y}$ is an $n \times 1$ vector with elements y_t $(t = 1, \ldots, n)$. The estimator $\hat{m}(x)$ is due to Nadaraya [118] and Watson [155] (NW) who derived this in an alternative way. Generally $\hat{m}(x)$ is calculated at the data points x_t, in which case we can write the leave-one out estimator as

$$\hat{m}(x) = \frac{\sum_{t'=1, t' \neq t}^{n} y_{t'} K_{t't}}{\sum_{t'=1, t' \neq t}^{n} K_{t't}},$$

where $K_{t't} = K\frac{x_{t'} - x_t}{h}$. The assumption that $h \to 0$ as $n \to \infty$ gives $x_t - x = O(h) \to 0$ and hence $\mathbb{E} e_t \to 0$ as $n \to \infty$. Thus the estimator $\hat{m}(x)$ will be consistent under certain smoothing conditions on h, K, and $m(x)$. In small samples however $\mathbb{E} e_t \neq 0$ so $\hat{m}(x)$ will be a biased estimator, see [122] for details on asymptotic and small sample properties.

An estimator which has a better small sample bias and hence the mean square error (MSE) behavior is the local linear LS (LLLS) estimator. In the LLLS estimator we take a first order Taylor-Series expansion of $m(x_t)$ around x so that

$$y_t = m(x_t) + u_t = m(x) + (x_t - x)m^{(1)}(x) + v_t$$
$$= \alpha(x) + x_t \beta(x) + v_t$$
$$= X_t \delta(x) + v_t$$

where $X_t = (1 \; x_t)$ and $\delta(x) = [\alpha(x) \; \beta(x)']'$ with $\alpha(x) = m(x) - x\beta(x)$ and $\beta(x) = m^{(1)}(x)$. The LLLS estimator of $\delta(x)$ is then obtained by minimizing

$$\sum_{t=1}^{n} v_t^2 K_{tx} = \sum_{t=1}^{n} (y_t - X_t \delta(x))^2 K_{tx}$$

sand it is given by

$$\tilde{\delta}(x) = (\mathbf{X}'\mathbf{K}(x)\mathbf{X})^{-1}\mathbf{X}'\mathbf{K}(x)\mathbf{y} . \tag{20}$$

where $\mathbf{X}$ is an $n \times (k + 1)$ matrix with the tth row X_t $(t = 1, \ldots, n)$.

The LLLS estimator of $\alpha(x)$ and $\beta(x)$ can be calculated as $\tilde{\alpha}(x) = (1 \; 0)\tilde{\delta}(x)$ and $\tilde{\beta}(x) = (0 \; 1)\tilde{\delta}(x)$. This gives

$$\tilde{m}(x) = (1 \; x)\tilde{\delta}(x) = \tilde{\alpha}(x) + x\tilde{\beta}(x) .$$

Obviously when $X = \mathbf{i}$, $\tilde{\delta}(x)$ reduces to the NW's LCLS estimator $\hat{m}(x)$. An extension of the LLLS is the local polynomial LS (LPLS) estimators, see [50].

In fact one can obtain the local estimators of a general nonlinear model $g(x_t, \delta)$ by minimizing

$$\sum_{t=1}^{n} [y_t - g(x_t, \delta(x))]^2 K_{tx}$$

with respect to $\delta(x)$. For $g(x_t, \delta(x)) = X_t \delta(x)$ we get the LLLS in (20). Further when $h = \infty$, $K_{tx} = K(0)$ is a constant so that the minimization of $K(0) \sum [y_t - g(x_t, \delta(x))]^2$ is the same as the minimization of $\sum [y_t - g(x_t, \delta)]^2$, that is the local LS becomes the global LS estimator $\hat{\delta}$.

The LLLS estimator in (20) can also be interpreted as the estimator of the functional coefficient (varying coefficient) linear regression model

$$y_t = m(x_t) + u_t$$
$$= X_t \delta(x_t) + u_t$$

where $\delta(x_t)$ is approximated locally by a constant $\delta(x_t) \simeq \delta(x)$. The minimization of $\sum u_t^2 K_{tx}$ with respect to $\delta(x)$ then gives the LLLS estimator in (20), which can be interpreted as the LC varying coefficient estimator. An extension of this is to consider the linear approximation $\delta(x_t) \simeq \delta(x) + D(x)(x_t - x)'$ where $D(x) = \frac{\partial \delta(x_t)}{\partial x_t'}$ evaluated at $x_t = x$. In this case

$$y_t = m(x_t) + u_t = X_t \delta(x_t) + u_t$$
$$\simeq X_t \delta(x) + X_t D(x)(x_t - x)' + u_t$$
$$= X_t \delta(x) + [(x_t - x) \otimes X_t] vec D(x) + u_t$$
$$= X_t^x \delta^x(x) + u_t$$

where $X_t^x = [X_t \; (x_t - x) \otimes X_t]$ and $\delta^x(x) = [\delta(x)' \; (vecD(x))']'$. The LL varying coefficient estimator of $\delta^x(x)$ can then be obtained by minimizing

$$\sum_{t=1}^{n} [y_t - X_t^x \delta^x(x)]^2 K_{tx}$$

with respect to $\delta^x(x)$ as

$$\dot{\delta}^x(x) = (\mathbf{X}^{x\prime}\mathbf{K}(x)\mathbf{X}^x)^{-1}\mathbf{X}^{x\prime}\mathbf{K}(x)\mathbf{y} \,. \tag{21}$$

From this $\dot{\delta}(x) = (\mathbf{I} \; 0)\dot{\delta}^x(x)$, and hence

$$\dot{m}(x) = (1 \; x \; 0)\dot{\delta}^x(x) = (1 \; x)\dot{\delta}(x) \,.$$

The above idea can be extended to the situations where $\xi_t = (x_t \; z_t)$ such that

$$\mathbb{E}(y_t | \xi_t) = m(\xi_t) = m(x_t, z_t) = X_t \delta(z_t) \,,$$

where the coefficients are varying with respect to only a subset of ξ_t; z_t is $1 \times l$ and ξ_t is $1 \times p$, $p = k + l$. Examples of these include functional coefficient autoregressive models of Chen and Tsay [26] and CFY [24], random coefficient models of Raj and Ullah [128], smooth transition autoregressive models of Granger and Teräsvirta [72], and threshold autoregressive models of Tong [149].

To estimate $\delta(z_t)$ we can again do a local constant approximation $\delta(z_t) \simeq \delta(z)$ and then minimize $\sum [y_t - X_t \delta(z)]^2 K_{tz}$ with respect to $\delta(z)$, where $K_{tz} = K(\frac{z_t - z}{h})$. This gives the LC varying coefficient estimator

$$\tilde{\delta}(z) = (\mathbf{X}'\mathbf{K}(z)\mathbf{X})^{-1}\mathbf{X}'\mathbf{K}(z)\mathbf{y} \tag{22}$$

where $\mathbf{K}(z)$ is a diagonal matrix of K_{tz}, $t = 1, \dots, n$. When $z = x$, (22) reduces to the LLLS estimator $\tilde{\delta}(x)$ in (20).

CFY [24] consider a local linear approximation $\delta(z_t) \simeq \delta(z) + D(z)(z_t - z)'$. The LL varying coefficient estimator of CFY is then obtained by minimizing

$$\sum_{t=1}^{n} [y_t - X_t \delta(z_t)]^2 K_{tz}$$
$$= \sum_{t=1}^{n} [y_t - X_t \delta(z) - [(z_t - z) \otimes X_t] vecD(z)]^2 K_{tz}$$
$$= \sum_{t=1}^{n} [y_t - X_t^z \delta^z(z)]^2 K_{tz}$$

with respect to $\delta^z(z) = [\delta(z)' \; (vecD(z))']'$ where $X_t^z = [X_t \; (z_t - z) \otimes X_t]$. This gives

$$\ddot{\delta}^z(z) = (\mathbf{X}^{z\prime}\mathbf{K}(z)\mathbf{X}^z)^{-1}\mathbf{X}^{z\prime}\mathbf{K}(z)\mathbf{y} \,, \tag{23}$$

and $\ddot{\delta}(z) = (\mathbf{I} \; 0)\ddot{\delta}^z(z)$. Hence

$$\ddot{m}(\xi) = (1 \; x \; 0)\ddot{\delta}^z(z) = (1 \; x)\ddot{\delta}(z) \,.$$

For the asymptotic properties of these varying coefficient estimators, see CFY [24]. When $z = x$, (23) reduces to the LL varying coefficient estimator $\dot{\delta}^x(x)$ in (21). See Lee and Ullah [98] for more discussion of these models and issues of testing nonlinearity.

Regime Switching Autoregressive Model Between Unit Root and Stationary Root

To avoid the usual dichotomy between unit-root non-stationarity and stationarity, we may consider models that permit two regimes of unit root nonstationarity and stationarity.

One model is the Innovation Regime-Switching (IRS) model of Kuan, Huang, and Tsay [96]. Intuitively, it may be implausible to believe that all random shocks exert only one effect (permanent or transitory) on future financial asset prices in a long time span. This intuition underpins the models that allow for breaks, stochastic unit root, or regime switching. As an alternative, Kuan, Huang, and Tsay [96] propose the IRS model that permits the random shock in each period to be permanent or transitory, depending on a switching mechanism, and hence admits distinct dynamics (unit-root nonstationarity or stationarity) in different periods. Under the IRS framework, standard unit-root models and stationarity models are just two extreme cases. By applying the IRS model to real exchange rate, they circumvent the difficulties arising from unit-root (or stationarity) testing. They allow the data to speak for themselves, rather than putting them in the straitjacket of unit-root nonstationarity or stationarity. Huang and Kuan [90] re-examine long-run PPP based on the IRS model and their empirical study on US/UK real exchange rates shows that there are both temporary and permanent influences on the real exchange rate such that approximately 42% of the shocks in the long run are more likely to have a permanent effect. They also found that transitory shocks dominate in the fixed-rate regimes, yet permanent shocks play a more important role during the floating regimes. Thus, the long-run PPP is rejected due to the presence of a significant amount of permanent shocks, but there are still long periods of time in which the deviations from long-run PPP are only transitory.

Another model is a threshold unit root (TUR) model or threshold integrated moving average (TIMA) model of Gonzalo and Martíneza [65]. Based on this model they examine whether large and small shocks have different long-

run effects, as well as whether one of them is purely transitory. They develop a new nonlinear permanent – transitory decomposition, that is applied to US stock prices to analyze the quality of the stock market.

Comparison of these two models with the linear autoregressive model with a unit root or a stationary AR model for the out-of-sample forecasting remains to be examined empirically.

Bagging Nonlinear Forecasts

To improve on unstable forecasts, bootstrap aggregating or bagging is introduced by Breiman [19]. Lee and Yang [100] show how bagging works for binary and quantile predictions. Lee and Yang [100] attributed part of the success of the bagging predictors to the small sample estimation uncertainties. Therefore, a question that may arise is that whether the good performance of bagging predictors critically depends on algorithms we employ in nonlinear estimation.

They find that bagging improves the forecasting performance of predictors on highly nonlinear regression models – e. g., artificial neural network models, especially when the sample size is limited. It is usually hard to choose the number of hidden nodes and the number of inputs (or lags), and to estimate the large number of parameters in an ANN model. Therefore, a neural network model generate poor predictions in a small sample. In such cases, bagging can do a valuable job to improve the forecasting performance as shown in [100], confirming the result of Breiman [20]. A bagging predictor is a combined predictor formed over a set of training sets to smooth out the "instability" caused by parameter estimation uncertainty and model uncertainty. A predictor is said to be "unstable" if a small change in the training set will lead to a significant change in the predictor [20].

As bagging would be valuable in nonlinear forecasting, in this section, we will show how a bagging predictor may improve the predicting performance of its underlying predictor. Let

$$\mathcal{D}_t \equiv \{(Y_s, \mathbf{X}_{s-1})\}_{s=t-R+1}^t \quad (t = R, \ldots, T)$$

be a training set at time t and let $\varphi(\mathbf{X}_t, \mathcal{D}_t)$ be a forecast of Y_{t+1} or of the binary variable $G_{t+1} \equiv \mathbf{1}(Y_{t+1} \geq 0)$ using this training set $\mathcal{D}_t$ and the explanatory variable vector $\mathbf{X}_t$. The optimal forecast $\varphi(\mathbf{X}_t, \mathcal{D}_t)$ for Y_{t+1} will be the conditional mean of Y_{t+1} given $\mathbf{X}_t$ under the squared error loss function, or the conditional quantile of Y_{t+1} on $\mathbf{X}_t$ if the loss is a tick function. Below we also consider the binary forecast for $G_{t+1} \equiv \mathbf{1}(Y_{t+1} \geq 0)$.

Suppose each training set $\mathcal{D}_t$ consists of R observations generated from the underlying probability distribution **P**. The forecast $\{\varphi(\mathbf{X}_t, \mathcal{D}_t)\}_{t=R}^T$ can be improved if more training sets were able to be generated from **P** and the forecast can be formed from averaging the multiple forecasts obtained from the multiple training sets. Ideally, if **P** were known and multiple training sets $\mathcal{D}_t^{(j)}$ ($j = 1, \ldots, J$) may be drawn from **P**, an ensemble aggregating predictor $\varphi_A(\mathbf{X}_t)$ can be constructed by the weighted averaging of $\varphi(\mathbf{X}_t, \mathcal{D}_t^{(j)})$ over j, i. e.,

$$\varphi_A(\mathbf{X}_t) \equiv \mathbb{E}_{\mathcal{D}_t} \varphi(\mathbf{X}_t, \mathcal{D}_t) \equiv \sum_{j=1}^J w_{j,t} \varphi(\mathbf{X}_t, \mathcal{D}_t^{(j)}),$$

where $\mathbb{E}_{\mathcal{D}_t}(\cdot)$ denotes the expectation over **P**, $w_{j,t}$ is the weight function with $\sum_{j=1}^J w_{j,t} = 1$, and the subscript A in φ_A denotes "aggregation".

Lee and Yang [100] show that the ensemble aggregating predictor $\varphi_A(X_t)$ has not a larger expected loss than the original predictor $\varphi(X_t, \mathcal{D}_t)$. For any convex loss function $c(\cdot)$ on the forecast error z_{t+1}, we will have

$$\mathbb{E}_{\mathcal{D}_t, Y_{t+1}, \mathbf{X}_t} c(z_{t+1}) \geq \mathbb{E}_{Y_{t+1}, \mathbf{X}_t} c(\mathbb{E}_{\mathcal{D}_t}(z_{t+1})),$$

where $\mathbb{E}_{\mathcal{D}_t}(z_{t+1})$ is the aggregating forecast error, and $\mathbb{E}_{\mathcal{D}_t, Y_{t+1}, \mathbf{X}_t}(\cdot) \equiv \mathbb{E}_{\mathbf{X}_t}[\mathbb{E}_{Y_{t+1}|\mathbf{X}_t}\{\mathbb{E}_{\mathcal{D}_t}(\cdot)|X_t\}]$ denotes the expectation $\mathbb{E}_{\mathcal{D}_t}(\cdot)$ taken over **P** (i. e., averaging over the multiple training sets generated from **P**), then taking an expectation of Y_{t+1} conditioning on X_t, and then taking an expectation of X_t. Similarly we define the notation $\mathbb{E}_{Y_{t+1}, \mathbf{X}_t}(\cdot) \equiv \mathbb{E}_{\mathbf{X}_t}[\mathbb{E}_{Y_{t+1}|\mathbf{X}_t}(\cdot)|X_t]$. Therefore, the aggregating predictor will always have no larger expected cost than the original predictor for a convex loss function $\varphi(X_t, D_t)$. The examples of the convex loss function includes the squared error loss and a tick (or check) loss $\rho_\alpha(z) \equiv [\alpha - \mathbf{1}(z < 0)]z$.

How much this aggregating predictor can improve depends on the distance between $\mathbb{E}_{\mathcal{D}_t, Y_{t+1}, \mathbf{X}_t} c(z_{t+1})$ and $\mathbb{E}_{Y_{t+1}, \mathbf{X}_t}(\mathbb{E}_{\mathcal{D}_t}(z_{t+1}))$. We can define this distance by $\Delta \equiv \mathbb{E}_{\mathcal{D}_t, Y_{t+1}, \mathbf{X}_t} c(z_{t+1}) - \mathbb{E}_{Y_{t+1}, \mathbf{X}_t} c(\mathbb{E}_{\mathcal{D}_t}(z_{t+1}))$. Therefore, the effectiveness of the aggregating predictor depends on the *convexity* of the cost function. The more convex is the cost function, the more effective this aggregating predictor can be. If the loss function is the squared error loss, then it can be shown that $\Delta = \mathbb{V}_{\mathcal{D}_t}\left[\varphi(\mathbf{X}_t, \mathcal{D}_t)\right]$ is the variance of the predictor, which measures the "instability" of the predictor. See Lee and Yang [100], Proposition 1, and Breiman [20]. If the loss is the tick function, the effectiveness of bagging is also different for different quantile predictions: bagging works better for tail-quantile predictions than for mid-quantile predictions.

In practice, however, **P** is not known. In that case we may estimate **P** by its empirical distribution, $\hat{\boldsymbol{P}}(\mathcal{D}_t)$, for a given $\mathcal{D}_t$. Then, from the empirical distribution $\hat{\boldsymbol{P}}(\mathcal{D}_t)$, multiple training sets may be drawn by the bootstrap method. Bagging predictors, $\varphi^B(\mathbf{X}_t, \mathcal{D}_t^*)$, can then be computed by taking weighted average of the predictors trained over a set of bootstrap training sets. More specifically, the bagging predictor $\varphi^B(\mathbf{X}_t, \mathcal{D}_t^*)$ can be obtained in the following steps:

1. Given a training set of data at time t, $\mathcal{D}_t \equiv \{(Y_s, \mathbf{X}_{s-1})\}_{s=t-R+1}^{t}$, construct the jth bootstrap sample $\mathcal{D}_t^{*(j)} \equiv \{(Y_s^{*(j)}, \mathbf{X}_{s-1}^{*(j)})\}_{s=t-R+1}^{t}$, $j = 1, \dots, J$, according to the empirical distribution of $\hat{\boldsymbol{P}}(\mathcal{D}_t)$ of $\mathcal{D}_t$.
2. Train the model (estimate parameters) from the jth bootstrapped sample $\mathcal{D}_t^{*(j)}$.
3. Compute the bootstrap predictor $\varphi^{*(j)}(\mathbf{X}_t, \mathcal{D}_t^{*(j)})$ from the jth bootstrapped sample $\mathcal{D}_t^{*(j)}$.
4. Finally, for mean and quantile forecast, the bagging predictor $\varphi^B(\mathbf{X}_t, \mathcal{D}_t^*)$ can be constructed by averaging over J bootstrap predictors

$$\varphi^B(\mathbf{X}_t, \mathcal{D}_t^*) \equiv \sum_{j=1}^{J} \hat{w}_{j,t} \varphi^{*(j)}(\mathbf{X}_t, \mathcal{D}_t^{*(j)}) ;$$

and for binary forecast, the bagging binary predictor $\varphi^B(\mathbf{X}_t, \mathcal{D}_t^*)$ can be constructed by majority voting over J bootstrap predictors:

$$\varphi^B(\mathbf{X}_t, \mathcal{D}_t^*) \equiv \mathbf{1}\left(\sum_{j=1}^{J} \hat{w}_{j,t} \varphi^{*(j)}(\mathbf{X}_t, \mathcal{D}_t^{*(j)}) > 1/2\right)$$

with $\sum_{j=1}^{J} \hat{w}_{j,t} = 1$ in both cases.

One concern of applying bagging to time series is whether a bootstrap can provide a sound simulation sample for dependent data, for which the bootstrap is required to be consistent. It has been shown that some bootstrap procedure (such as moving block bootstrap) can provide consistent densities for moment estimators and quantile estimators. See, e. g., Fitzenberger [54].

Nonlinear Forecasting Models for the Conditional Variance

Nonlinear Parametric Models for Volatility

Volatility models are of paramount importance in financial economics. Issues such as portfolio allocation, op-

tion pricing, risk management, and generally any decision making under uncertainty rely on the understanding and forecasting of volatility. This is one of the most active ares of research in time series econometrics. Important surveys as in Bollerslev, Chou, and Kroner [15], Bera and Higgins [13], Bollerslev, Engle, and Nelson [16], Poon and Granger [125], and Bauwens, Laurent, and Rombouts [12] attest to the variety of issues in volatility research. The motivation for the introduction of the first generation of volatility models namely the ARCH models of Engle [44] was to account for clusters of activity and fat-tail behavior of financial data. Subsequent models accounted for more complex issues. Among others and without being exclusive, we should mention issues related to asymmetric responses of volatility to news, probability distribution of the standardized innovations, i.i.d. behavior of the standardized innovation, persistence of the volatility process, linkages with continuous time models, intraday data and unevenly spaced observations, seasonality and noise in intraday data. The consequence of this research agenda has been a vast array of specifications for the volatility process.

Suppose that the return series $\{y_t\}_{t=1}^{T+1}$ of a financial asset follows the stochastic process $y_{t+1} = \mu_{t+1} + \varepsilon_{t+1}$, where $\mathbb{E}(y_{t+1}|\mathcal{F}_t) = \mu_{t+1}(\theta)$ and $\mathbb{E}(\varepsilon_{t+1}^2|\mathcal{F}_t) = \sigma_{t+1}^2(\theta)$ given the information set $\mathcal{F}_t$ (σ-field) at time t. Let $z_{t+1} \equiv \varepsilon_{t+1}/\sigma_{t+1}$ have the conditional normal distribution with zero conditional mean and unit conditional variance. Volatility models can be classified in three categories: MA family, ARCH family, and stochastic volatility (SV) family.

The simplest method to forecast volatility is to calculate a historical moving average variance, denoted as $MA(m)$, or an exponential weighted moving average (EWMA):

MA(m)	$\sigma_t^2 = \frac{1}{m}\sum_{j=1}^{m}(y_{t-j} - \hat{\mu}_t^m)^2, \quad \hat{\mu}_t^m = \frac{1}{m}\sum_{j=1}^{m} y_{t-j}$
EWMA	$\sigma_t^2 = (1-\lambda)\sum_{j=1}^{t-1} \lambda^{j-1}(y_{t-j} - \hat{\mu}_t)^2,$ $\hat{\mu}_t = \frac{1}{t-1}\sum_{j=1}^{t-1} y_{t-j}$

In the EWMA specification, a common practice is to fix the λ parameter, for instance $\lambda = 0.94$ [129]. For these two MA family models, there are not parameters to estimate.

Second, the ARCH family is very extensive with many variations on the original model $ARCH(p)$ of Engle [44]. Some representative models are: GARCH model of Bollerslev [14]; Threshold GARCH (T-GARCH) of Glosten et al. [60]; Exponential GARCH (E-GARCH) of Nelson [120]; quadratic GARCH models (Q-GARCH) as in Sentana [135]; Absolute GARCH (ABS-GARCH) of

Taylor [143] and Schwert [134] and Smooth Transition GARCH (ST-GARCH) of González-Rivera [61].

ARCH(p)	$\sigma_t^2 = \omega + \sum_{i=1}^p \alpha_i \varepsilon_{t-i}^2$
GARCH	$\sigma_t^2 = \omega + \beta \sigma_{t-1}^2 + \alpha \varepsilon_{t-1}^2$
I-GARCH	$\sigma_t^2 = \omega + \beta \sigma_{t-1}^2 + \alpha \varepsilon_{t-1}^2,\ \alpha + \beta = 1$
T-GARCH	$\sigma_t^2 = \omega + \beta \sigma_{t-1}^2 + \alpha \varepsilon_{t-1}^2 + \gamma \varepsilon_{t-1}^2 \mathbf{1}(\varepsilon_{t-1} \geq 0)$
ST-GARCH	$\sigma_t^2 = \omega + \beta \sigma_{t-1}^2 + \alpha \varepsilon_{t-1}^2 + \gamma \varepsilon_{t-1}^2 F(\varepsilon_{t-1}, \delta)$ with $F(\varepsilon_{t-1}, \delta) = [1 + \exp(\delta \varepsilon_{t-1})]^{-1} - 0.5$
E-GARCH	$\ln \sigma_t^2 = \omega + \beta \ln \sigma_{t-1}^2 + \alpha[|z_{t-1}| - c z_{t-1}]$
Q-GARCH	$\sigma_t^2 = \omega + \beta \sigma_{t-1}^2 + \alpha(\varepsilon_{t-1} + \gamma)^2$
ABS-GARCH	$\sigma_t = \omega + \beta \sigma_{t-1} + \alpha|\varepsilon_{t-1}|$

The EWMA specification can be viewed as an integrated GARCH model with $\omega = 0, \alpha = \lambda$, and $\beta = 1 - \lambda$. In the T-GARCH model, the parameter γ allows for possible asymmetric effects of positive and negative innovations. In Q-GARCH models, the parameter γ measures the extent of the asymmetry in the news impact curve. For the ST-GARCH model, the parameter γ measures the asymmetric effect of positive and negative shocks, and the parameter $\delta > 0$ measures the smoothness of the transition between regimes, with a higher value of δ making ST-GARCH closer to T-GARCH.

Third, the stationary SV model of Taylor [143] with η_t is i.i.d. N $(0, \sigma_\eta^2)$ and ξ_t is i.i.d. N$(0, \pi^2/2)$ is a representative member of the SV family.

SV	$\sigma_t^2 = \exp(0.5 h_t), \quad \ln(y_t^2) = -1.27 + h_t + \xi_t,$ $h_t = \gamma + \phi h_{t-1} + \eta_t.$

With so many models, the natural question becomes which one to choose. There is not a universal answer to this question. The best model depends upon the objectives of the user. Thus, given an objective function, we search for the model(s) with the best predictive ability controlling for possible biases due to "data snooping" [105]. To compare the relative performance of volatility models, it is customary to choose either a statistical loss function or an economic loss function.

The preferred statistical loss functions are based on moments of forecast errors (mean-error, mean-squared error, mean absolute error, etc.). The best model will minimize a function of the forecast errors. The volatility forecast is often compared to a measure of realized volatility. With financial data, the common practice has been to take squared returns as a measure of realized volatility. However, this practice is questionable. Andersen and Bollerslev [2] argued that this measure is a noisy estimate, and proposed the use of the intra-day (at each five min-

utes interval) squared returns to calculate the daily realized volatility. This measure requires intra-day data, which is subject to the variation introduced by the bid-ask spread and the irregular spacing of the price quotes.

Some other authors have evaluated the performance of volatility models with criteria based on economic loss functions. For example, West, Edison, and Cho [157] considered the problem of portfolio allocation based on models that maximize the utility function of the investor. Engle, Kane, and Noh [46] and Noh, Engle, and Kane [121] considered different volatility forecasts to maximize the trading profits in buying/selling options. Lopez [107] considered probability scoring rules that were tailored to a forecast user's decision problem and confirmed that the choice of loss function directly affected the forecast evaluation of different models. Brooks and Persand [21] evaluated volatility forecasting in a financial risk management setting in terms of Value-at-Risk (VaR). The common feature to these branches of the volatility literature is that none of these has controlled for forecast dependence across models and the inherent biases due to data-snooping.

Controlling for model dependence [160], González-Rivera, Lee, and Mishra [62] evaluate fifteen volatility models for the daily returns to the SP500 index according to their out-of-sample forecasting ability. The forecast evaluation is based, among others, on two economic loss functions: an option pricing formula and a utility function; and a statistical loss function: a goodness-of-fit based on a Value-at-Risk (VaR) calculation. For option pricing, volatility is the only component that is not observable and it needs to be estimated. The loss function assess the difference between the actual price of a call option and the estimated price, which is a function of the estimated volatility of the stock. The second economic loss function refers to the problem of wealth allocation. An investor wishes to maximize her utility allocating wealth between a risky asset and a risk-free asset. The loss function assesses the performance of the volatility estimates according to the level of utility they generate. The statistical function based on the goodness-of-fit of a VaR calculation is important for risk management. The main objective of VaR is to calculate extreme losses within a given probability of occurrence, and the estimation of the volatility is central to the VaR measure. The preferred models depend very strongly upon the loss function chosen by the user. González-Rivera, Lee, and Mishra [62] find that, for option pricing, simple models such as the exponential weighted moving average (EWMA) proposed by Riskmetrics [64] performed as well as any GARCH model. For an utility loss function, an asymmetric quadratic GARCH model is the most pre-

ferred. For VaR calculations, a stochastic volatility model dominates all other models.

Nonparametric Models for Volatility

Ziegelmann [163] considers the kernel smoothing techniques that free the traditional parametric volatility estimators from the constraints related to their specific models. He applies the nonparametric local 'exponential' estimator to estimate conditional volatility functions, ensuring its nonnegativity. Its asymptotic properties are established and compared with those for the local linear estimator for the volatility model of Fan and Yao [51]. Long, Su, and Ullah [106] extend this idea to semiparametric multivariate GARCH and show that there may exist substantial out-of-sample forecasting gain over the parametric models. This gain accounts for the presence of nonlinearity in the conditional variance-covariance that is neglected in parametric linear models.

Forecasting Volatility Using High Frequency Data

Using high-frequency data, quadratic variation may be estimated using realized volatility (RV). Andersen, Bollerslev, Diebold, and Labys [3] and Barndorff-Nielsen and Shephard [11] establish that RV, defined as the sum of squared intraday returns of small intervals, is an asymptotically unbiased estimator of the unobserved quadratic variation as the interval length approaches zero. Besides the use of high frequency information in volatility estimation, volatility forecasting using high frequency information has been addressed as well. In an application to volatility prediction, Ghysels, Santa-Clara, and Valkanov [58] investigate the predictive power of various regressors (lagged realized volatility, squared return, realized power, and daily range) for future volatility forecasting. They find that the best predictor is realized power (sum of absolute intraday returns), and more interestingly, direct use of intraday squared returns in mixed data sampling (MIDAS) regressions does not necessarily lead to better volatility forecasts.

Andersen, Bollerslev, Diebold, and Labys [4] represent another approach to forecasting volatility using RV. The model they propose is a fractional integrated AR model: ARFI(5, d) for logarithmic RV's obtained from foreign exchange rates data of 30-minute frequency and demonstrate the superior predictive power of their model.

Alternatively, Corsi [32] proposes the heterogeneous autoregressive (HAR) model of RV, which is able to reproduce long memory. McAleer and Medeiros [115] propose a new model that is a multiple regime smooth transition (ST) extension of the HAR model, which is specifically designed to model the behavior of the volatility inherent in financial time series. The model is able to describe simultaneously long memory as well as sign and size asymmetries. They apply the model to several Dow Jones Industrial Average index stocks using transaction level data from the Trades and Quotes database that covers ten years of data, and find strong support for long memory and both sign and size asymmetries. Furthermore, they show that the multiple regime smooth transition HAR model, when combined with the linear HAR model, is flexible for the purpose of forecasting volatility.

Forecasting Beyond Mean and Variance

In the previous section, we have surveyed the major developments in nonlinear time series, mainly modeling the conditional mean and the conditional variance of financial returns. However it is not clear yet that any of those nonlinear models may generate profits after accounting for various market frictions and transactions costs. Therefore, some research efforts have been directed to investigate other aspects of the conditional density of returns such as higher moments, quantiles, directions, intervals, and the density itself. In this section, we provide a brief survey on forecasting these other features.

Forecasting Quantiles

The optimal forecast of a time series model depends on the specification of the loss function. A symmetric quadratic loss function is the most prevalent in applications due to its simplicity. Under symmetric quadratic loss, the optimal forecast is simply the conditional mean. An asymmetric loss function implies a more complicated forecast that depends on the distribution of the forecast error as well as the loss function itself [67].

Consider a stochastic process $Z_t \equiv (Y_t, X_t')'$ where Y_t is the variable of interest and X_t is a vector of other variables. Suppose there are $T + 1$ ($\equiv R + P$) observations. We use the observations available at time t, $R \leq t < T + 1$, to generate P forecasts using each model. For each time t in the prediction period, we use either a rolling sample $\{Z_{t-R+1}, \ldots, Z_t\}$ of size R or the whole past sample $\{Z_1, \ldots, Z_t\}$ to estimate model parameters $\hat{\beta}_t$. We can then generate a sequence of one-step-ahead forecasts $\{f(Z_t, \hat{\beta}_t)\}_{t=R}^T$.

Suppose that there is a decision maker who takes an one-step point forecast $f_{t,1} \equiv f(Z_t, \hat{\beta}_t)$ of Y_{t+1} and uses it in some relevant decision. The one-step forecast error $e_{t+1} \equiv Y_{t+1} - f_{t,1}$ will result in a cost of $c(e_{t+1})$, where the function $c(e)$ will increase as e increases in size, but not necessarily symmetrically or continuously. The optimal forecast $f_{t,1}^*$ will be chosen to produce the forecast er-

rors that minimize the expected loss

$$\min_{f_{t,1}} \int_{-\infty}^{\infty} c(y - f_{t,1}) \mathrm{d}F_t(y) \,,$$

where $F_t(y) \equiv \Pr(Y_{t+1} \leq y | I_t)$ is the conditional distribution function, with I_t being some proper information set at time t that includes Z_{t-j}, $j \geq 0$. The corresponding optimal forecast error will be

$$e_{t+1}^* = Y_{t+1} - f_{t,1}^* \,.$$

Then the optimal forecast would satisfy

$$\frac{\partial}{\partial f_{t,1}} \int_{-\infty}^{\infty} c(y - f_{t,1}^*) \mathrm{d}F_t(y) = 0 \,.$$

When we interchange the operations of differentiation and integration,

$$\int_{-\infty}^{\infty} \frac{\partial}{\partial f_{t,1}} c(y - f_{t,1}^*) \mathrm{d}F_t(y) \equiv \mathbb{E}\left(\frac{\partial}{\partial f_{t,1}} c(Y_{t+1} - f_{t,1}^*) | I_t \right)$$

Based on the "generalized forecast error", $g_{t+1} \equiv \frac{\partial}{\partial f_{t,1}} c(Y_{t+1} - f_{t,1}^*)$, the condition for forecast optimality is:

$$H_0 : \mathbb{E}\left(g_{t+1} | I_t \right) = 0 \quad a.s. \,,$$

that is a martingale difference (MD) property of the generalized forecast error. This forms the optimality condition of the forecasts and gives an appropriate regression function corresponding to the specified loss function $c(\cdot)$.

To see this we consider the following two examples. First, when the loss function is the squared error loss

$$c(Y_{t+1} - f_{t,1}) = (Y_{t+1} - f_{t,1})^2 \,,$$

the generalized forecast error will be $g_{t+1} \equiv \frac{\partial}{\partial f_t} c(Y_{t+1} - f_{t,1}^*) = -2e_{t+1}^*$ and thus $\mathbb{E}\left(e_{t+1}^* | I_t \right) = 0$ $a.s.$, which implies that the optimal forecast

$$f_{t,1}^* = \mathbb{E}\left(Y_{t+1} | I_t \right)$$

is the conditional mean. Next, when the loss is the check function, $c(e) = \left[\alpha - \mathbf{1}(e < 0) \right] \cdot e \equiv \rho_\alpha(e_{t+1})$, the optimal forecast $f_{t,1}$, for given $\alpha \in (0, 1)$, minimizing

$$\min_{f_{t,1}} \mathbb{E}\left[c(Y_{t+1} - f_{t,1}) | I_t \right]$$

can be shown to satisfy

$$\mathbb{E}\left[\alpha - \mathbf{1}(Y_{t+1} < f_{t,1}^*) | I_t \right] = 0 \quad a.s.$$

Hence, $g_{t+1} \equiv \alpha - \mathbf{1}(Y_{t+1} < f_{t,1}^*)$ is the generalized forecast error. Therefore,

$$\alpha = \mathbb{E}\left[\mathbf{1}(Y_{t+1} < f_{t,1}^*) | I_t \right] = \Pr(Y_{t+1} \leq f_{t,1}^* | I_t) \,,$$

and the optimal forecast $f_{t,1}^* = q^\alpha \left(Y_{t+1} | I_t \right) \equiv q_t^\alpha$ is the conditional α-quantile.

Forecasting conditional quantiles are of paramount importance for risk management, which nowdays is key activity in financial institutions due to the increasing financial fragility in emerging markets and the extensive use of derivative products over the last decade. A risk measurement methodology called Value-at-Risk (VaR) has received a great attention from both regulatory and academic fronts. During a short span of time, numerous papers have studied various aspects of the VaR methodology. Bao, Lee, and Saltoglu [8] examine the relative out-of-sample predictive performance of various VaR models.

An interesting VaR model is the CaViaR (conditional autoregressive Value-at-Risk) model suggested by Engle and Manganelli [47]. They estimate the VaR from a quantile regression rather than inverting a conditional distribution. The idea is similar to the GARCH modeling in that VaR is modeled autoregressively

$$q_t(\alpha) = a_0 + a_1 q_{t-1}(\alpha) + h(x_t | \theta) \,,$$

where $x_t \in \mathcal{F}_{t-1}$, θ is a parameter vector, and $h(\cdot)$ is a function to explain the VaR model. Depending on the specification of $h(\cdot)$, the CaViaR model may be

$$q_t(\alpha) = a_0 + a_1 q_{t-1}(\alpha) + a_2 |r_{t-1}| \,,$$

$$q_t(\alpha) = a_0 + a_1 q_{t-1}(\alpha) + a_2 |r_{t-1}| + a_3 |r_{t-1}| \cdot \mathbf{1}(r_{t-1} < 0),$$

where the second model allow nonlinearity (asymmetry) similarly to the asymmetric GARCH models.

Bao, Lee, and Saltoglu [8] compare various VaR models. Their results show that the CaViaR quantile regression models of Engle and Manganelli [47] have shown some success in predicting the VaR risk measure for various periods of time, and it is generally more stable than the models that invert a distribution function.

Forecasting Directions

It is well known that, while financial returns $\{Y_t\}$ may not be predictable, their variance, sign, and quantiles may be predictable. Christofferson and Diebold [27] show that binary variable $G_{t+1} \equiv \mathbf{1}(Y_{t+1} > 0)$, where $\mathbf{1}(\cdot)$ takes the value of 1 if the statement in the parenthesis is true, and 0 otherwise, is predictable when some conditional moments are time varying, Hong and Lee [86], Hong and

Chung [85], Linton and Whang [104], Lee and Yang [100] among many others find some evidence that the directions of stock returns and foreign exchange rate changes are predictable.

Lee and Yang [100] also show that forecasting quantiles and forecasting binary (directional) forecasts are related, in that the former may lead to the latter. As noted by Powell [126], using the fact that for any monotonic function $h(\cdot)$, $q_t^\alpha(h(Y_{t+1})|\mathbf{X}_t) = h(q_t^\alpha(Y_{t+1}|\mathbf{X}_t))$, which follows immediately from observing that $\Pr(Y_{t+1} < y|\mathbf{X}_t) = \Pr[h(Y_{t+1}) < h(y)|\mathbf{X}_t]$, and noting that the indicator function is monotonic, $q_t^\alpha(G_{t+1}|\mathbf{X}_t) = q_t^\alpha(\mathbf{1}(Y_{t+1} > 0)|\mathbf{X}_t) = \mathbf{1}(q_t^\alpha(Y_{t+1}|\mathbf{X}_t) > 0)$. Therefore, predictability of conditional quantiles of financial returns may imply predictability of conditional direction.

Probability Forecasts

Diebold and Rudebush [38] consider the probability forecasts for the turning points of the business cycle. They measure the accuracy of predicted probabilities, that is the average distance between the predicted probabilities and observed realization (as measured by a zero-one dummy variable). Suppose there are $T + 1 \, (\equiv R + P)$ observations. We use the observations available at time $t \, (R \le t < T + 1)$, to estimate a model. We then have time series of $P = T - R + 1$ probability forecasts $\{p_{t+1}\}_{t=R}^T$ where p_t is the predicted probability of the occurrence of an event (e. g., business cycle turning point) in the next period $t + 1$. Let $\{d_{t+1}\}_{t=R}^T$ be the corresponding realization with $d_t = 1$ if a business cycle turning point (or any defined event) occurs in period t and $d_t = 0$ otherwise. The loss function analogous to the squared error is the Brier's score based on quadratic probability score (QPS):

$$QPS = P^{-1} \sum_{t=R}^T 2(p_t - d_t)^2 \, .$$

The QPS ranges from 0 to 2, with 0 for perfect accuracy. As noted by Diebold and Rudebush [38], the use of the symmetric loss function may not be appropriate as a forecaster may be penalized more heavily for missing a call (making a type II error) than for signaling a false alarm (making a type I error). Another loss function is given by the log probability score (LPS)

$$LPS = -P^{-1} \sum_{t=R}^T \ln \left(p_t^{d_t}(1 - p_t)^{(1-d_t)} \right) \, ,$$

which is similar to the loss for the interval forecast. A large mistake is penalized more heavily under LPS than under QPS. More loss functions are discussed in Diebold and Rudebush [38].

Another loss function useful in this context is the Kuipers score (KS), which is defined by

$$KS = \text{Hit Rate} - \text{False Alarm Rate} \, ,$$

where Hit Rate is the fraction of the bad events that were correctly predicted as good events (power, or $1-$ probability of type II error), and False Alarm Rate is the fraction of good events that had been incorrectly predicted as bad events (probability of type I error).

Forecasting Interval

Suppose Y_t is a stationary series. Let the one-period ahead conditional interval forecast made at time t from a model be denoted as

$$J_{t,1}(\alpha) = (L_{t,1}(\alpha), U_{t,1}(\alpha)), \quad t = R, \dots, T \, ,$$

where $L_{t,1}(\alpha)$ and $U_{t,1}(\alpha)$ are the lower and upper limits of the ex ante interval forecast for time $t + 1$ made at time t with the coverage probability α. Define the indicator variable $X_{t+1}(\alpha) = \mathbf{1}[Y_{t+1} \in J_{t,1}(\alpha)]$. The sequence $\{X_{t+1}(\alpha)\}_{t=R}^T$ is i.i.d. Bernoulli (α). The optimal interval forecast would satisfy $\mathbb{E}(X_{t+1}(\alpha)|I_t) = \alpha$, so that $\{X_{t+1}(\alpha) - \alpha\}$ will be an MD. A better model has a larger expected Bernoulli log-likelihood

$$\mathbb{E}\alpha^{X_{t+1}(\alpha)}(1 - \alpha)^{[1-X_{t+1}(\alpha)]} \, .$$

Hence, we can choose a model for interval forecasts with the largest out-of-sample mean of the predictive log-likelihood, which is defined by

$$P^{-1} \sum_{t=R}^T \ln \left(\alpha^{x_{t+1}(\alpha)}(1 - \alpha)^{[1-x_{t+1}(\alpha)]} \right) \, .$$

Evaluation of Nonlinear Forecasts

In order to evaluate the possible superior predictive ability of nonlinear models, we need to compare competing models in terms of a certain loss function. The literature has recently been exploding on this issue. Examples are Granger and Newbold [69], Diebold and Mariano [37], West [156], White [160], Hansen [81], Romano and Wolf [130], Giacomini and White [59], etc. In different perspective, to test the optimality of a given model, Patton and Timmermann [123] examine various testable properties that should hold for an optimal forecast.

Loss Functions

The loss function (or cost function) is a crucial ingredient for the evaluation of nonlinear forecasts. When a forecast $f_{t,h}$ of a variable Y_{t+h} is made at time t for h periods ahead, the loss (or cost) will arise if a forecast turns out to be different from the actual value. The loss function of the forecast error $e_{t+h} = Y_{t+h} - f_{t,h}$ is denoted as $c(Y_{t+h}, f_{t,h})$. The loss function can depend on the time of prediction and so it can be $c_{t+h}(Y_{t+h}, f_{t,h})$. If the loss function is not changing with time and does not depend on the value of the variable Y_{t+h}, the loss can be written simply as a function of the error only, $c_{t+h}(Y_{t+h}, f_{t,h}) = c(e_{t+h})$.

Granger [67] discusses the following required properties for a loss function: (i) $c(0) = 0$ (no error and no loss), (ii) $\min_e c(e) = 0$, so that $c(e) \geq 0$, and (iii) $c(e)$ is monotonically nondecreasing as e moves away from zero so that $c(e_1) \geq c(e_2)$ if $e_1 > e_2 > 0$ and if $e_1 < e_2 < 0$.

When $c_1(e), c_2(e)$ are both loss functions, Granger [67] shows that further examples of loss functions can be generated: $c(e) = ac_1(e) + bc_2(e), a \geq 0, b \geq 0$ will be a loss function. $c(e) = c_1(e)^a c_2(e)^b, a > 0, b > 0$ will be a loss function. $c(e) = 1(e > 0)c_1(e) + 1(e < 0)c_2(e)$ will be a loss function. If $h(\cdot)$ is a positive monotonic nondecreasing function with $h(0)$ finite, then $c(e) = h(c_1(e)) - h(0)$ is a loss function.

Granger [68] notes that an expected loss (a risk measure) of financial return Y_{t+1} that has a conditional predictive distribution $F_t(y) \equiv \Pr(Y_{t+1} \leq y|I_t)$ with $\mathbf{X}_t \in I_t$ may be written as

$$\mathbb{E}c(e) = A_1 \int_0^\infty |y-f|^p dF_t(y) + A_2 \int_{-\infty}^0 |y-f|^p \mathrm{d}F_t(y),$$

with A_1, A_2 both > 0 and some $\theta > 0$. Considering the symmetric case $A_1 = A_2$, one has a class of volatility measures $V_p = \mathbb{E}\left[|y - f|^p\right]$, which includes the variance with $p = 2$, and mean absolute deviation with $p = 1$.

Ding, Granger, and Engle [39] study the time series and distributional properties of these measures empirically and show that the absolute deviations are found to have some particular properties such as the longest memory. Granger remarks that given that the financial returns are known to come from a long tail distribution, $p = 1$ may be more preferable.

Another problem raised by Granger is how to choose optimal L_p-norm in empirical works, to minimize $\mathbb{E}[|\varepsilon_t|^p]$ for some p to estimate the regression model $Y_t = \mathbb{E}(Y_t|X_t; \beta) + \varepsilon_t$. As the asymptotic covariance matrix of $\hat{\beta}$ depends on p, the most appropriate value of p can be chosen to minimize the covariance matrix. In particular, Granger [68] refers to a trio of papers [84,116,117] who

find that the optimal $p = 1$ from Laplace and Cauchy distribution, $p = 2$ for Gaussian and $p = \infty$ (min/max estimator) for a rectangular distribution. Granger [68] also notes that in terms of the kurtosis κ, Harter [84] suggests to use $p = 1$ for $\kappa > 3.8$; $p = 2$ for $2.2 \leq \kappa \leq 3.8$; and $p = 3$ for $\kappa < 2.2$. In finance, the kurtosis of returns can be thought of as being well over 4 and so $p = 1$ is preferred.

Forecast Optimality

Optimal forecast of a time series model extensively depends on the specification of the loss function. Symmetric quadratic loss function is the most prevalent in applications due to its simplicity. The optimal forecast under quadratic loss is simply the conditional mean, but an asymmetric loss function implies a more complicated forecast that depends on the distribution of the forecast error as well as the loss function itself [67], as the expected loss function if formulated with the expectation taken with respect to the conditional distribution. Specification of the loss function defines the model under consideration.

Consider a stochastic process $Z_t \equiv (Y_t, X_t')'$ where Y_t is the variable of interest and X_t is a vector of other variables. Suppose there are $T + 1 (\equiv R + P)$ observations. We use the observations available at time t, $R \leq t < T + 1$, to generate P forecasts using each model. For each time t in the prediction period, we use either a rolling sample $\{Z_{t-R+1}, \ldots, Z_t\}$ of size R or the whole past sample $\{Z_1, \ldots, Z_t\}$ to estimate model parameters $\hat{\beta}_t$. We can then generate a sequence of one-step-ahead forecasts $\{f(Z_t, \hat{\beta}_t)\}_{t=R}^T$.

Suppose that there is a decision maker who takes an one-step point forecast $f_{t,1} \equiv f(Z_t, \hat{\beta}_t)$ of Y_{t+1} and uses it in some relevant decision. The one-step forecast error $e_{t+1} \equiv Y_{t+1} - f_{t,1}$ will result in a cost of $c(e_{t+1})$, where the function $c(e)$ will increase as e increases in size, but not necessarily symmetrically or continuously. The optimal forecast $f_{t,1}^*$ will be chosen to produce the forecast errors that minimize the expected loss

$$\min_{f_{t,1}} \int_{-\infty}^\infty c(y - f_{t,1})\mathrm{d}F_t(y),$$

where $F_t(y) \equiv \Pr(Y_{t+1} \leq y|I_t)$ is the conditional distribution function, with I_t being some proper information set at time t that includes $Z_{t-j}, j \geq 0$. The corresponding optimal forecast error will be

$$e_{t+1}^* = Y_{t+1} - f_{t,1}^*.$$

Then the optimal forecast would satisfy

$$\frac{\partial}{\partial f_{t,1}} \int_{-\infty}^{\infty} c(y - f_{t,1}^*) \mathrm{d}F_t(y) = 0 \, .$$

When we may interchange the operations of differentiation and integration,

$$\int_{-\infty}^{\infty} \frac{\partial}{\partial f_{t,1}} c(y - f_{t,1}^*) \mathrm{d}F_t(y) \equiv \mathbb{E}\left(\frac{\partial}{\partial f_{t,1}} c(Y_{t+1} - f_{t,1}^*) | I_t\right)$$

the "generalized forecast error", $g_{t+1} \equiv \frac{\partial}{\partial f_{t,1}} c(Y_{t+1} - f_{t,1}^*)$, forms the condition of forecast optimality:

$$H_0 : \mathbb{E}\left(g_{t+1} | I_t\right) = 0 \quad a.s.,$$

that is a martingale difference (MD) property of the generalized forecast error. This forms the optimality condition of the forecasts and gives an appropriate regression function corresponding to the specified loss function $c(\cdot)$.

Forecast Evaluation of Nonlinear Transformations

Granger [67] note that it is implausible to use the same loss function for forecasting Y_{t+h} and for forecasting $h_{t+1} = h(Y_{t+h})$ where $h(\cdot)$ is some function, such as the log or the square, if one is interested in forecasting volatility. Suppose the loss functions $c_1(\cdot), c_2(\cdot)$ are used for forecasting Y_{t+h} and for forecasting $h(Y_{t+h})$, respectively. Let $e_{t+1} \equiv Y_{t+1} - f_{t,1}$ will result in a cost of $c_1(e_{t+1})$, for which the optimal forecast $f_{t,1}^*$ will be chosen from $\min_{f_{t,1}} \int_{-\infty}^{\infty} c_1(y - f_{t,1}) \mathrm{d}F_t(y)$, where $F_t(y) \equiv \Pr(Y_{t+1} \leq y | I_t)$. Let $\varepsilon_{t+1} \equiv h_{t+1} - h_{t,1}$ will result in a cost of $c_2(\varepsilon_{t+1})$, for which the optimal forecast $h_{t,1}^*$ will be chosen from $\min_{h_{t,1}} \int_{-\infty}^{\infty} c_2(h - h_{t,1}) \mathrm{d}H_t(h)$, where $H_t(h) \equiv \Pr(h_{t+1} \leq h | I_t)$. Then the optimal forecasts for Y and h would respectively satisfy

$$\int_{-\infty}^{\infty} \frac{\partial}{\partial f_{t,1}} c_1(y - f_{t,1}^*) \mathrm{d}F_t(y) = 0 \, ,$$
$$\int_{-\infty}^{\infty} \frac{\partial}{\partial h_{t,1}} c_2(h - h_{t,1}^*) \mathrm{d}H_t(h) = 0 \, .$$

It is easy to see that the optimality condition for $f_{t,1}^*$ does not imply the optimality condition for $h_{t,1}^*$ in general. Under some strong conditions on the functional forms of the transformation $h(\cdot)$ and of the two loss functions $c_1(\cdot), c_2(\cdot)$, the above two conditions may coincide. Granger [67] remarks that it would be strange behavior to use the same loss function for Y and $h(Y)$. We leave this for further analysis in a future research.

Density Forecast Evaluation

Most of the classical finance theories, such as asset pricing, portfolio selection and option valuation, aim to model the surrounding uncertainty via a parametric distribution function. For example, extracting information about market participants' expectations from option prices can be considered another form of density forecasting exercise [92]. Moreover, there has also been increasing interest in evaluating forecasting models of inflation, unemployment and output in terms of density forecasts [29]. While evaluating each density forecast model has become versatile since Diebold et al. [35], there has been much less effort in comparing alternative density forecast models.

Given the recent empirical evidence on volatility clustering and asymmetry and fat-tailedness in financial return series, relative adequacy of a given model among alternative models would be useful measure of evaluating forecast models. Deciding on which distribution and/or volatility specification to use for a particular asset is a common task even for finance practitioners. For example, despite the existence of many volatility specifications, a consensus on which model is most appropriate has yet to be reached. As argued in Poon and Granger [125], most of the (volatility) forecasting studies do not produce very conclusive results because only a subset of alternative models are compared, with a potential bias towards the method developed by the authors. Poon and Granger [125] argue that lack of a uniform forecast evaluation technique makes volatility forecasting a difficult task. They wrote (p. 507), " … it seems clear that one form of study that is included is conducted just to support a viewpoint that a particular method is useful. It might not have been submitted for publication if the required result had not been reached. This is one of the obvious weaknesses of a comparison such as this; the papers being prepared for different reasons, use different data sets, many kinds of assets, various intervals between readings, and a variety of evaluation techniques".

Following Diebold et al. [35], it has become common practice to evaluate the adequacy of a forecast model based on the probability integral transform (PIT) of the process with respect to the model's density forecast. If the density forecast model is correctly specified, the PIT follows an i.i.d. uniform distribution on the unit interval and, equivalently, its inverse normal transform follows an i.i.d. normal distribution. We can therefore evaluate a density forecast model by examining the departure of the transformed PIT from this property (i.i.d. and normality). The departure can be quantified by the Kullback-Leibler [97] information criterion, or KLIC, which is the expected logarithmic

value of the likelihood ratio (LR) of the transformed PIT and the i.i.d. normal variate. Thus the LR statistic measures the distance of a candidate model to the unknown true model.

Consider a financial return series $\{y_t\}_{t=1}^T$. This observed data on a univariate series is a realization of a stochastic process $\mathbf{Y}^T \equiv \{Y_\tau : \Omega \to \mathbb{R}, \tau = 1, 2, \ldots, T\}$ on a complete probability space $(\Omega, \mathcal{F}_T, P_0^T)$, where $\Omega = \mathbb{R}^T \equiv \times_{\tau=1}^T \mathbb{R}$ and $\mathcal{F}_T = \mathcal{B}(\mathbb{R}^T)$ is the Borel σ-field generated by the open sets of $\mathbb{R}^T$, and the *joint* probability measure $P_0^T(B) \equiv P_0[\mathbf{Y}^T \in B]$, $B \in \mathcal{B}(\mathbb{R}^T)$ completely describes the stochastic process. A sample of size T is denoted as $\mathbf{y}^T \equiv (y_1, \ldots, y_T)'$.

Let σ-finite measure ν^T on $\mathcal{B}(\mathbb{R}^T)$ be given. Assume $P_0^T(B)$ is absolutely continuous with respect to ν^T for all $T = 1, 2, \ldots$, so that there exists a measurable Radon–Nikodým density $g^T(\mathbf{y}^T) = \mathrm{d}P_0^T/\mathrm{d}\nu^T$, unique up to a set of zero measure-ν^T.

Following White [159], we define a probability model $\mathcal{P}$ as a collection of distinct probability measures on the measurable space $(\Omega, \mathcal{F}_T)$. A probability model $\mathcal{P}$ is said to be correctly specified for $\mathbf{Y}^T$ if $\mathcal{P}$ contains P_0^T. Our goal is to evaluate and compare a set of parametric probability models $\{P_\theta^T\}$, where $P_\theta^T(B) \equiv P_\theta[\mathbf{Y}^T \in B]$. Suppose there exists a measurable Radon–Nikodým density $f^T(\mathbf{y}^T) = \mathrm{d}P_\theta^T/\mathrm{d}\nu^T$ for each $\theta \in \Theta$, where θ is a finite-dimensional vector of parameters and is assumed to be identified on Θ, a compact subset of $\mathbb{R}^k$. See Theorem 2.6 in White [159].

In the context of forecasting, instead of the joint density $g^T(\mathbf{y}^T)$, we consider forecasting the *conditional* density of $\mathbf{Y}^t$, given the information $\mathcal{F}_{t-1}$ generated by $\mathbf{Y}^{t-1}$. Let $\varphi_t(y_t) \equiv \varphi_t(y_t|\mathcal{F}_{t-1}) \equiv g^t(\mathbf{y}^t)/g^{t-1}(\mathbf{y}^{t-1})$ for $t = 2, 3, \ldots$ and $\varphi_1(y_1) \equiv \varphi_1(y_1|\mathcal{F}_0) \equiv g^1(\mathbf{y}^1) = g^1(y_1)$. Thus the goal is to forecast the (true, unknown) conditional density $\varphi_t(y_t)$.

For this, we use an one-step-ahead conditional density forecast model $\psi_t(y_t; \theta) \equiv \psi_t(y_t|\mathcal{F}_{t-1}; \theta) \equiv f^t(\mathbf{y}^t)/f^{t-1}(\mathbf{y}^{t-1})$ for $t = 2, 3, \ldots$ and $\psi_1(y_1) \equiv \psi_1(y_1|\mathcal{F}_0) \equiv f^1(\mathbf{y}^1) = f^1(y_1)$. If $\psi_t(y_t; \theta_0) = \varphi_t(y_t)$ almost surely for some $\theta_0 \in \Theta$, then the one-step-ahead density forecast is correctly specified, and it is said to be optimal because it dominates all other density forecasts for any loss functions as discussed in the previous section (see [35,67,70]).

In practice, it is rarely the case that we can find an optimal model. As it is very likely that "the true distribution is in fact too complicated to be represented by a simple mathematical function" [133], all the models proposed by different researchers can be possibly misspecified and thereby we regard each model as an approximation to the truth. Our task is then to investigate which density forecast model can approximate the true conditional density most closely. We have to first define a metric to measure the distance of a given model to the truth, and then compare different models in terms of this distance.

The adequacy of a density forecast model can be measured by the conditional Kullback-Leibler [97] Information Criterion (KLIC) divergence measure between two conditional densities,

$$\mathbb{I}_t(\varphi : \psi, \theta) = \mathbb{E}_{\varphi_t}[\ln \varphi_t(y_t) - \ln \psi_t(y_t; \theta)],$$

where the expectation is with respect to the true conditional density $\varphi_t(\cdot|\mathcal{F}_{t-1})$, $\mathbb{E}_{\varphi_t} \ln \varphi_t(y_t|\mathcal{F}_{t-1}) < \infty$, and $\mathbb{E}_{\varphi_t} \ln \psi_t(y_t|\mathcal{F}_{t-1}; \theta) < \infty$. Following White [159], we define the distance between a density model and the true density as the minimum of the KLIC

$$\mathbb{I}_t(\varphi : \psi, \theta_{t-1}^*) = \mathbb{E}_{\varphi_t}\left[\ln \varphi_t(y_t) - \ln \psi_t(y_t; \theta_{t-1}^*)\right],$$

where $\theta_{t-1}^* = \arg\min \mathbb{I}_t(\varphi : \psi, \theta)$ is the pseudo-true value of θ [133]. We assume that θ_{t-1}^* is an interior point of Θ. The smaller this distance is, the closer the density forecast $\psi_t(\cdot|\mathcal{F}_{t-1}; \theta_{t-1}^*)$ is to the true density $\varphi_t(\cdot|\mathcal{F}_{t-1})$.

However, $\mathbb{I}_t(\varphi : \psi, \theta_{t-1}^*)$ is unknown since θ_{t-1}^* is not observable. We need to estimate θ_{t-1}^*. If our purpose is to compare the out-of-sample predictive abilities among competing density forecast models, we split the data into two parts, one for estimation and the other for out-of-sample validation. At each period t in the out-of-sample period ($t = R + 1, \ldots, T$), we estimate the unknown parameter vector θ_{t-1}^* and denote the estimate as $\hat{\theta}_{t-1}$. Using $\{\hat{\theta}_{t-1}\}_{t=R+1}^T$, we can obtain the out-of-sample estimate of $\mathbb{I}_t(\varphi : \psi, \theta_{t-1}^*)$ by

$$\mathbb{I}_P(\varphi : \psi) \equiv \frac{1}{P} \sum_{t=R+1}^T \ln[\varphi_t(y_t)/\psi_t(y_t; \hat{\theta}_{t-1})]$$

where $P = T - R$ is the size of the out-of-sample period. Note that

$$\mathbb{I}_P(\varphi : \psi) = \frac{1}{P} \sum_{t=R+1}^T \ln\left[\varphi_t(y_t)/\psi_t(y_t; \theta_{t-1}^*)\right]$$
$$+ \frac{1}{P} \sum_{t=R+1}^T \ln[\psi_t(y_t; \theta_{t-1}^*)/\psi_t(y_t; \hat{\theta}_{t-1})],$$

where the first term in $\mathbb{I}_P(\varphi : \psi)$ measures model uncertainty (the distance between the optimal density $\varphi_t(y_t)$ and the model $\psi_t(y_t; \theta_{t-1}^*)$) and the second term mea-

sures parameter estimation uncertainty due to the distance between $\boldsymbol{\theta}^{*}_{t-1}$ and $\hat{\boldsymbol{\theta}}_{t-1}$.

Since the KLIC measure takes on a smaller value when a model is closer to the truth, we can regard it as a loss function and use $\mathbb{I}_P(\varphi : \psi)$ to formulate the loss-differential. The out-of-sample average of the loss-differential between model 1 and model 2 is

$$\mathbb{I}_P(\varphi : \psi^1) - \mathbb{I}_P(\varphi : \psi^2)$$
$$= \frac{1}{P} \sum_{t=R+1}^{T} \ln \left[\psi_t^2 \left(y_t; \hat{\boldsymbol{\theta}}_{t-1}^2 \right) / \psi_t^1 \left(y_t; \hat{\boldsymbol{\theta}}_{t-1}^1 \right) \right] \,,$$

which is the ratio of the two predictive log-likelihood functions. With treating model 1 as a benchmark model (for model selection) or as the model under the null hypothesis (for hypothesis testing), $\mathbb{I}_P(\varphi : \psi^1) - \mathbb{I}_P(\varphi : \psi^2)$ can be considered as a loss function to minimize. To sum up, the KLIC differential can serve as a *loss* function for density forecast evaluation as discussed in Bao, Lee, and Saltoglu [10]. See Corradi and Swanson [31] for the related ideas using different loss functions.

Using the KLIC divergence measure to characterize the extent of misspecification of a forecast model, Bao, Lee, and Saltoglu [10], in an empirical study with the S&P500 and NASDAQ daily return series, find strong evidence for rejecting the Normal-GARCH benchmark model, in favor of the models that can capture skewness in the conditional distribution and asymmetry and long-memory in the conditional variance. Also, Bao and Lee [8] investigate the nonlinear predictability of stock returns when the density forecasts are evaluated/compared instead of the conditional mean point forecasts. The conditional mean models they use for the daily closing S&P500 index returns include the martingale difference model, the linear ARMA models, the STAR and SETAR models, the ANN model, and the polynomial model. Their empirical findings suggest that the out-of-sample predictive abilities of nonlinear models for stock returns are asymmetric in the sense that the right tails of the return series are predictable via many of the nonlinear models while we find no such evidence for the left tails or the entire distribution.

Conclusions

In this article we have selectively reviewed the state-of-the-art in nonlinear time series models that are useful in forecasting financial variables. Overall financial returns are difficult to forecast, and this may just be a reflection of the efficiency of the markets on processing information. The success of nonlinear time series on producing better fore-

casts than linear models depends on how persistent the nonlinearities are in the data. We should note that though many of the methodological developments are concerned with the specification of the conditional mean and conditional variance, there is an active area of research investigating other aspects of the conditional density – quantiles, directions, intervals – that seem to be promising from a forecasting point of view.

For a more extensive coverage to complement this review, the readers may find the following additional references useful. Campbell, Lo, and MacKinlay [22], Chapter 12, provides a brief but excellent summary of nonlinear time series models for the conditional mean and conditional variance as well and various methods such as ANN and nonparametric methods. Similarly, the interested readers may also refer to the books and monographs of Granger and Teräsvirta [72], Franses and van Dijk [55], Fan and Yao [52], Tsay [153], Gao [57], and some book chapters such as Stock [139], Tsay [152], Teräsvirta [145], and White [161].

Future Directions

Methodological developments in nonlinear time series have happened without much guidance from economic theory. Nonlinear models are for most part ad hoc specifications that, from a forecasting point of view, are validated according to some statistical loss function. Though we have surveyed some articles that employ some economic rationale to evaluate the model and/or the forecast – bull/bear cycles, utility function, profit/loss function –, there is still a vacuum on understanding why, how, and when nonlinearities may show up in the data.

From a methodological point of view, future developments will focus on multivariate nonlinear time series models and their associated statistical inference. Nonlinear VAR-type models for the conditional mean and high-dimensional multivariate volatility models are still in their infancy. Dynamic specification testing in a multivariate setting is paramount to the construction of a multivariate forecast and though multivariate predictive densities are inherently difficult to evaluate, they are most important in financial economics.

Another area of future research will deal with the econometrics of a data-rich environment. The advent of large databases begs the introduction of new techniques and methodologies that permits the reduction of the many dimensions of a data set to a parsimonious but highly informative set of variables. In this sense, criteria on how to combine information and how to combine models to produce more accurate forecasts are highly desirable.

Finally, there are some incipient developments on defining new stochastic processes where the random variables that form the process are of a symbolic nature, i. e. intervals, boxplots, histograms, etc. Though the mathematics of these processes are rather complex, future developments in this area will bring exciting results for the area of forecasting.

Bibliography

1. Ait-Sahalia Y, Hansen LP (2009) Handbook of Financial Econometrics. Elsevier Science, Amsterdam
2. Andersen TG, Bollerslev T (1998) Answering the Skeptics: Yes, Standard Volatility Models Do Provide Accurate Forecasts. Int Econ Rev 39(4):885–905
3. Andersen TG, Bollerslev T, Diebold FX, Labys P (2001) The Distribution of Realized Exchange Rate Volatility. J Amer Stat Assoc 96:42–55
4. Andersen TG, Bollerslev T, Diebold FX, Labys P (2003) Modeling and Forecasting Realized Volatility. Econometrica 71: 579–625
5. Ang A, Bekaert G (2002) Regime Switccheds in Interest Rates. J Bus Econ Stat 20:163–182
6. Bachelier L (1900) Theory of Speculation. In: Cootner P (ed) The Random Character of Stock Market Prices. MIT Press, Cambridge. (1964) reprint
7. Bai J, Ng S (2007) Forecasting Economic Time Series Using Targeted Predictors. Working Paper, New York University and Columbia University
8. Bao Y, Lee TH (2006) Asymmetric Predictive Abilities of Non-linear Models for Stock Returns: Evidence from Density Forecast Comparison. Adv Econ 20 B:41–62
9. Bao Y, Lee TH, Saltoglu B (2006) Evaluating Predictive Performance of Value-at-Risk Models in Emerging Markets: A Reality Check. J Forecast 25(2):101–128
10. Bao Y, Lee TH, Saltoglu B (2007) Comparing Density Forecast Models. J Forecast 26(3):203–225
11. Barndorff-Nielsen OE, Shephard N (2002) Econometric Analysis of Realised Volatility and Its Use in Estimating Stochastic Volatility Models. J Royal Stat Soc B 64:853–223
12. Bauwens L, Laurent S, Rombouts JVK (2006) Multivariate GARCH Models: A Survey. J Appl Econ 21:79–109
13. Bera AK, Higgins ML (1993) ARCH Models: Properties, Estimation, and Testing. J Econ Surv 7:305–366
14. Bollerslev T (1986) Generalized Autoregressive Conditional Heteroskedasticity. J Econ 31:307–327
15. Bollerslev T, Chou RY, Kroner KF (1992) ARCH Models in Finance. J Econ 52:5–59
16. Bollerslev T, Engle RF, Nelson DB (1994) ARCH Models. In: Engle RF, McFadden DL (eds) Handbook of Econometrics, vol 4. Elsevier Science, Amsterdam
17. Bollerslev T, Engle RF, Wooldridge J (1988) A Capital Asset Pricing Model with Time Varying Covariances. J Political Econ 96:116–131
18. Boero G, Marrocu E (2004) The Performance of SETAR Models: A Regime Conditional Evaluation of Point, Interval, and Density Forecasts. Int J Forecast 20:305–320
19. Breiman L (1996) Bagging Predictors. Machine Learning 24:123–140
20. Breiman L (1996) Heuristics of Instability and Stabilization in Model Selection. Ann Stat 24(6):2350–2383
21. Brooks C, Persand G (2003) Volatility Forecasting for Risk Management. J Forecast 22(1):1–22
22. Campbell JY, Lo AW, MacKinlay AC (1997) The Econometrics of Financial Markets. Princeton University Press, New Jersey
23. Campbell JY, Thompson SB (2007) Predicting Excess Stock Returns Out of Sample: Can Anything Beat the Historical Average? Harvard Institute of Economic Research, Discussion Paper No. 2084
24. Cai Z, Fan J, Yao Q (2000) Functional-coefficient Regression Models for Nonlinear Time Series. J Amer Stat Assoc 95: 941–956
25. Chen X (2006) Large Sample Sieve Estimation of Semi-Nonparametric Models. In: Heckman JJ, Leamer EE (eds) Handbook of Econometrics, vol 6. Elsevier Science, Amsterdam, Chapter 76
26. Chen R, Tsay RS (1993) Functional-coefficient Autoregressive Models. J Amer Stat Assoc 88:298–308
27. Christofferson PF, Diebold FX (2006) Financial Asset Returns, Direction-of-Change Forecasting, and Volatility Dynamics. Manag Sci 52:1273–1287
28. Clements MP, Franses PH, Swanson NR (2004) Forecasting Economic and Financial Time-Series with Non-linear Models. Int J Forecast 20:169–183
29. Clements MP, Smith J (2000) Evaluating the forecast densities of linear and non-linear models: applications to output growth and unemployment. J Forecast 19:255–276
30. Cleveland WS (1979) Robust Locally Weighted Regression and Smoothing Scatter Plots. J Amer Stat Assoc 74: 829–836
31. Corradi V, Swanson NR (2006) Predictive Density Evaluation. In: Granger CWJ, Elliot G, Timmerman A (eds) Handbook of Economic Forecasting. Elsevier, Amsterdam, pp 197–284
32. Corsi F (2004) A Simple Long Memory Model of Realized Volatility. Working Paper, University of Lugano
33. Dahl CM, González-Rivera G (2003) Testing for Neglected Nonlinearity in Regression Models based on the Theory of Random Fields. J Econ 114:141–164
34. Dahl CM, González-Rivera G (2003) Identifying Nonlinear Components by Random Fields in the US GNP Growth. Implications for the Shape of the Business Cycle. Stud Nonlinear Dyn Econ 7(1):art2
35. Diebold FX, Gunther TA, Tay AS (1998) Evaluating Density Forecasts with Applications to Financial Risk Management. Int Econ Rev 39:863–883
36. Diebold FX, Li C (2006) Forecasting the Term Structure of Government Bond Yields. J Econom 130:337–364
37. Diebold FX, Mariano R (1995) Comparing predictive accuracy. J Bus Econ Stat 13:253–265
38. Diebold FX, Rudebusch GD (1989) Scoring the Leading Indicators. J Bus 62(3):369–391
39. Ding Z, Granger CWJ, Engle RF (1993) A Long Memory Property of Stock Market Returns and a New Model. J Empir Finance 1:83–106
40. Dueker M, Neely CJ (2007) Can Markov Switching Models Predict Excess Foreign Exchange Returns? J Bank Finance 31:279–296
41. Durland JM, McCurdy TH (1994) Duration-Dependent Transi-

tions in a Markov Model of US GNP Growth. J Bus Econ Stat 12:279–288

42. Efron B, Hastie T, Johnstone I, Tibshirani R (2004) Least Angle Regression. Ann Stat 32(2):407–499

43. Engel C, Hamilton JD (1990) Long Swings in the Dollar: Are they in the Data and Do Markets Know it? Amer Econ Rev 80(4):689–713

44. Engle RF (1982) Autoregressive Conditional Heteroscedasticity with Estimates of the Variance of UK Inflation. Econometrica 50:987–1008

45. Engle RF, Hendry DF, Richard J-F (1983) Exogeneity. Econometrica 51:277–304

46. Engle RF, Kane A, Noh J (1997) Index-Option Pricing with Stochastic Volatility and the Value of Accurate Variance Forecasts. Rev Deriv Res 1:139–157

47. Engle RF, Manganelli S (2004) CaViaR: Conditional autoregressive Value at Risk by regression quantiles. J Bus Econ Stat 22(4):367–381

48. Engle RF, Ng VK, Rothschild M (1990) Asset Pricing with a Factor ARCH Covariance Structure: Empirical Estimates for Treasury Bills. J Econ 45:213–238

49. Engle RF, Russell JR (1998) Autoregressive Conditional Duration: A New Model for Irregularly Spaced Transaction Data. Econometrica 66:1127–1162

50. Fan J, Gijbels I (1996) Local Polynomial Modelling and Its Applications. Chapman and Hall, London

51. Fan J, Yao Q (1998) Efficient estimation of conditional variance functions in stochastic regression. Biometrika 85: 645–660

52. Fan J, Yao Q (2003) Nonlinear Time Series. Springer, New York

53. Fan J, Yao Q, Cai Z (2003) Adaptive varying-coefficient linear models. J Royal Stat Soc B 65:57–80

54. Fitzenberger B (1997) The Moving Blocks Bootstrap and Robust Inference for Linear Least Squares and Quantile Regressions. J Econ 82:235–287

55. Franses PH, van Dijk D (2000) Nonlinear Time Series Models in Empirical Finance. Cambridge University Press, Cambridge

56. Gallant AR, Nychka DW (1987) Semi-nonparametric maximum likelihood estimation. Econometrica 55:363–390

57. Gao J (2007) Nonlinear Time Series: Semiparametric and Nonparametric Methods. Chapman and Hall, Boca Raton

58. Ghysels E, Santa-Clara P, Valkanov R (2006) Predicting Volatility: How to Get Most out of Returns Data Sampled at Different Frequencies. J Econ 131:59–95

59. Giacomini R, White H (2006) Tests of Conditional Predictive Ability. Econometrica 74:1545–1578

60. Glosten LR, Jaganathan R, Runkle D (1993) On the Relationship between the Expected Value and the Volatility of the Nominal Excess Return on Stocks. J Finance 48:1779–1801

61. González-Rivera G (1998) Smooth-Transition GARCH Models. Stud Nonlinear Dyn Econ 3(2):61–78

62. González-Rivera G, Lee TH, Mishra S (2004) Forecasting Volatility: A Reality Check Based on Option Pricing, Utility Function, Value-at-Risk, and Predictive Likelihood. Int J Forecast 20(4):629–645

63. González-Rivera G, Lee TH, Mishra S (2008) Jumps in Cross-Sectional Rank and Expected Returns: A Mixture Model. J Appl Econ; forthcoming

64. González-Rivera G, Lee TH, Yoldas E (2007) Optimality of the Riskmetrics VaR Model. Finance Res Lett 4:137–145

65. Gonzalo J, Martíneza O (2006) Large shocks vs. small shocks. (Or does size matter? May be so). J Econ 135:311–347

66. Goyal A, Welch I (2006) A Comprehensive Look at The Empirical Performance of Equity Premium Prediction. Working Paper, Emory and Brown, forthcoming in Rev Financ Stud

67. Granger CWJ (1999) Outline of Forecast Theory Using Generalized Cost Functions. Span Econ Rev 1:161–173

68. Granger CWJ (2002) Some Comments on Risk. J Appl Econ 17:447–456

69. Granger CWJ, Newbold P (1986) Forecasting Economic Time Series, 2nd edn. Academic Press, San Diego

70. Granger CWJ, Pesaran MH (2000) A Decision Theoretic Approach to Forecasting Evaluation. In: Chan WS, Li WK, Tong H (eds) Statistics and Finance: An Interface. Imperial College Press, London

71. Granger CWJ, Lee TH (1999) The Effect of Aggregation on Nonlinearity. Econ Rev 18(3):259–269

72. Granger CWJ, Teräsvirta T (1993) Modelling Nonlinear Economic Relationships. Oxford University Press, New York

73. Guidolin M, Timmermann A (2006) An Econometric Model of Nonlinear Dynamics in the Joint Distribution of Stock and Bond Returns. J Appl Econ 21:1–22

74. Haggan V, Ozaki T (1981) Modeling Nonlinear Vibrations Using an Amplitude-dependent Autoregressive Time Series Model. Biometrika 68:189–196

75. Hamilton JD (1994) Time Series Analysis. Princeton University Press, New Jersey

76. Hamilton JD (1989) A New Approach to the Economic Analysis of Nonstationary Time Series and the Business Cycle. Econometrica 57:357–384

77. Hamilton JD (1996) Specification Testing in Markov-Switching Time Series Models. J Econ 70:127–157

78. Hamilton JD (2001) A Parametric Approach to Flexible Nonlinear Inference. Econometrica 69:537–573

79. Hamilton JD, Jordà O (2002) A Model of the Federal Funds Target. J Political Econ 110:1135–1167

80. Hansen BE (1996) Inference when a Nuisance Parameter is not Identified under the Null Hypothesis. Econometrica 64: 413–430

81. Hansen PR (2005) A test for superior predictive ability. J Bus Econ Stat 23:365–380

82. Harding D, Pagan A (2002) Dissecting the Cycle: A Methodological Investigation. J Monet Econ 49:365–381

83. Härdle W, Tsybakov A (1997) Local polynomial estimators of the volatility function in nonparametric autoregression. J Econ 81:233–242

84. Harter HL (1977) Nonuniqueness of Least Absolute Values Regression. Commun Stat – Theor Methods A6:829–838

85. Hong Y, Chung J (2003) Are the Directions of Stock Price Changes Predictable? Statistical Theory and Evidence. Working Paper, Department of Economics, Cornell University

86. Hong Y, Lee TH (2003) Inference on Predictability of Foreign Exchange Rates via Generalized Spectrum and Nonlinear Time Series Models. Rev Econ Stat 85(4):1048–1062

87. Hong Y, Lee TH (2003b) Diagnostic Checking for Adequacy of Nonlinear Time Series Models. Econ Theor 19(6):1065–1121

88. Hornik K, Stinchcombe M, White H (1989) Multi-Layer Feedforward Networks Are Universal Approximators. Neural Netw 2:359–366

89. Huang H, Tae-Hwy L, Canlin L (2007) Forecasting Output Growth and Inflation: How to Use Information in the Yield

Curve. Working Paper, University of California, Riverside, Department of Economics

90. Huang YL, Kuan CM (2007) Re-examining Long-Run PPP under an Innovation Regime Switching Framework. Academia Sinica, Taipei

91. Inoue A, Kilian L (2008) How Useful is Bagging in Forecasting Economic Time Series? A Case Study of US CPI Inflation, forthcoming. J Amer Stat Assoc 103(482):511–522

92. Jackwerth JC, Rubinstein M (1996) Recovering probability distributions from option prices. J Finance 51:1611–1631

93. Judd KL (1998) Numerical Methods in Economics. MIT Press, Cambridge

94. Kanas A (2003) Non-linear Forecasts of Stock Returns. J Forecast 22:299–315

95. Koenker R, Bassett Jr G (1978) Regression Quantiles. Econometrica 46(1):33–50

96. Kuan CM, Huang YL, Tsayn RS (2005) An unobserved component model with switching permanent and transitory innovations. J Bus Econ Stat 23:443–454

97. Kullback L, Leibler RA (1951) On Information and Sufficiency. Ann Math Stat 22:79–86

98. Lee TH, Ullah A (2001) Nonparametric Bootstrap Tests for Neglected Nonlinearity in Time Series Regression Models. J Nonparametric Stat 13:425–451

99. Lee TH, White H, Granger CWJ (1993) Testing for Neglected Nonlinearity in Time Series Models: A Comparison of Neural Network Methods and Alternative Tests. J Econ 56:269–290

100. Lee TH, Yang Y (2006) Bagging Binary and Quantile Predictors for Time Series. J Econ 135:465–497

101. Lettau M, Ludvigson S (2001) Consumption, Aggregate Wealth, and Expected Stock Returns. J Finance 56:815–849

102. Lewellen J (2004) Predicting Returns with Financial Ratios. J Financial Econ 74:209–235

103. Lintner J (1965) Security Prices, Risk and Maximal Gains from Diversification. J Finance 20:587–615

104. Linton O, Whang YJ (2007) A Quantilogram Approach to Evaluating Directional Predictability. J Econom 141:250-282

105. Lo AW, MacKinlay AC (1999) A Non-Random Walk Down Wall Street. Princeton University Press, Princeton

106. Long X, Su L, Ullah A (2007) Estimation and Forecasting of Dynamic Conditional Covariance: A Semiparametric Multivariate Model. Working Paper, Department of Economics, UC Riverside

107. Lopez JA (2001) Evaluating the Predictive Accuracy of Volatility Models. J Forecast 20:87–109

108. Ludvigson S, Ng S (2007) The Empirical Risk Return Relation: A Factor Analysis Approach. J Financ Econ 83:171–222

109. Lundbergh S, Teräsvirta T (2002) Forecasting with smooth transition autoregressive models. In: Clements MP, Hendry DF (eds) A Companion to Economic Forecasting. Blackwell, Oxford, Chapter 21

110. Luukkonen R, Saikkonen P, Teräsvirta T (1988) Testing Linearity in Univariate Time Series Models. Scand J Stat 15:161–175

111. Maheu JM, McCurdy TH (2000) Identifying Bull and Bear Markets in Stock Returns. J Bus Econ Stat 18:100–112

112. Manski CF (1975) Maximum Score Estimation of the Stochastic Utility Model of Choice. J Econ 3(3):205–228

113. Markowitz H (1959) Portfolio Selection: Efficient Diversification of Investments. John Wiley, New York

114. Marsh IW (2000) High-frequency Markov Switching Models in the Foreign Exchange Market. J Forecast 19:123–134

115. McAleer M, Medeiros MC (2007) A multiple regime smooth transition heterogeneous autoregressive model for long memory and asymmetries. J Econ; forthcoming

116. Money AH, Affleck-Graves JF, Hart ML, Barr GDI (1982) The Linear Regression Model and the Choice of p. Commun Stat – Simul Comput 11(1):89–109

117. Nyguist H (1983) The Optimal L_p-norm Estimation in Linear Regression Models. Commun Stat – Theor Methods 12: 2511–2524

118. Nadaraya ÉA (1964) On Estimating Regression. Theor Probab Appl 9:141–142

119. Nelson CR, Siegel AF (1987) Parsimonious Modeling of Yield Curves. J Bus 60:473–489

120. Nelson DB (1991) Conditional Heteroscedasticity in Asset Returns: A New Approach. Econometrica 59(2):347–370

121. Noh J, Engle RF, Kane A (1994) Forecasting Volatility and Option Prices of the S&P 500 Index. J Deriv 17–30

122. Pagan AR, Ullah A (1999) Nonparametric Econometrics. Cambridge University Press, Cambridge

123. Patton AJ, Timmermann A (2007) Testing Forecast Optimality Under Unknown Loss. J Amer Stat Assoc 102(480): 1172–1184

124. Perez-Quiros G, Timmermann A (2001) Business Cycle Asymmetries in Stock Returns: Evidence form Higher Order Moments and Conditional Densities. J Econ 103:259–306

125. Poon S, Granger CWJ (2003) Forecasting volatility in financial markets. J Econ Lit 41:478–539

126. Powell JL (1986) Censored Regression Quantiles. J Econ 32:143–155

127. Priestley MB (1980) State-dependent Models: A General Approach to Nonlinear Time Series Analysis. J Time Ser Anal 1:47–71

128. Raj B, Ullah A (1981) Econometrics: A Varying Coefficients Approach. Croom Helm, London

129. Riskmetrics (1995) Technical Manual, 3rd edn. New York

130. Romano JP, Wolf M (2005) Stepwise multiple testing as formalized data snooping. Econometrica 73:1237–1282

131. Ross S (1976) The Arbitrage Theory of Capital Asset Pricing. J Econ Theor 13:341–360

132. Ruppert D, Wand MP (1994) Multivariate Locally Weighted Least Squares Regression. Ann Stat 22:1346–1370

133. Sawa T (1978) Information Criteria for Discriminating among Alternative Regression Models. Econometrica 46:1273–1291

134. Schwert GW (1990) Stock Volatility and the Crash of '87. Rev Financ Stud 3(1):77–102

135. Sentana E (1995) Quadratic ARCH models. Rev Econ Stud 62(4):639–661

136. Sichel DE (1994) Inventories and the Three Phases of the Business Cycle. J Bus Econ Stat 12:269–277

137. Sharpe W (1964) Capital Asset Prices: A Theory of Market Equilibrium under Conditions of Risk. J Finance 19:425–442

138. Stinchcombe M, White H (1998) Consistent Specification Testing with Nuisanse Parameters Present only under the Alternative. Econ Theor 14:295–325

139. Stock JH (2001) Forecasting Economic Time Series. In: Baltagi BP (ed) A Companion to Theoretical Econometrics. Blackwell, Oxford, Chapter 27

140. Stock JH, Watson MW (2002) Forecasting Using Principal Components from a Large Number of Predictors. J Amer Stat Assoc 97:1167–1179

141. Stock JH, Watson MW (2006) Forecasting Using Many Predic-

tors. In: Elliott G, Granger CWJ, Timmermann A (eds) Handbook of Economic Forecasting, vol 1. Elsevier, Amsterdam

142. Stone CJ (1977) Consistent Nonparametric Regression. Ann Stat 5:595–645

143. Taylor SJ (1986) Modelling Financial Time Series. Wiley, New York

144. Teräsvirta T (1994) Specification, Estimation and Evaluation of Smooth Transition Autoregressive Models. J Amer Stat Assoc 89:208–218

145. Teräsvirta T (2006) Forecasting economic variables with nonlinear models. In: Elliott G, Granger CWJ, Timmermann A (eds) Handbook of Economic Forecasting, vol 1. Elsevier, Amsterdam, pp 413–457

146. Teräsvirta T, Anderson H (1992) Characterizing Nonlinearities in Business Cycles using Smooth Transition Autoregressive Models. J Appl Econ 7:119–139

147. Teräsvirta T, Lin CF, Granger CWJ (1993) Power of the Neural Network Linearity Test. J Time Ser Analysis 14:209–220

148. Tong H (1983) Threshold Models in Nonlinear Time Series Analysis. Springer, New York

149. Tong H (1990) Nonlinear Time Series: A Dynamical Systems Approach. Oxford University Press, Oxford

150. Trippi R, Turban E (1992) Neural Networks in Finance and Investing: Using Artificial Intelligence to Improve Real World Performance. McGraw-Hill, New York

151. Tsay RS (1998) Testing and Modeling Multivariate Threshold Models. J Amer Stat Assoc 93:1188–1202

152. Tsay RS (2002) Nonlinear Models and Forecasting. In: Clements MP, Hendry DF (eds) A Companion to Economic Forecasting. Blackwell, Oxford, Chapter 20

153. Tsay RS (2005) Analysis of Financial Time Series, 2nd edn. Wiley, New York

154. Varian HR (1975) A Bayesian Approach to Real Estate Assessment. In: Fienberg SE, Zellner A (eds) Studies in Bayesian Econometrics and Statistics in Honor of L.J. Savage. North Holland, Amsterdam, pp 195–208

155. Watson GS (1964) Smooth Regression Analysis. Sankhya Series A 26:359–372

156. West KD (1996) Asymptotic inference about predictive ability. Econometrica 64:1067–1084

157. West KD, Edison HJ, Cho D (1993) A Utility Based Comparison of Some Models of Exchange Rate Volatility. J Int Econ 35:23–45

158. White H (1989) An Additional Hidden Unit Tests for Neglected Nonlinearity in Multilayer Feedforward Networks. In: Proceedings of the International Joint Conference on Neural Networks, Washington, DC. IEEE Press, New York, II, pp 451–455

159. White H (1994) Estimation, Inference, and Specification Analysis. Cambridge University Press, Cambridge

160. White H (2000) A Reality Check for Data Snooping. Econometrica 68(5):1097–1126

161. White H (2006) Approximate Nonlinear Forecasting Methods. In: Elliott G, Granger CWJ, Timmermann A (eds) Handbook of Economic Forecasting, vol 1. Elsevier, Amsterdam, chapter 9

162. Zellner A (1986) Bayesian Estimation and Prediction Using Asymmetric Loss Functions. J Amer Stat Assoc 81:446–451

163. Ziegelmann FA (2002) Nonparametric estimation of volatility functions: the local exponential estimator. Econ Theor 18:985–991

Financial Forecasting, Sensitive Dependence

MOTOTSUGU SHINTANI
Vanderbilt University, Nashville, USA

Article Outline

Glossary

Global Lyapunov exponent A global stability measure of the nonlinear dynamic system. It is a long-run average of the exponential growth rate of infinitesimally small initial deviation and is uniquely determined in the ergodic and stationary case. In this sense, this initial value sensitivity measure does not depend on the initial value. A system with positive Lyapunov exponents is considered chaotic for both deterministic and stochastic cases.

Local Lyapunov exponent A local stability measure based on a short-run average of the exponential growth rate of infinitesimally small initial deviations. Unlike the global Lyapunov exponent, this initial value sensitivity measure depends both on the initial value and the horizon for the average calculation. A smaller local Lyapunov exponent implies a better performance at the point of forecast.

Noise amplification In a stochastic system with the additive noise, the effect of shocks can either grow, remain, or die out with the forecast horizon. If the system is nonlinear, this effect depends both on the initial value and size of the shock. For a chaotic system, the degree of noise amplification is so high that it makes the forecast almost identical to the iid forecast within the next few steps ahead.

Nonlinear impulse response function In a stochastic system with the additive noise, the effect of shocks on the variable in subsequent periods can be summarized in impulse response functions. If the system is

linear, the impulse response does not depend on the initial value and its shape is proportional to the size of shocks. If the system is nonlinear, however, the impulse response depends on the initial value, or the history, and its shape is no longer proportional to the size of shocks.

Definition of the Subject

Empirical studies show that there are at least some components in future asset returns that are predictable using information that is currently available. When the linear time series models are employed in prediction, the accuracy of the forecast does not depend on the current return or the initial condition. In contrast, with nonlinear time series models, properties of the forecast error depend on the initial value or the history. The effect of the difference in initial values in a stable nonlinear model, however, usually dies out quickly as the forecast horizon increases. For both deterministic and stochastic cases, the dynamic system is chaos if a small difference in the initial value is amplified at an exponential rate. In a chaotic nonlinear model, the reliability of the forecast can decrease dramatically even for a moderate forecast horizon. Thus, the knowledge of the sensitive dependence on initial conditions in a particular financial time series offers practically useful information on its forecastability. The most frequently used measure of initial value sensitivity is the largest Lyapunov exponent, defined as the long-run average growth rate of the difference between two nearby trajectories. It is a global initial value sensitivity measure in the sense that it contains the information on the global dynamic property of the whole system. The dynamic properties around a single point in the system can be also described using other local measures. Both global and local measures of the sensitive dependence on initial conditions can be estimated nonparametrically from data without specifying the functional form of the nonlinear autoregressive model.

Introduction

When the asset market is efficient, all the information contained in the history of the asset price is already reflected in the current price of the asset. Mathematically, the conditional mean of asset returns becomes independent of the conditioning information set, and thus price changes must be unpredictable (a martingale property). A convenient model to have such a characteristic is a random walk model with independent and identically distributed (iid) increments given by

$$\ln P_t - \ln P_{t-1} = x_t$$

for $t = 0, 1, 2, \ldots$, where P_t is the asset price and x_t is an iid random variable with mean μ_x and variance σ_x^2. When $\mu_x = 0$, the model becomes a random walk without drift, otherwise, it is a random walk with drift μ_x.

Chaos is a nonlinear deterministic process that can generate a random-like fluctuation. In principle, if a purely deterministic model, instead of a random walk process, is used to describe the dynamics of the asset return x_t, all future asset returns should be completely predictable. However, in the case of chaos, a small perturbation can make the performance of a few steps ahead forecast almost identical to that of a random walk forecast. A leading example is the tent map:

$$x_t = 1 - |2x_{t-1} - 1|$$

with some initial value x_0 between 0 and 1. This map almost surely has the uniform distribution $U(0, 1)$ as its natural measure, defined as the distribution of a typical trajectory of x_t. This dynamic system thus provides aperiodic trajectory or random-like fluctuation of x_t as the number of iteration increases. By introducing a randomness in the initial value x_0, marginal distribution of x_t approaches the natural measure. This property, referred to as ergodicity, implies that the temporal average of any smooth function of a trajectory x_t, $M^{-1} \sum_{t=0}^{M-1} h(x_t)$, converges to a mathematical expectation $E[h(x_t)] = \int_{-\infty}^{\infty} h(x)\pi(x)\mathrm{d}x$ as M tends to infinity, where the marginal distribution of x_t is expressed in terms of the probability density function (pdf) $\pi(x)$. The natural measure $U(0, 1)$ is also a stationary or invariant distribution since the marginal distribution of x_t for any $t \geq 1$, is $U(0, 1)$ whenever initial value x_0 follows $U(0, 1)$. In this case, the mean μ_x and variance σ_x^2 are 1/2 and 1/12, respectively. Furthermore, almost all the trajectories are second-order white noise in the sense that they have a flat spectrum and zero autocorrelation at all leads and lags.

Therefore, the knowledge of the marginal distribution $\pi(x)$ or spectrum of asset returns cannot be directly used to distinguish between the case of a random walk combined with an iid random variable and the case of the tent map generating the returns. Yet, the two cases have significantly different implications on the predictability of asset returns, at least for the extremely short horizon. When the initial value x_0 is given, using $\mu_x = 1/2$ as a one-period-ahead forecast at $t = 0$ provides a minimum mean square forecast error (MSFE) of $\sigma_x^2 = 1/2$ for the former case. In contrast, using $1 - |2x_0 - 1|$ as the forecast gives zero forecast error for the latter case. With a very tiny perturbation, however, the MSFE of the multiple-period-ahead forecast for the latter case quickly approaches $\sigma_x^2 = 1/2$, which is identical to that of the random walk case.

Another example is the logistic map:

$$x_t = 4x_{t-1}(1 - x_{t-1})$$

with some initial value x_0 between 0 and 1. This map again provides chaotic fluctuation with the natural measure almost surely given by beta distribution $B(1/2, 1/2)$. Provided the same distribution as its stationary or invariant distribution, the mean μ_x and variance σ_x^2 are 1/2 and 1/8, respectively (see [37] for the invariant distribution of the logistic map in general). Again, the random walk model combined with an iid random variable with the same marginal distribution $B(1/2, 1/2)$ is not distinguishable from the logistic map based only on the marginal distribution nor spectra. But the two have very different predictive implications.

The key feature of chaos that is not observed in the iid random variable is the sensitivity of the trajectories to the choice of initial values. This sensitivity can be measured by the Lyapunov exponent which is defined as the average rate of divergence (or convergence) of two nearby trajectories. Indeed, the positivity of the Lyapunov exponent in a bounded dissipative nonlinear system is a widely used formal definition of chaos. To derive this measure for the two examples above, first consider a one-dimensional general nonlinear system

$$x_t = f(x_{t-1})$$

where $f : R \to R$ is a continuously differentiable map, with two initial values $x_0 = \overline{x}_0$ and $x_0^* = \overline{x}_0 + \delta$ where δ represents infinitesimal difference in the initial condition. When the distance between two trajectories $\{x_t\}_{t=0}^{\infty}$ and $\{x_t^*\}_{t=0}^{\infty}$ after M steps is measured by the exponential growth rate of δ using $x_M^* - x_M = \delta \exp(M\lambda_M(\overline{x}_0))$, the average of the growth rate in each iteration is given by

$$\lambda_M(\overline{x}_0) = \frac{1}{M} \ln \left| \frac{x_M^* - x_M}{\delta} \right| .$$

Further, let $f^{(M)}$ be the M-fold composition of f. Then from the first order term in the Taylor series expansion of $x_M^* - x_M = f^{(M)}(\overline{x}_0 + \delta) - f^{(M)}(\overline{x}_0)$ around $\overline{x}_0$, combined with the chain rule applied to $df^{(M)}(x)/dx|_{x=\overline{x}_0}$ yields $[f'(x_0)f'(x_1)\cdots f'(x_{M-1})]\delta = [\prod_{t=1}^{M} f'(x_{t-1})]\delta$. Thus, the product $\prod_{t=1}^{M} f'(x_{t-1})$ is the amplifying factor to the initial difference after M periods. Substituting this approximation result into the average growth rate formula yields $\lambda_M(\overline{x}_0) = M^{-1} \sum_{t=1}^{M} \ln |f'(x_{t-1})|$. This measure is called a *local Lyapunov exponent* (of order M) and in general depends on both $\overline{x}_0$ and M (see Fig. 1).

Next, consider the case M tending to infinity. If x_t is ergodic and stationary, $M^{-1} \sum_{t=1}^{M} \ln |f'(x_{t-1})|$ converges
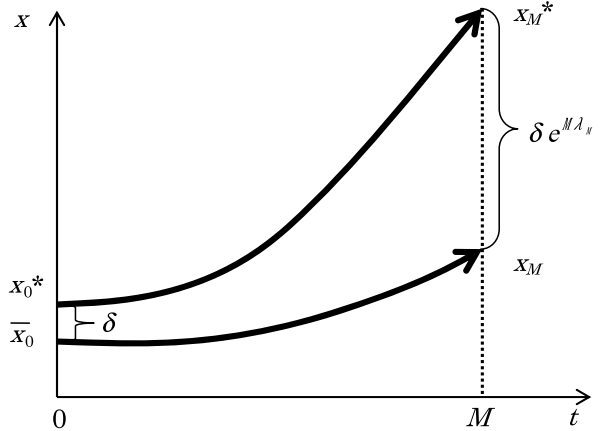
Financial Forecasting, Sensitive Dependence, Figure 1
Lyapunov exponent is an exponential growth rate

to $E[\ln |f'(x_{t-1})|] = \int_{-\infty}^{\infty} \ln |f'(x)| \, \pi(x)dx$, which does not depend on $\overline{x}_0$. Thus, a *global Lyapunov exponent*, or simply a *Lyapunov exponent*, of a one-dimensional system is defined as $\lambda = \lim_{M\to\infty} \lambda_M(\overline{x}_0)$ or

$$\lambda = \lim_{M\to\infty} \frac{1}{M} \sum_{t=1}^{M} \ln \left| f'(x_{t-1}) \right| .$$

According to this definition, the computation of the Lyapunov exponent of the tent map is straightforward. Since the tent map is $x_t = 2x_{t-1}$ for $0 \le x_{t-1} \le 1/2$ and $x_t = 2 - 2x_{t-1}$ for $1/2 < x_{t-1} \le 1$, its first derivative $f'(x_{t-1})$ is 2 for $0 \le x_{t-1} \le 1/2$ and -2 for $1/2 < x_{t-1} \le 1$. Using the uniform distribution as its stationary distribution, we have

$$\lambda = \int_0^{1/2} \ln |2| \, dx + \int_{1/2}^1 \ln |-2| \, dx = \ln 2 \, (\approx 0.693) .$$

Similarly, for the logistic map $x_t = ax_{t-1}(1 - x_{t-1})$ with $a = 4$,

$$\lambda = \int_0^1 \frac{\ln |4 - 8x|}{\pi \sqrt{x(1-x)}} dx = \ln 4 - \ln 2 = \ln 2 .$$

Thus, both the tent map and the logistic map with $a = 4$ have a common positive Lyapunov exponent. The value $\ln 2$ implies that, on average, the effect of an initial deviation doubles each time of iteration. Such a rapid rate of divergence is the source of the fact that their trajectories become unpredictable very quickly. Chaos is thus characterized by sensitive dependence on initial conditions measured by a positive Lyapunov exponent.

Let us next consider a simple linear difference equation $x_t = \rho x_{t-1}$ with $|\rho| < 1$. Since its first derivative

$f'(x)$ is a constant ρ, not only the Lyapunov exponent but also the local Lyapunov exponent $\lambda_M(\overline{x}_0)$ does not depend on the initial value $\overline{x}_0$. For example, when $\rho = 0.5$, $\lambda = \lambda_M(\overline{x}_0) = -\ln 2 \ (\approx -0.693)$. The logistic map $x_t = a x_{t-1}(1 - x_{t-1})$, can be either chaotic or stable depending on the choice of a. When $a = 1.5$, all the trajectories converge to a point mass at $x_t = 1/3$, where the first derivative is $1/2$ thus $\lambda = -\ln 2$. For these two examples, the system has a common negative Lyapunov exponent. In this case, the effect of the initial condition is short-lived and the system is not sensitive to initial conditions. The value $-\ln 2$ implies that, on average, the effect of initial deviation reduces by one half each time of iteration.

Knowing the Lyapunov exponent of the asset returns, or their transformation, thus offers a useful information regarding the predictability of a financial market. In particular, for a system with sensitive dependence (namely, the one with a positive Lyapunov exponent), the performance of a multiple step forecast can worsen quickly as the forecast horizon increases if there are (i) a small uncertainty about the current value at the time of forecast (observation noise) and/or (ii) a small additive noise in the system (system noise).

Lyapunov Exponent and Forecastability

Lyapunov Spectrum

As a global measure of initial value sensitivity in a multi-dimensional system, the largest Lyapunov exponent and Lyapunov spectrum will first be introduced. For the p-dimensional deterministic nonlinear system,

$$x_t = f\left(x_{t-1}, \ldots, x_{t-p}\right) ,$$

where $f \colon R^p \to R$ is continuously differentiable, the (global) largest Lyapunov exponent of the system is defined as

$$\lambda = \lim_{M \to \infty} \frac{1}{2M} \ln \left| \nu_1(\mathsf{T}'_M \mathsf{T}_M) \right|$$

where $\nu_1(\mathsf{T}'_M \mathsf{T}_M)$ is the largest eigenvalue of $\mathsf{T}'_M \mathsf{T}_M$, and $\mathsf{T}_M = J_{M-1} \cdot J_{M-2} \cdot \cdots \cdot J_0$. Here J_{t-1}'s are Jacobian matrices defined as

$$J_{t-1} =$$

$$\begin{bmatrix} \Delta f_1(\mathsf{X}_{t-1}) & \Delta f_2(\mathsf{X}_{t-1}) & \cdots & \Delta f_{p-1}(\mathsf{X}_{t-1}) & \Delta f_p(\mathsf{X}_{t-1}) \\ 1 & 0 & \cdots & 0 & 0 \\ 0 & 1 & \cdots & 0 & 0 \\ \vdots & \vdots & \ddots & \vdots & \vdots \\ 0 & 0 & \cdots & 1 & 0 \end{bmatrix}$$

for $t = 1, \ldots, M$, where $\Delta f_j(\mathsf{X}_{t-1}) = \partial f(\mathsf{X}_{t-1})/\partial x_{t-j}$, for $j = 1, \ldots, p$, are partial derivatives of the conditional mean function evaluated at $\mathsf{X}_{t-1} = (x_{t-1}, \ldots, x_{t-p})'$.

Using an analogy to the one-dimensional case, the local Lyapunov exponent can be defined similarly by $\lambda_M(\mathbf{x}) = (2M)^{-1} \ln \left| \nu_1(\mathsf{T}'_M \mathsf{T}_M) \right|$ with initial value $\mathbf{x} = (\overline{x}_0, \overline{x}_{-1}, \ldots, \overline{x}_{-p+1})'$. Note that $(2M)^{-1} \ln \left| \nu_1(\mathsf{T}'_M \mathsf{T}_M) \right|$ reduces to the sum of absolute derivatives in logs used for the one-dimensional case since $(2M)^{-1} \sum_{t=1}^{M} \ln[f'(x_{t-1})^2] = M^{-1} \sum_{t=1}^{M} \ln |f'(x_{t-1})|$.

In the multi-dimensional case, the whole spectrum of Lyapunov exponents can be also considered using ith Lyapunov exponent λ_i, for $i = 1, \ldots, p$, defined by replacing the largest eigenvalue ν_1 with the ith largest eigenvalue ν_i. A set of all Lyapunov exponents is called the Lyapunov spectrum. Geometrically, each Lyapunov exponent represents the rate of growth (or contraction) of the corresponding principal axis of a growing (or shrinking) ellipsoid. An attracting set of a dynamic system, or simply the attractor, is defined as the set to which x_t approaches in the limit. The attractor can be a point, a curve, a manifold, or more complicated set. The Lyapunov spectrum contains information on the type of the attractor. For example, a system with all negative Lyapunov exponents has an equilibrium point as an attracting set. To understand this claim, let $\mathbf{x}_{\text{EQ}}$ be an equilibrium point and consider a small initial deviation $\boldsymbol{\delta} = (\delta_1, \ldots, \delta_p)'$ from $\mathbf{x}_{\text{EQ}}$. By the linearization of $f \colon R^p \to R$ at $\mathbf{x}_{\text{EQ}}$, the deviation from $\mathbf{x}_{\text{EQ}}$ after M periods is approximated by $c_1 \eta_1 \exp\{\widetilde{\lambda}_1 M\} + \cdots + c_p \eta_p \exp\{\widetilde{\lambda}_p M\}$, where $\widetilde{\lambda}_i$'s and η_i's are the eigenvalues and eigenvectors of J_{EQ}, respectively, where J_{EQ} is J_{t-1} evaluated at $\mathsf{X}_{t-1} = \mathbf{x}_{\text{EQ}}$, and c_i's are scalar constants. The real part of $\widetilde{\lambda}_i$, denoted by $\text{Re}[\widetilde{\lambda}_i]$, represents the rate of growth (contraction) around the equilibrium point $\mathbf{x}_{\text{EQ}}$ along the direction of η_i if $\text{Re}[\widetilde{\lambda}_i]$ is positive (negative). Thus if $\text{Re}[\widetilde{\lambda}_i] < 0$ for all $i = 1, \ldots, p$, $\mathbf{x}_{\text{EQ}}$ is asymptotically stable and is an attractor. Otherwise, $\mathbf{x}_{\text{EQ}}$ is either unstable with $\text{Re}[\widetilde{\lambda}_i] > 0$ for all i, or a saddle point with $\text{Re}[\widetilde{\lambda}_i] > 0$ for some i, provided that none of $\text{Re}[\widetilde{\lambda}_i]$ is zero. In this simple case, ith Lyapunov exponent λ_i corresponds to $\text{Re}[\widetilde{\lambda}_i]$.

Among all the Lyapunov exponents, the largest Lyapunov exponent λ_1, or simply λ, is a key measure to distinguish chaos from other stable systems. By using an analogy to the equilibrium point example, $\lambda > 0$ implies the expansion in the direction of η_1. An attractor requires that the sum of all the Lyapunov exponents be negative since contraction on the whole must be stronger than the expansion. When this condition is met with some positive λ_i's, the system is said to have a strange attractor. Chaos is thus excluded if the largest Lyapunov exponent is not positive.

Financial Forecasting, Sensitive Dependence, Table 1
Lyapunov spectrum and attractors

Attractor	Point	Closed curve	k-torus	Strange attractor
Steady state	equilibrium point	limit cycle (periodic)	k-periodic	chaotic
Dimension	0	1	k (integer)	noninteger
Lyapunov exponents	$\lambda_i < 0\ (i = 1, \ldots, p)$	$\lambda_1 = 0$ $\lambda_i < 0\ (i = 2, \ldots, p)$	$\lambda_1 = \cdots = \lambda_k = 0$ $\lambda_i < 0\ (i = k + 1, \ldots, p)$	$\lambda_1 > 0$

A system with a zero largest Lyapunov exponent implies that the average distance of two orbits (along some directions) is same as their initial deviation, the property often referred to as Lyapunov stability. A zero largest Lyapunov exponent and strictly negative remaining Lyapunov exponents lead to a system with a limit cycle. If only the first two (k) largest Lyapunov exponents are zero, the system has a two-torus (k-torus) attractor. The types of attractors and their relationship to the signs of Lyapunov exponents are summarized in Table 1.

Entropy and Dimension

In a deterministic system with initial value sensitivity, the information on how quickly trajectories separate on the whole has a crucial implication in the predictability. This is because if two different trajectories which are initially indistinguishable become distinguishable after a finite number of steps, the knowledge of the current state is useful in forecasting only up to a finite times ahead. Kolmogorov entropy of the system measures the rate at which information is produced and has a close relationship to the Lyapunov exponents. In general, the sum of all positive Lyapunov exponents provides an upper bound to Kolmogorov entropy, which contains the information on how quickly trajectories separate on the whole. Under some conditions, both the entropy and the sum become identical (see [22,56]). This fact can intuitively be understood as follows. Suppose a system with k positive Lyapunov exponents and an attractor of size L. Here, the size of an attractor roughly refers to the range of an invariant distribution of an attractor, which becomes unpredictable as a result of magnified small initial deviation of size d. Note that the length of the first k principal axes after M steps of iteration is proportional to $\exp(M \sum_{i=1}^{k} \lambda_i)$. From $d \exp(M \sum_{i=1}^{k} \lambda_i) = L$, the expected time M to reach the size of attractor is given by $(1/\sum_{i=1}^{k} \lambda_i) \ln(L/d)$. This result implies that the larger $\sum_{i=1}^{k} \lambda_i$ becomes, the shorter the period during which the path is predictable.

Lyapunov exponents are also closely related to the notion of dimension designed to classify the type of attractors. An equilibrium point has zero dimension. A limit cycle is one-dimensional since it resembles an interval in a neighborhood of any point. A k-torus is k-dimensional since it locally resembles an open subset of R^k. However, the neighborhood of any point of a strange attractor does not resemble any Euclidean space and does not have integer dimension. Among many possibilities of introducing a non-integer dimension, one can consider the Lyapunov dimension, or Kaplan–Yorke dimension, defined as

$$D_L = k + \frac{1}{|\lambda_{k+1}|} \sum_{i=1}^{k} \lambda_i$$

where λ_i is the ith Lyapunov exponent and k is the largest integer for which $\sum_{i=1}^{k} \lambda_i \geq 0$. This definition provides the dimension of zero for an equilibrium point, one for a limit cycle, and k for a k-torus. For a chaotic example, suppose a three-dimensional system with a positive Lyapunov exponent ($\lambda_1 = \lambda_+ > 0$), a zero Lyapunov exponent ($\lambda_2 = 0$), and a negative Lyapunov exponent ($\lambda_3 = \lambda_- < 0$). The Lyapunov dimension D_L is then given by $2 + \lambda_+/|\lambda_-|$ which is a fraction that lies strictly between 2 and 3 since an attractor should satisfy $\lambda_+ + \lambda_- < 0$. Likewise, in general, the Lyapunov dimension D_L will be a fraction between k and $k + 1$ since $\sum_{i=1}^{k} \lambda_i \leq |\lambda_{k+1}|$ always holds by the definition of k (see Fig. 2).

Since the Lyapunov spectrum contains richer information than the largest Lyapunov exponent alone, several empirical studies reported the Lyapunov spectrum [18], or the transformation such as Kolmogorov entropy and Lyapunov dimension [1,2] of financial time series. However, one must be careful on the interpretation of these quantities since their properties under noisy environment is not rigorously established. In addition, it should be noted that some other forms of the entropy and the dimension can be computed without estimating each Lyapunov exponent separately. For example, [36] recommended using an approximation to Kolmogorov entropy, given by

$$K_2 = \lim_{\substack{\delta \to 0 \\ p \to \infty}} \ln \left(\frac{C^p(\delta)}{C^{p+1}(\delta)} \right)$$
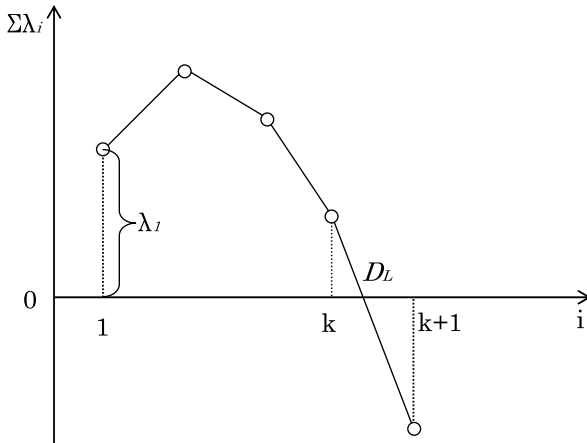
Financial Forecasting, Sensitive Dependence, Figure 2
Lyapunov dimension

where $C^p(\delta)$ is the correlation integral defined by

$$C^p(\delta) = \lim_{T \to \infty} \sharp\{(t,s)| \, \|\mathsf{X}_t - \mathsf{X}_s\| < \delta\}/T^2$$

where $\mathsf{X}_t = (x_t, \ldots, x_{t-p+1})'$ and $\|\cdot\|$ is a vector norm. The approximation given by K_2 provides a lower bound of Kolmogorov entropy (see [22]). The correlation dimension, a type of dimension, can also be defined as

$$D_C = \lim_{\delta \to 0} \frac{\ln C^p(\delta)}{\ln \delta} \, .$$

Both the K_2 entropy and correlation dimension can be estimated by replacing $C^p(\delta)$ with its sample analogue. In applications to financial time series, these two measures are computed in [30] and [50]. Finally, note that the correlation integral has been used as the basis of the BDS test, a well-established nonlinear dependence test frequently used in economic application, developed by [9]. Formally, the test statistic relies on the sample analogue of $C^p(\delta) - [C^1(\delta)]^p$ and follows normal distribution under the null hypothesis of iid randomness. The BDS test appears to have a good power against the alternative of linear or nonlinear dependence including some low-dimensional chaotic process. Thus, the BDS test is useful in providing the indirect evidence of sensitive dependence and can be complementarily used along with a more direct test based on Lyapunov exponents (see [5] for an example on the comparison between the two approaches).

System Noise and Noisy Chaos

Unlike the data generated from a purely deterministic system, economic and financial data are more likely to be contaminated by noise. There are two main types of random noise used to extend the deterministic model to the stochastic model in the analysis of initial value sensitivity: observation noise and system noise. In the case of the observation noise, or measurement noise, observables are given as the sum of stochastic noise and the unobservables generated from the deterministic model. In contrast, with the system noise, or dynamic noise, observables are generated directly from a nonlinear autoregressive (AR) model. In practice, it is often convenient to introduce the system noise in the additive manner. Theoretically, system noise can make the system to have a unique stationary distribution. Note that for the examples of tent map and logistic map, aperiodic trajectory, or random-like fluctuation, could not be obtained with some choice of initial condition with measure zero. In general, the deterministic system can have infinitely many stationary distributions. However, typically, the presence of additive noise can exclude all degenerate marginal distributions. Furthermore, additive system noise is convenient to generalize the use of the Lyapunov exponents, originally defined in the deterministic system as a measure of sensitive dependence, to the case of a stochastic system.

To see this point, first, consider the following simple linear system with an additive system noise. Adding an iid stochastic error term ε_t, with $E(\varepsilon_t) = 0$ and $E(\varepsilon_t^2) = \sigma^2$, in the previously introduced linear difference equation leads to a linear AR model of order one,

$$x_t = \rho x_{t-1} + \varepsilon_t \, .$$

The model has a stationary distribution if $|\rho| < 1$. Even if the error term is present, since $f'(x_{t-1}) = \rho$, a one-dimensional Lyapunov exponent can be computed as $\lambda = \ln |\rho| < 0$, the value identical to the case of the deterministic linear difference equation. Thus, the stationarity condition $|\rho| < 1$ in the linear model always implies a negative Lyapunov exponent, while a unit root process $\rho = 1$ implies zero Lyapunov exponent.

Next, consider the introduction of a system noise to a nonlinear system. A general (stationary) nonlinear AR model of order one is defined as

$$x_t = f(x_{t-1}) + \varepsilon_t$$

where $f : R \to R$ is a smooth function. For a known unique stationary marginal distribution $\pi(x)$, Lyapunov exponent can be computed as $E[\ln |f'(x_{t-1})|] = \int_{-\infty}^{\infty} \ln |f'(x)| \, \pi(x) \mathrm{d}x$. Thus, by using an analogy of the definition of deterministic chaos, *noisy chaos* can be defined as a stationary nonlinear AR model with a positive Lyapunov exponent. Even if an analytical solution is

not available, the value of Lyapunov exponent is typically obtained numerically or by simulation. Similarly, for the multidimensional nonlinear AR model,

$$x_t = f(x_{t-1}, \ldots, x_{t-p}) + \varepsilon_t \,,$$

(noisy) chaos can be defined by a positive largest Lyapunov exponent computed from the Jacobian and the stationary joint distribution of $\mathsf{X}_{t-1} = (x_{t-1}, \ldots, x_{t-p})'$. Furthermore, as long as the process has a stationary distribution, for both the chaotic and non-chaotic case, M-period ahead least squares predictor $f_M(\mathbf{x}) \equiv E[x_{t+M} | \mathsf{X}_t = \mathbf{x}]$ and its conditional MSFE $\sigma_M^2(\mathbf{x}) \equiv E[\{x_{t+M} - f_M(\mathbf{x})\}^2 | \mathsf{X}_t = \mathbf{x}]$ depend on the initial condition $\mathbf{x} = (\overline{x}_0, \overline{x}_{-1}, \ldots, \overline{x}_{-p+1})'$ but do not depend on the timing of forecast t.

Noise Amplification

The next issue involves the prediction in the stochastic dynamic system. When additive noise is present in the nonlinear system, the amplification of noise can depend on the initial values and is not necessarily monotonic in horizon. This feature is not unique to the chaotic model but holds for general nonlinear models. However, a small noise is expected to be amplified rapidly in time if the nonlinear system is chaotic.

To understand the process of noise amplification, consider the previously introduced linear AR model of order one with a non-zero coefficient ρ and an initial condition $x_0 = \overline{x}_0$. Then, at the period M,

$$
\begin{aligned}
x_M &= \rho\{\rho x_{M-2} + \varepsilon_{M-1}\} + \varepsilon_M \\
&= \rho^2 x_{M-2} + \rho \varepsilon_{M-1} + \varepsilon_M \\
&= \rho^M \overline{x}_0 + \varepsilon_M + \cdots + \rho^{M-1} \varepsilon_1 \,.
\end{aligned}
$$

Since $\{\varepsilon_1, \varepsilon_2, \ldots, \varepsilon_M\}$ are not predictable at period 0, the least square M-period ahead predictor is $\rho^M \overline{x}_0$ with its MSFE σ_M^2 given by $\mu_M \sigma^2$ where

$$\mu_M = 1 + \cdots + \rho^{2(M-1)} = 1 + \sum_{j=1}^{M-1} \rho^{2j}$$

is a monotonically increasing proportional factor that does not depend on $\overline{x}_0$. Since $\mu_M > 1$, MSFE is strictly greater than the variance of the noise for all M. However, for a stationary process with $|\rho| < 1$, increments in such a noise amplification become smaller and μ_M converges to $1/(1 - \rho^2)$ as M tends to infinity. Thus, eventually, the MSFE converges to the unconditional variance $\sigma_x^2 = \sigma^2/(1 - \rho^2)$. In a special case with $\rho = 0$, when the asset price have iid increments, the proportional factor

becomes 1 for all M giving its MSFE $\sigma_M^2 = \sigma_x^2 = \sigma^2$ for all M.

Suppose, instead, a general nonlinear AR model of order one with an initial condition $x_0 = \overline{x}_0$. In addition, let $|\varepsilon_t| \leq \zeta$ almost surely, where $\zeta > 0$ is a small constant. By Taylor series expansion, for $M \geq 1$,

$$
\begin{aligned}
x_M &= f\{f(x_{M-2}) + \varepsilon_{M-1}\} + \varepsilon_M \\
&= f^{(2)}(x_{M-2}) + f'\{f(x_{M-2})\}\varepsilon_{M-1} + \varepsilon_M + O(\zeta^2) \,.
\end{aligned}
$$

Using the fact that $x_{M-2} = f^{(M-2)}(\overline{x}_0) + O(\zeta)$, and repeating applications of Taylor series expansion,

$$
\begin{aligned}
x_M &= f^{(2)}(x_{M-2}) + f'\{f^{(M-1)}(\overline{x}_0)\}\varepsilon_{M-1} + \varepsilon_M + O(\zeta^2) \\
&= f^{(M)}(\overline{x}_0) + \varepsilon_M + f'\{f^{(M-1)}(\overline{x}_0)\}\varepsilon_{M-1} + \cdots \\
&\quad + \prod_{k=1}^{M-1} f'\{f^{(k)}(\overline{x}_0)\}\varepsilon_1 + O(\zeta^2) \,.
\end{aligned}
$$

Thus the least square M-period ahead predictor is $f^{(M)}(\overline{x}_0)$ with its conditional MSFE given by

$$\sigma_M^2(\overline{x}_0) = \mu_M(\overline{x}_0)\sigma^2 + O(\zeta^3)$$

where

$$\mu_M(\overline{x}_0) = 1 + \sum_{j=1}^{M-1} \left[\prod_{k=j}^{M-1} f'\{f^{(k)}(\overline{x}_0)\} \right]^2 \,.$$

A comparison of μ_M for the linear model and $\mu_M(\overline{x}_0)$ for the nonlinear model provides some important features of the nonlinear prediction. First, unlike the linear case, the proportional factor now depends not only on the forecast horizon M but also on the initial condition $\overline{x}_0$. Thus, in general, performance of the nonlinear prediction depends on where you are.

Second, $\mu_M(\overline{x}_0)$ does not need to be monotonically increasing with M in nonlinear case. The formula for $\mu_M(\overline{x}_0)$ can be rewritten as

$$\mu_{M+1}(\overline{x}_0) = 1 + \mu_M(\overline{x}_0)f'\{f^{(M)}(\overline{x}_0)\}^2 \,.$$

Thus, $\mu_{M+1}(\overline{x}_0) < \mu_M(\overline{x}_0)$ is possible when $f'\{f^{(M)}(\overline{x}_0)\}^2 < 1 - 1/\mu_M(\overline{x}_0)$. Therefore, with some initial value and M, the $(M + 1)$-period ahead MSFE can be smaller than the M-period ahead MSFE.

Third, and most importantly, unlike the stationary linear model, which imposes the restriction $|\rho| < 1$, $|f'(x)| > 1$ is possible for a large range of values of x in the nonlinear model even if it has a bounded and stationary distribution. In such a case, $\mu_M(\overline{x}_0)$ can grow rapidly

for the moderate or short forecast horizon M. The rapid noise amplification makes the long-horizon forecast very unreliable especially when the model is chaotic. To see this point, it is convenient to rewrite the proportional factor $\mu_M(\overline{x}_0)$ in terms of the local Lyapunov exponent as

$$\mu_M(\overline{x}_0) = 1 + \sum_{j=1}^{M-1} \exp\left\{2(M-j)\lambda_{M-j}(f^{(j)}(\overline{x}_0))\right\} .$$

When the local Lyapunov exponent is positive, the proportional factor grows at an exponential rate as M grows. Recall that in the case of iid forecast (random walk forecast in terms of price level), the MSFE σ_M^2 becomes σ_x^2. Likewise, for the chaotic case with infinitesimally small σ^2, the MSFE σ_M^2 reaches σ_x^2 only after a few steps even if the MSFE is close to zero for the one-step ahead forecast. Thus, the global Lyapunov exponent or other local measures of sensitive dependence contain important information on the predictability in the nonlinear time series framework.

Nonparametric Estimation of the Global Lyapunov Exponent

Local Linear Regression

The measures of initial value sensitivity can be computed from the observed data. Since the Lyapunov exponent is by definition the average growth rate of initial deviations between two trajectories, it can be directly computed by finding pairs of neighbors and then averaging growth rates of the subsequent deviations of such pairs [77]. This 'direct' method, however, provides a biased estimator when there is a random component in the system [51]. A modified regression method proposed by [63] is considered more robust to the presence of measurement noise but not necessarily when the system noise is present. A natural approach to compute the Lyapunov exponent in the nonlinear AR model framework is to rely on the estimation of the nonlinear conditional mean function $f: R^p \to R$. For example, based on an argument similar to the deterministic case, the noisy logistic map, $x_t = ax_{t-1}(1 - x_{t-1}) + \varepsilon_t$, can be either chaotic or stable depending on the value of the parameter a. The first derivative $f'(x) = a - 2ax$ can be evaluated at each data point once an estimate of a is provided. Thus, the parametric approach in the estimation of Lyapunov exponents has been considered in some cases (e. g., [7]). In practice, however, information on the functional form is rarely available and the nonparametric approach is a reasonable alternative. In principle, any nonparametric estimator can be used to estimate the function f

and its partial derivatives in the nonlinear AR model,

$$x_t = f\left(x_{t-1}, \ldots, x_{t-p}\right) + \varepsilon_t$$

where f is smooth and ε_t is a martingale difference sequence with $E[\varepsilon_t | x_{t-1}, x_{t-2}, \ldots] = 0$ and $E[\varepsilon_t^2 | x_{t-1}, x_{t-2}, \ldots] = \sigma^2(x_{t-1}, \ldots, x_{t-p}) = \sigma^2(\mathbf{x})$. To simplify the discussion, here, the one based on a particular type of the kernel regression estimator is explained in detail. Methods based on other types of nonparametric estimators will be later mentioned briefly (see, for example, [27], on the nonparametric approach in time series analysis).

The local linear estimator of the conditional mean function and its first partial derivatives at a point $\mathbf{x}$ can be obtained by minimizing the weighted least squares criterion $\sum_{t=1}^{T}(x_t - \beta_0 - \beta_1'(X_{t-1} - \mathbf{x}))^2 K_H(X_{t-1} - \mathbf{x})$, where H is the $d \times d$ bandwidth matrix, K is d-variate kernel function such that $\int K(u)du = 1$, and $K_H(u) = |H|^{-1/2} K(H^{-1/2}u)$. For example, the standard p-variate normal density

$$K(u) = \frac{1}{2\pi^{-p/2}} \exp(-||u||^2/2)$$

with H given by hI_p where h is a scalar bandwidth and I_p is an identity matrix of order p, can be used in the estimation. The solution to the minimization problem is given by $\widehat{\beta}(x) = (X_x' \mathbf{W}_x X_x)^{-1} X_x' \mathbf{W}_x \mathbf{Y}$ where

$$X_x = \begin{bmatrix} 1 & (X_0 - \mathbf{x})' \\ \vdots & \vdots \\ 1 & (X_{T-1} - \mathbf{x})' \end{bmatrix} ,$$

$\mathbf{Y} = (x_1, \ldots, x_T)'$ and $W_x = \text{diag}\{K_H(X_0 - \mathbf{x}), \ldots, K_H(X_{T-1} - \mathbf{x})\}$. The local linear estimator of the nonlinear function $f(\mathbf{x})$ and its first derivatives $(\partial f)/(\partial x_{t-j})(\mathbf{x})$ for $j = 1, \ldots, p$ are given by $\widehat{\beta}_0(\mathbf{x}) = \widehat{f}(\mathbf{x})$ and

$$\widehat{\beta}_1(\mathbf{x}) = \begin{bmatrix} \widehat{\beta}_{11}(\mathbf{x}) \\ \vdots \\ \widehat{\beta}_{1p}(\mathbf{x}) \end{bmatrix} = \begin{bmatrix} \Delta\widehat{f}_1(\mathbf{x}) \\ \vdots \\ \Delta\widehat{f}_p(\mathbf{x}) \end{bmatrix} ,$$

respectively. [22] and [21] proposed a method, known as the 'Jacobian' method, to estimate the Lyapunov exponent by substituting $\Delta f_i(\mathbf{x})$ in the Jacobian formula by its nonparametric estimator $\Delta\widehat{f}_i(\mathbf{x})$. It should be noted that, in general, the "sample size" T used for estimating Jacobian $\widehat{J}_t$ and the "block length" M, which is the number of evaluation points used for estimating the Lyapunov exponent, can be different. Formally, the Lyapunov exponent estima-

tor of λ is given by

$$\widehat{\lambda}_M = \frac{1}{2M} \ln \nu_1 \left(\widehat{\mathsf{T}}'_M \widehat{\mathsf{T}}_M \right),$$

$$\widehat{\mathsf{T}}_M = \prod_{t=1}^{M} \widehat{J}_{M-t} = \widehat{J}_{M-1} \cdot \widehat{J}_{M-2} \cdots \widehat{J}_0,$$

where

$$\widehat{J}_{t-1} =$$

$$\begin{bmatrix} \Delta\widehat{f}_1(\mathsf{X}_{t-1}) & \Delta\widehat{f}_2(\mathsf{X}_{t-1}) & \cdots & \Delta\widehat{f}_{p-1}(\mathsf{X}_{t-1}) & \Delta\widehat{f}_p(\mathsf{X}_{t-1}) \\ 1 & 0 & \cdots & 0 & 0 \\ 0 & 1 & \cdots & 0 & 0 \\ \vdots & \vdots & \ddots & \vdots & \vdots \\ 0 & 0 & \cdots & 1 & 0 \end{bmatrix},$$

for $t = 0, 1, \dots, M-1$, where $\Delta\widehat{f}_j(\mathbf{x})$ is a nonparametric estimator of $\Delta f_j(\mathbf{x}) = \partial f(\mathbf{x})/\partial x_{t-j}$ for $j = 1, \dots, p$.

As an estimator for the global Lyapunov exponent, setting $M = T$ gives the maximum number of Jacobians and thus the most accurate estimation can be expected. Theoretically, however, it is often convenient to have a block length M smaller than T. For a fixed M, with T tends to infinity, $\widehat{\lambda}_M$ is a consistent estimator of the local Lyapunov exponent with initial value $\mathbf{x} = (\overline{x}_0, \overline{x}_{-1}, \dots, \overline{x}_{-p+1})'$ (see [48]). In case both M and T increase with M/T tends to zero, $\widehat{\lambda}_M$ is still a consistent estimator of the global Lyapunov exponent.

Statistical Inference on the Sign of Lyapunov Exponent

Since the positive Lyapunov exponent is the condition that distinguishes the chaotic process from the stable system without high initial value sensitivity, conducting the inference regarding the sign of the Lyapunov exponent is often of practical interest. For such inference, a consistent standard error formula for $\widehat{\lambda}_M$ is available. Under the condition that M grows at a sufficiently slow rate, a standard error can be computed by $\sqrt{\widehat{\Phi}/M}$ where

$$\widehat{\Phi} = \sum_{j=-M+1}^{M-1} w(j/S_M)\widehat{\gamma}(j)$$

$$\text{with} \quad \widehat{\gamma}(j) = \frac{1}{M} \sum_{t=|j|+1}^{M} \widehat{\eta}_t \widehat{\eta}_{t-|j|},$$

$$\widehat{\eta}_t = \widehat{\xi}_t - \widehat{\lambda}_M \quad \text{with} \quad \widehat{\xi}_t = \frac{1}{2} \ln \left(\frac{\nu_1\left(\widehat{\mathsf{T}}'_t\widehat{\mathsf{T}}_t\right)}{\nu_1\left(\widehat{\mathsf{T}}'_{t-1}\widehat{\mathsf{T}}_{t-1}\right)} \right)$$

$$\text{for} \quad t \geq 2 \quad \text{and} \quad \widehat{\xi}_1 = \frac{1}{2} \ln \nu_1\left(\widehat{\mathsf{T}}'_1\widehat{\mathsf{T}}_1\right),$$

where $w(u)$ and S_M denote a kernel function and a lag truncation parameter, respectively (see [67,68,74]). An example of $w(u)$ is the triangular (Bartlett) kernel given by $w(u) = 1 - |u|$ for $|u| < 1$ and $w(u) = 0$, otherwise. The lag truncation parameter S_M should grow at a rate slower than the rate of M.

The procedure above relies on the asymptotic normality of the Lyapunov exponent estimator. Therefore, if the number of Jacobians, M, is not large, an approximation by the normal distribution may not be appropriate. An alternative approach to computing the standard error is to use the resample methods, such as bootstrapping or subsampling. See [32,35] and [79] for the applications of resampling methods to the evaluation of the global Lyapunov exponent estimates.

Consistent Lag Selection

Performance of the nonparametric Lyapunov exponent estimator is often influenced by the choice of lag length p in the nonlinear AR model when the true lag is not known in practice. To see this point, artificial data is generated from a noisy logistic map with an additive system error given by

$$x_t = ax_{t-1}(1 - x_{t-1}) + \sigma(x_{t-1})\varepsilon_t$$

where $\varepsilon_t \sim$ iid $U(-1/2, 1/2)$ and $\sigma(x_{t-1}) = 0.5 \times \min\{ax_{t-1}(1 - x_{t-1}), 1 - ax_{t-1}(1 - x_{t-1})\}$. Note that the conditional heteroskedasticity function $\sigma(x)$ here ensures that the process x_t is restricted to the unit interval $[0, 1]$. When $a = 4.0$, the system has a positive Lyapunov exponent 0.699. Figure 3 shows an example of a sample path from a deterministic logistic map (left) and a noisy logistic map with the current specification of an error term (right). When $a = 1.5$, the system has a negative Lyapunov exponent -0.699. Table 2 reports the mean and median of nonparametric estimates of Lyapunov exponents using the lags from 1 to 4, $M = T = 50$, based on 1,000 replications.

The simulation results show that overfitting has relatively small effect when the true Lyapunov exponent is positive. On the other hand, in case of negative Lyapunov

Financial Forecasting, Sensitive Dependence, Table 2
Lyapunov exponent estimates when $T = 50$: logistic map

		$p = 1$	$p = 2$	$p = 3$	$p = 4$
Logistic map with $a = 4.0$ (true $\lambda = 0.699$)	Mean	0.694	0.706	0.713	0.720
	Median	0.696	0.704	0.710	0.715
Logistic map with $a = 1.5$ (true $\lambda = -0.699$)	Mean	−0.560	−0.046	0.115	0.179
	Median	−0.661	−0.152	0.060	0.149
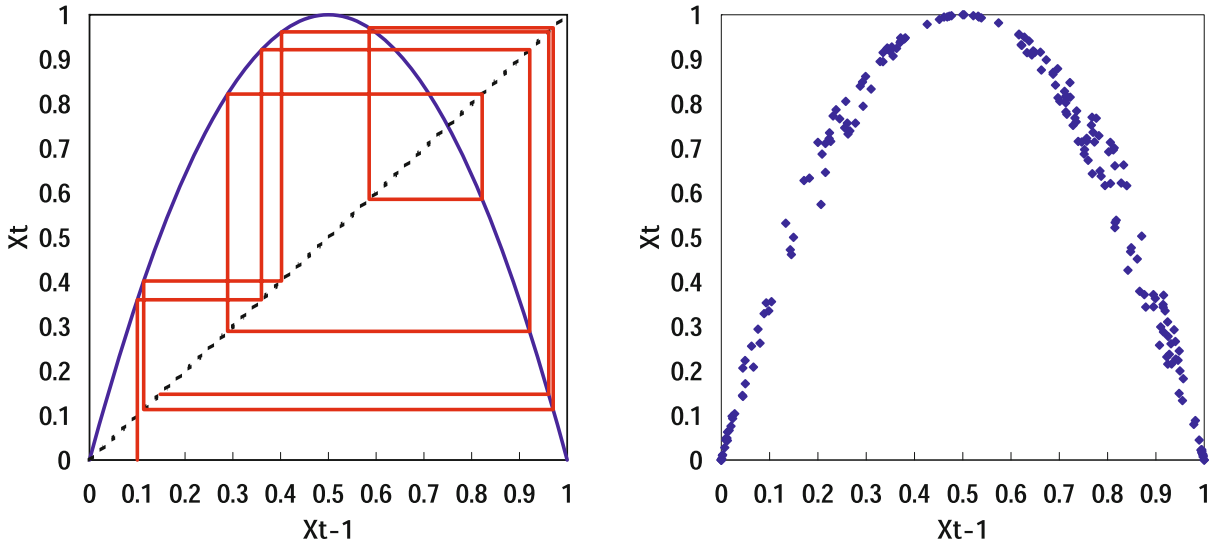
Financial Forecasting, Sensitive Dependence, Figure 3
Logistic map and noisy logistic map

exponent, the upward bias caused by including redundant lags in the nonparametric regression can result in positive Lyapunov exponent estimates. Therefore, when the true lag length of the system is not known, lag selection procedure will be an important part of the analysis of sensitive dependence.

There are several alternative criteria that are designed to select lag length p in the nonparametric kernel autoregressions. With respect to lag selection in the nonparametric analysis of chaos, [15] suggested minimizing the cross-validation (CV) defined by

$$\widehat{CV}(p) = T^{-1} \sum_{t=1}^{T} \left\{ x_t - \widehat{f}_{-(t-1)}(X_{t-1}) \right\}^2 W^2(X_{t-1})$$

where $\widehat{f}_{-(t-1)}(X_{t-1})$ is the leave-one-out estimator evaluated at X_{t-1} and $W^2(\mathbf{x})$ is a weight function. [70] suggested minimizing the nonparametric version of the final prediction error (FPE) defined by

$$\widehat{FPE}(p) = T^{-1} \sum_{t=1}^{T} \left\{ x_t - \widehat{f}(X_{t-1}) \right\}^2 W^2(X_{t-1})$$
$$+ \frac{2}{Th^p} K(0)^p T^{-1} \sum_{t=1}^{T} \{ x_t - \widehat{f}(X_{t-1}) \}^2$$
$$\cdot W^2(X_{t-1}) / \widehat{\pi}(X_{t-1})$$

where $\widehat{\pi}(\mathbf{x})$ is a nonparametric joint density estimator at $\mathbf{x}$. [71] proposed a modification to the FPE to prevent

overfitting in a finite sample with a multiplicative correction term $\{1 + p(T - p + 1)\}^{-4/(p+4)}$. All three nonparametric criteria, the CV, FPE, and the corrected version of the FPE (CFPE) are proved to be consistent lag selection criteria so that the probability of selecting the correct p converges to one as T increases. Table 3 reports frequencies of selected lags based on these criteria among 1,000 iterations.

The simulation results show that all the lag selection criteria perform reasonably well when the data is generated from a noisy logistic map.

While a noisy logistic map has the nonlinear AR(1) form, it should be informative to examine the performance of the procedures when the true process is the AR model of a higher lag order. [15] considered a nonlinear AR(2) model of the form,

$$x_t = 1 - 1.4x_{t-1}^2 + 0.3x_{t-2} + \varepsilon_t$$

Financial Forecasting, Sensitive Dependence, Table 3
Frequencies of selected lags when $T = 50$: logistic map

		$p = 1$	$p = 2$	$p = 3$	$p = 4$
Logistic map with $a = 4.0$ (true $\lambda = 0.699$)	CV	0.989	0.011	0.000	0.000
	FPE	0.998	0.002	0.000	0.000
	CFPE	1.000	0.000	0.000	0.000
Logistic map with $a = 1.5$ (true $\lambda = -0.699$)	CV	0.697	0.168	0.080	0.055
	FPE	0.890	0.085	0.017	0.008
	CFPE	0.989	0.011	0.000	0.000

Financial Forecasting, Sensitive Dependence, Table 4
Lyapunov exponent estimates when $T = 50$: Hénon map

		$p = 1$	$p = 2$	$p = 3$	$p = 4$
Hénon map	Mean	0.411	0.419	0.424	0.431
(true $\lambda = 0.409$)	Median	0.407	0.423	0.427	0.425

Financial Forecasting, Sensitive Dependence, Table 5
Frequencies of selected lags when $T = 50$: Hénon map

		$p = 1$	$p = 2$	$p = 3$	$p = 4$
Hénon map	CV	0.006	0.740	0.250	0.004
(true $\lambda = 0.409$)	FPE	0.028	0.717	0.253	0.002
	CFPE	0.043	0.762	0.194	0.001

where $\varepsilon_t \sim$ iid $U(-0.01, 0.01)$. This is a noisy Hénon map with a positive Lyapunov exponent, $\lambda = 0.409$. Table 4 shows the mean and median of 1,000 nonparametric estimates of Lyapunov exponents using the lags from 1 to 4, $M = T = 50$, when the data is artificially generated from this higher order noisy chaos process.

As in the finding from a chaotic logistic map example, estimates do not seem to be very sensitive to the choice of lags. The results on the lag selection criteria are provided in Table 5.

The table shows that frequencies of selecting the true lag ($p = 2$) becomes less than in the case of the chaotic logistic map in Table 3. However, the performance of CV improves when it is compared to the case of stable logistic map.

The results from this small-scale simulation exercise show that when the true lag length is not known, combining the automatic lag selection method with Lyapunov exponent estimation is recommended in practice.

Other Nonparametric Estimators

In addition to the class of kernel regression estimators, which includes Nadaraya–Watson, local linear or local polynomial estimators, other estimators have also been employed in the estimation of the Lyapunov exponent. With the kernel regression method, Jacobians are evaluated using a local approximation to the nonlinear function at the lagged point X_{t-1}. Another example of the local smoothing method used in Lyapunov exponent estimation is the local thin-plate splines suggested by [51,54]. The local estimation method, however, is subject to the data sparseness problem in the high-dimensional system. Alternatively, Jacobians can be evaluated using a global approximation to the unknown function. As a global estimation method, a global spline function may be used to smooth all the available sample. However, the most frequently used global method in Lyapunov exponent estimation in practice is the neural nets ([2,18,68], among others). A single hidden-layer, feedforward neural network is given by

$$f(X_{t-1}) = \beta_0 + \sum_{j=1}^{k} \beta_j \psi(a_j' X_{t-1} + b_j)$$

where ψ is an activation function (most commonly a logistic distribution function) and k is a number of hidden units. The neural network estimator $\widehat{f}$ can be obtained by minimizing the (nonlinear) least square criterion. Jacobians are then evaluated using the analytical first derivative of neural net function. Compared to other functional approximations, the neural net form is less sensitive to increasing lag length, p. Thus, it has a merit in terms of the effective sample size.

Four Local Measures of Sensitive Dependence

Local Lyapunov Exponent

The global Lyapunov exponent measures the initial value sensitivity of long horizon forecast. For the ergodic and stationary case, this initial value sensitivity measure does not depend on the initial value. By definition, the global Lyapunov exponent is the limit of the local Lyapunov exponent when its order M tends to infinity. Unlike the global Lyapunov exponent, the local Lyapunov exponent is a function of an initial value and thus the initial value sensitivity of the short-term forecast depends on where you are. In this sense, local measures of sensitive dependence contain more detailed information on the predictability in the nonlinear dynamic system.

Recall that both the deterministic tent map and the logistic map with $a = 4.0$ have a common positive Lyapunov exponent 0.693. Thus in terms of long-horizon predictability, two processes have exactly the same degree of initial value sensitivity. Yet, in terms of short term forecast, it is possible that predictability at the same point differs among two processes.

The sign of local Lyapunov exponents of the single process can also be different in some range of initial values. Figure 4 shows the local Lyapunov exponents of the deterministic logistic map with $a = 4.0$ for different values of M. Consistent with the definition, as M grows, it approaches to a flat line at the value of 0.693. However, when M is finite, there is a range of initial values associated with a negative local Lyapunov exponent. Within such a range of initial values, sensitive dependence is low
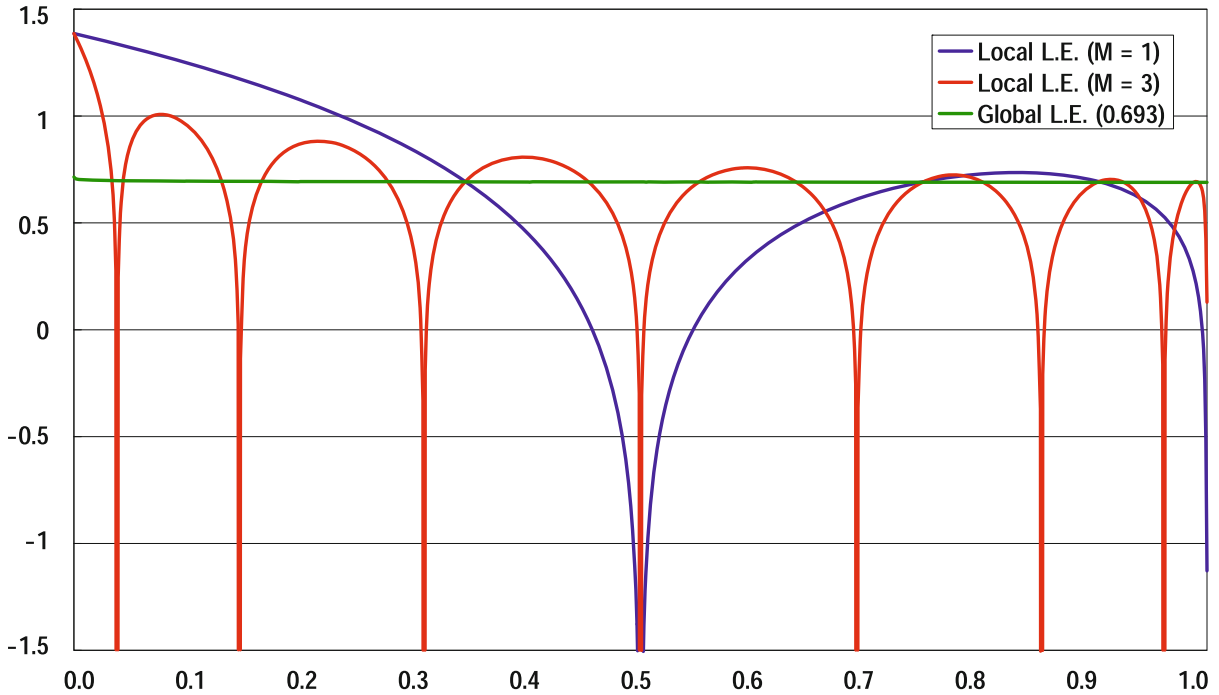
Financial Forecasting, Sensitive Dependence, Figure 4
Local and global Lyapunov exponents of logistic map

and predictability is high even if it is a globally chaotic process.

Analysis of local Lyapunov exponent is also valid in the presence of noise. Studies by [4,48,78], among others, investigate the properties of the local Lyapunov exponent in a noisy system.

The local Lyapunov exponent can be estimated non-parametrically from data using the following procedure. First, obtain the nonparametric Jacobian estimate $\widehat{J}_{t-1}$ for each t using a full sample, as in the case of global Lyapunov exponent estimation. Second, choose a single horizon M of interest. Third, choose the p-dimensional initial value $\mathbf{x} = (x_{t^*}, x_{t^*-1}, \ldots, x_{t^*-p+1})'$ from the data subsequence $\{x_t\}_{t=-p+1}^{T-M}$. Finally, the local Lyapunov exponent estimator at $\mathbf{x}$ is given by $\widehat{\lambda}_M(\mathbf{x}) = (2M)^{-1} \ln \nu_1(\widehat{T}'_M \widehat{T}_M)$ where $\widehat{T}_M = \prod_{t=t^*}^{t^*+M} \widehat{J}_{M-t}$.

While the local Lyapunov exponent is a simple and straightforward local measure of the sensitive dependence, three other useful local measures will be introduced below.

Nonlinear Impulse Response Function

The impulse response function (IRF) is a widely used measure of the persistence effect of shocks in the analysis of economic time series. Here, it is useful to view the IRF as the difference between the two expected future paths: one with and the other without a shock occurred at the current period. When the shock, or the initial deviation, is very small, the notion of impulse responses is thus closely related to the concept of sensitive dependence on initial conditions. To verify this claim, a simple example of a one-dimensional linear IRF is first provided below, followed by the generalization of the IRF to the case of nonlinear time-series model.

For a linear AR model of order one, $x_t = \rho x_{t-1} + \varepsilon_t$, the M-period ahead IRF to a unit shock is defined as

$$\text{IRF}_M = \rho^M.$$

Let $\{x_t^*\}_{t=0}^{\infty}$ be a sample path that contains a single unit shock whereas $\{x_t\}_{t=0}^{\infty}$ is a sample path without any shock. Also let $x_0 = \overline{x}_0$ be an initial condition for the latter path. Then, this linear IRF can be interpreted in two ways. One interpretation is the sequence of the responses to a shock defined to increase one unit of x_0 at time 0 ($x_1 = \rho \overline{x}_0, x_1^* = \rho(\overline{x}_0 + 1), x_1^* - x_1 = \rho, \ldots, x_M^* - x_M = \rho^M$). In this case, the initial value of x_t^* is given as $x_0^* = \overline{x}_0 + 1$, so the shock can be simply viewed as the deviation of two paths at the initial condition. The other interpretation is the sequence of the responses to a shock defined to increase one unit of x_1 at time 1 ($x_1 = \rho \overline{x}_0, x_1^* = \rho \overline{x}_0 + 1$,

$x_1^* - x_1 = 1, \ldots, x_{M+1}^* - x_{M+1} = \rho^M$). In contrast to the first case, two paths have a common initial condition $x_0^* = x_0 = \overline{x}_0$, but the second path is perturbed as if a shock of $\varepsilon_1 = 1$ is realized at time 1 through the dynamic system of $x_t = \rho x_{t-1} + \varepsilon_t$. In either interpretation, however, IRF_M is the difference between x_t^* and x_t at exactly M-period after the shock has occurred and the IRF does not depend on the initial condition $\overline{x}_0$. In addition, the shape of IRF is preserved even if we replace the unit shock with a shock of size δ. The IRF becomes $\rho^M \delta$ and thus the IRF to a unit shock can be considered as a ratio of $\rho^M \delta$ to δ or the normalized IRF.

In the linear framework, the choice between the two interpretations does not matter in practice since the two cases yield exactly the same IRF. However, for nonlinear models, two alternative interpretations lead to different definitions of the IRF. Depending on the objective of the analysis, one may use the former version [31] or the latter version [42,58] of the nonlinear IRFs. The M-period ahead nonlinear impulse response based on the first interpretation considered by [31] is defined as

$$
\begin{aligned}
\text{IRF}_M(\delta, \mathbf{x}) &= E\left[x_{t+M-1} | \mathsf{X}_{t-1} = \mathbf{x}^*\right] \\
&\quad - E\left[x_{t+M-1} | \mathsf{X}_{t-1} = \mathbf{x}\right] \\
&= E\left[x_M | \mathsf{X}_0 = \mathbf{x}^*\right] - E\left[x_M | \mathsf{X}_0 = \mathbf{x}\right] \\
&= f_M(\mathbf{x}^*) - f_M(\mathbf{x})
\end{aligned}
$$

where $\mathsf{X}_{t-1} = (x_{t-1}, \ldots, x_{t-p})'$, $\mathbf{x}^* = (\overline{x}_0 + \delta, \overline{x}_{-1}, \ldots, \overline{x}_{-p+1})'$ and $\mathbf{x} = (\overline{x}_0, \overline{x}_{-1}, \ldots, \overline{x}_{-p+1})'$. Unlike the linear IRF, the nonlinear IRF depends on the size of shock δ and the initial condition (or the history) $\mathsf{X}_0 = \mathbf{x}$. Interestingly, the partial derivative $\Delta f_{M,1}(\mathbf{x}) = \partial f_M(\mathbf{x})/\partial x_{t-1}$ corresponds to normalized IRF (proportional to the nonlinear IRF) for small δ since

$$
\begin{aligned}
&\lim_{\delta \to 0} \frac{\text{IRF}_M(\delta, \mathbf{x})}{\delta} \\
&= \lim_{\delta \to 0} \frac{f_M(\overline{x}_0 + \delta, \overline{x}_{-1}, \ldots, \overline{x}_{-p+1})}{\delta} \\
&\quad - \frac{f(\overline{x}_0, \overline{x}_{-1}, \ldots, \overline{x}_{-p+1})}{\delta} = \Delta f_{M,1}(\mathbf{x}) .
\end{aligned}
$$

In the one-dimensional case, the IRF simplifies to

$$
\begin{aligned}
\text{IRF}_M(\delta, \overline{x}_0) &= E\left[x_{t+M-1} | x_{t-1} = \overline{x}_0 + \delta\right] \\
&\quad - E\left[x_{t+M-1} | x_{t-1} = \overline{x}_0\right] \\
&= E\left[x_M | x_0 = \overline{x}_0 + \delta\right] - E\left[x_M | x_0 = \overline{x}_0\right] \\
&= f_M(\overline{x}_0 + \delta) - f_M(\overline{x}_0) .
\end{aligned}
$$

The first derivative $f_M'(x)$, thus corresponds to the IRF to an infinitesimally small deviation since

$$
\lim_{\delta \to 0} \frac{\text{IRF}_M(\delta, \overline{x}_0)}{\delta} = \lim_{\delta \to 0} \frac{f_M(\overline{x}_0 + \delta) - f(\overline{x}_0)}{\delta} = f_M'(\overline{x}_0).
$$

Recall that $\lambda_M(\overline{x}_0) = M^{-1} \ln | \prod_{t=1}^{M} f'(x_{t-1})|$. If $\prod_{t=1}^{M} f'(x_{t-1})$ can be approximated by $f_M'(\overline{x}_0)$, both normalized IRF and the local Lyapunov exponent contain the same information regarding the initial value sensitivity.

Next, based on the second interpretation, IRF can be alternatively defined as

$$
\begin{aligned}
\text{IRF}_M^*(\delta, \mathbf{x}) &= E\left[x_{t+M-1} | x_t = f(\mathbf{x}) + \delta, \mathsf{X}_{t-1} = \mathbf{x}\right] \\
&\quad - E(x_{t+M-1} | \mathsf{X}_{t-1} = \mathbf{x}) \\
&= E\left[x_M | x_1 = f(\mathbf{x}) + \delta, \mathsf{X}_0 = \mathbf{x}\right] \\
&\quad - E(x_M | \mathsf{X}_0 = \mathbf{x}) \\
&= f_{M-1}(\mathbf{x}^*) - f_M(\mathbf{x})
\end{aligned}
$$

where $\mathsf{X}_{t-1} = (x_{t-1}, \ldots, x_{t-p})'$ and $\mathbf{x} = (\overline{x}_0, \overline{x}_{-1}, \ldots, \overline{x}_{-p+1})'$ and $\mathbf{x}^* = (f(\mathbf{x}) + \delta, \overline{x}_0, \overline{x}_{-1}, \ldots, \overline{x}_{-p+2})'$. This version of nonlinear IRF is sometimes referred to as the generalized impulse response function [42,58]. Using the fact that

$$
f_M(\mathbf{x}) = f_{M-1}\left(f(\mathbf{x}), \overline{x}_0, \overline{x}_{-1}, \ldots, \overline{x}_{-p+2}\right) ,
$$

the equivalence of the partial derivative $\Delta f_{M-1,1}(\mathbf{x}) = \partial f_{M-1}(\mathbf{x})/\partial x_{t-1}$ and the small deviation IRF can be also shown as

$$
\begin{aligned}
&\lim_{\delta \to 0} \frac{\text{IRF}_M^*(\delta, \mathbf{x})}{\delta} \\
&= \lim_{\delta \to 0} \frac{f_{M-1}\left(f(\mathbf{x}) + \delta, \overline{x}_0, \overline{x}_{-1}, \ldots, \overline{x}_{-p+2}\right)}{\delta} \\
&\quad - \frac{f_{M-1}\left(f(\mathbf{x}), \overline{x}_0, \overline{x}_{-1}, \ldots, \overline{x}_{-p+2}\right)}{\delta} \\
&= \Delta f_{M-1,1}(f(\mathbf{x}), \overline{x}_0, \overline{x}_{-1}, \ldots, \overline{x}_{-p+2}) .
\end{aligned}
$$

In the one dimensional case, the IRF formula reduces to

$$
\begin{aligned}
\text{IRF}_M^*(\delta, \overline{x}_0) &= E\left[x_{t+M-1} | x_t = f(\overline{x}_0) + \delta, x_{t-1} = \overline{x}_0\right] \\
&\quad - E(x_{t+M-1} | x_{t-1} = \overline{x}_0) \\
&= E\left[x_M | x_1 = f(\overline{x}_0) + \delta, x_0 = \overline{x}_0\right] \\
&\quad - E(x_M | x_0 = \overline{x}_0) \\
&= E\left[x_{M-1} | x_0 = f(\overline{x}_0) + \delta\right] \\
&\quad - E(x_M | x_0 = \overline{x}_0) \\
&= f_{M-1}(f(\overline{x}_0) + \delta) - f_M(\overline{x}_0) \\
&= f_{M-1}(f(\overline{x}_0) + \delta) - f_{M-1}(f(\overline{x}_0)) .
\end{aligned}
$$

Similarly, the small deviation IRF is given by

$$
\begin{aligned}
\lim_{\delta \to 0} & \frac{\mathrm{IRF}^*_M(\delta, \overline{x}_0)}{\delta} \\
&= \lim_{\delta \to 0} \frac{f_{M-1}(f(\overline{x}_0) + \delta) - f_{M-1}(f(\overline{x}_0))}{\delta} \\
&= f'_{M-1}(f(\overline{x}_0)) \,.
\end{aligned}
$$

The nonlinear impulse response function can be estimated nonparametrically without specifying the functional form by an analogy to Lyapunov exponent estimation (see [72] and [66]). Instead of minimizing $\sum_{t=1}^{T}(x_t - \beta_0 - \beta'_1(X_{t-1} - \mathbf{x}))^2 K_H(X_{t-1} - \mathbf{x})$, the local linear estimator of M-period ahead predictor $f_M(\mathbf{x})$ and its partial derivatives $(\partial f_M)/(\partial x_{t-j})(\mathbf{x})$ for $j = 1, \ldots, p$ can be obtained by minimizing, $\sum_{t=1}^{T-M+1}(x_{t+M-1} - \beta_{M,0} - \beta'_{M,1}(X_{t-1} - \mathbf{x}))^2 K_H(X_{t-1} - \mathbf{x})$, or $\widehat{\beta}_{M,0}(\mathbf{x}) = \widehat{f}_M(\mathbf{x})$ and

$$
\widehat{\beta}_{M,1}(\mathbf{x}) = \begin{bmatrix} \widehat{\beta}_{M,11}(\mathbf{x}) \\ \vdots \\ \widehat{\beta}_{M,1p}(\mathbf{x}) \end{bmatrix} = \begin{bmatrix} \Delta\widehat{f}_{M,1}(\mathbf{x}) \\ \vdots \\ \Delta\widehat{f}_{M,p}(\mathbf{x}) \end{bmatrix},
$$

respectively, where $\widehat{\beta}_M(x) = (\widehat{\beta}_{M,0}(\mathbf{x}), \widehat{\beta}_{M,1}(\mathbf{x})')' = (X'_x \mathbf{W}_x X_x)^{-1} X'_x \mathbf{W}_x \mathbf{Y}$,

$$
X_x = \begin{bmatrix} 1 & (X_0 - \mathbf{x})' \\ \vdots & \vdots \\ 1 & (X_{T-M} - \mathbf{x})' \end{bmatrix},
$$

$\mathbf{Y} = (x_M, \ldots, x_T)'$ and $W_x = \mathrm{diag}\{K_H(X_0 - \mathbf{x}), \ldots, K_H(X_{T-M} - \mathbf{x})\}$. The local linear estimator of the IRF is then given by

$$
\widehat{\mathrm{IRF}}_M(\delta, \mathbf{x}) = \widehat{f}_M(\mathbf{x}^*) - \widehat{f}_M(\mathbf{x})
$$

where $\mathbf{x}^* = (\overline{x}_0 + \delta, \overline{x}_{-1}, \ldots, \overline{x}_{-p+1})'$ and $\mathbf{x} = (\overline{x}_0, \overline{x}_{-1}, \ldots, \overline{x}_{-p+1})'$. Similarly, the estimator of the alternative IRF is given by

$$
\widehat{\mathrm{IRF}}^*_M(\delta, \mathbf{x}) = \widehat{f}_{M-1}(\mathbf{x}^*) - \widehat{f}_M(\mathbf{x})
$$

where $\mathbf{x}^* = (\widehat{f}(\mathbf{x}) + \delta, \overline{x}_0, \overline{x}_{-1}, \ldots, \overline{x}_{-p+2})'$ and $\mathbf{x} = (\overline{x}_0, \overline{x}_{-1}, \ldots, \overline{x}_{-p+1})'$. When $\mathbf{x}$ and δ are given, computing nonparametric IRFs for a sequence of M provide a useful information on the persistence of deviation without specifying the autoregressive function. However, instead of reporting IRFs for many possible combinations of $\mathbf{x}$ and δ, one can also compute the small deviation IRF based on the nonparametric estimate of the first partial derivative at $\mathbf{x}$. The local linear estimator of the small deviation IRF is given by $\Delta\widehat{f}_{M,1}(\mathbf{x})$ for the first version, and

$\Delta\widehat{f}_{M-1,1}(\widehat{f}(\mathbf{x}), \overline{x}_0, \overline{x}_{-1}, \ldots, \overline{x}_{-p+2})$ for the second version, respectively. A large change in the value of derivatives with increasing M represents the sensitive dependence on initial conditions.

Yao and Tong's Variance Decomposition

The initial value sensitivity of the system with dynamic noise also has an implication in the presence of additional observation noise. Suppose that current observation is subject to a measurement error, a rounding error, or when only preliminary estimates of aggregate economic variables announced by the statistical agency are available. When the true current position deviates slightly from $\mathbf{x} = (\overline{x}_0, \overline{x}_{-1}, \ldots, \overline{x}_{-p+1})'$ by $\boldsymbol{\delta} = (\delta_1, \ldots, \delta_p)'$, the performance of the same predictor may be measured by $E[\{x_{t+M} - f_M(\mathbf{x})\}^2 | X_t = \mathbf{x} + \boldsymbol{\delta}]$. Under a certain condition, this MSFE can be decomposed as follows:

$$
\begin{aligned}
E&\left[\{x_{t+M} - f_M(\mathbf{x})\}^2 | X_t = \mathbf{x} + \boldsymbol{\delta}\right] \\
&= \sigma^2_M(\mathbf{x} + \boldsymbol{\delta}) + \{f_M(\mathbf{x} + \boldsymbol{\delta}) - f_M(\mathbf{x})\}^2 \\
&= \sigma^2_M(\mathbf{x} + \boldsymbol{\delta}) + \{\boldsymbol{\delta}' \Delta f_M(\mathbf{x})\}^2 + o(||\boldsymbol{\delta}||^2)
\end{aligned}
$$

where $\Delta f_M(\mathbf{x}) = (\Delta f_{M,1}(\mathbf{x}), \ldots, \Delta f_{M,p}(\mathbf{x}))'$, $\Delta f_{M,j}(\mathbf{x}) = (\partial f_M)/(\partial x_{t-j})(\mathbf{x})$ for $j = 1, \ldots, p$. This decomposition shows two dominant components in the MSFE. The first component represents the prediction error caused by the randomness in the system at point $\mathbf{x} + \boldsymbol{\delta}$. This component will be absent in the case where there is no dynamic noise ε_t in the system. The second component represents the difference caused by the deviation $\boldsymbol{\delta}$ from the initial point $\mathbf{x}$. When the non-zero deviation $\boldsymbol{\delta}$ is much smaller than σ, the standard deviation of ε_t, the first component $\sigma^2_M(\mathbf{x} + \boldsymbol{\delta}) = \sigma^2_M(\mathbf{x}) + O(||\boldsymbol{\delta}||)$ is the dominant term because the second component $\{\boldsymbol{\delta}' \Delta f_M(\mathbf{x})\}^2$ is of order $O(||\boldsymbol{\delta}||^2)$. However, for a nonlinear system with a very small error ε_t, the contribution of the second term can become nonnegligible. Thus, [80] considered $\Delta f_M(\mathbf{x})$ as a measure of sensitivity to initial conditions for the M-period ahead forecast (They referred to the M-step Lyapunov-like index).

If $f_M(\mathbf{x})$ is replaced by a mean square consistent estimator $\widehat{f}_M(\mathbf{x})$, such as a local linear estimator $\widehat{\beta}_{M,0}(\mathbf{x})$,

$$
\begin{aligned}
\lim_{T \to \infty} E&\left[\{x_{T+M} - \widehat{f}_M(\mathbf{x})\}^2 | X_T = \mathbf{x} + \boldsymbol{\delta}\right] \\
&= \sigma^2_M(\mathbf{x} + \boldsymbol{\delta}) + \{\boldsymbol{\delta}' \Delta f_M(\mathbf{x})\}^2 + o(||\boldsymbol{\delta}||^2) \,.
\end{aligned}
$$

Thus the decomposition is still valid. For the estimation of the sensitivity measure $\Delta f_M(\mathbf{x})$, the local linear estimator $\widehat{\beta}_{M,0}(\mathbf{x}) = \Delta\widehat{f}_M(\mathbf{x})$ can be used. In practice, it

is convenient to consider a norm version of the measure $\mathrm{LI}_M(\mathbf{x}) = ||\Delta f_M(\mathbf{x})||$ and report its estimator

$$\widehat{\mathrm{LI}}_M(\mathbf{x}) = ||\widehat{\Delta f}_M(\mathbf{x})||$$

evaluated at various $\mathbf{x}$. In a one-dimensional case, they are $\mathrm{LI}_M(\overline{x}_0) = |f'_M(\overline{x}_0)|$ and $\widehat{\mathrm{LI}}_M(\overline{x}_0) = |\widehat{f}'_M(\overline{x}_0)|$, respectively. Note that $\mathrm{LI}_M(\mathbf{x})$ is related to the derivative of the normalized nonlinear impulse response function $\mathrm{IRF}_M(\delta, \mathbf{x})$. Recall that, in the one-dimensional case, a normalized IRF to infinitesimal shocks becomes the first derivative. Thus, $\mathrm{LI}_M(\overline{x}_0)$ is the absolute value of the estimator of the corresponding IRF. In the multidimensional case, IRF to small shocks becomes the partial derivative with respect to the first components. If shocks are also given to other initial values in IRF, computing the norm of the estimator of all IRFs yields $\mathrm{LI}_M(\mathbf{x})$.

This sensitivity measure is also related to the local Lyapunov exponent. In the one-dimensional case, with a fixed M, the local Lyapunov exponent can be written as $\lambda_M(\overline{x}_0) = M^{-1} \ln |\prod_{t=1}^{M} f'(x_{t-1})|$. If the contribution of ε_t is very small, $\mathrm{d}f^{(M)}(x_0)/\mathrm{d}x \approx \prod_{t=1}^{M} f'(x_{t-1})$ and then the estimator $\widehat{f}'_M(x_0)$ becomes an estimator of $\mathrm{d}f^{(M)}(x_0)/\mathrm{d}x$. Thus $\lambda_M(\overline{x}_0)$ can be also estimated by $M^{-1} \ln \widehat{\mathrm{LI}}_M(\overline{x}_0)$.

Information Matrix

The last measure of the initial value sensitivity is the one based on the distance between two distributions of M-steps ahead forecast, conditional on two nearby initial values $\mathbf{x} = (\overline{x}_0, \overline{x}_{-1}, \ldots, \overline{x}_{-p+1})'$ and $\mathbf{x} + \boldsymbol{\delta}$ where $\boldsymbol{\delta} = (\delta_1, \ldots, \delta_p)'$. Let $\pi_M(y|\mathbf{x})$ and $\Delta \pi_M(y|\mathbf{x})$ be the conditional density function of x_M given $\mathsf{X}_0 = \mathbf{x}$ and a $p \times 1$ vector of its partial derivatives. [81] suggested using Kullback–Leibler information to measure the distance, which is given by

$$K_M(\delta, \mathbf{x}) = \int_{-\infty}^{+\infty} \{\pi_M(y|\mathbf{x} + \boldsymbol{\delta}) - \pi_M(y|\mathbf{x})\}$$
$$\cdot \ln \{\pi_M(y|\mathbf{x} + \boldsymbol{\delta})/\pi_M(y|\mathbf{x})\} \mathrm{d}y .$$

Assuming the smoothness of conditional distribution and interchangeability of integration and differentiation, Taylor series expansion around $\mathbf{x}$ for small $\boldsymbol{\delta}$ yields

$$K_M(\delta, \mathbf{x}) = \boldsymbol{\delta}' I_M(\mathbf{x})\boldsymbol{\delta} + o(||\boldsymbol{\delta}||^2)$$

where

$$I_M(\mathbf{x}) = \int_{-\infty}^{+\infty} \Delta \pi_M(y|\mathbf{x}) \Delta \pi_M(y|\mathbf{x})'/\pi_M(y|\mathbf{x}) \mathrm{d}y .$$

If initial value $\mathbf{x}$ is treated as a parameter vector of the distribution, $I_M(\mathbf{x})$ is the Fisher's information matrix, which represents the information on $\mathbf{x}$ contained in x_M. This quantity can be used as an initial value sensitivity measure since more information on $\mathbf{x}$ implies more sensitivity of distribution of x_M to the initial condition $\mathbf{x}$. This information matrix measure and the M-step Lyapunov-like index are related via the following inequality when the system is one-dimensional,

$$I_M(\overline{x}_0) \geq \frac{\mathrm{LI}_M^2(\overline{x}_0)}{\sigma_M^2(\overline{x}_0)} .$$

Thus, for a given M-step Lyapunov-like index, a larger conditional MSFE implies more sensitivity. In addition, because that $\lambda_M(\overline{x}_0) \approx M^{-1} \ln \mathrm{LI}_M(\overline{x}_0)$ and $\sigma_M^2(\overline{x}_0) \approx \sigma^2[1 + \sum_{j=1}^{M-1} \exp\{2(M-j)\lambda_{M-j}(f^{(j)}(\overline{x}_0))\}]$,

$$\ln I_M(\overline{x}_0) \geq 2M\lambda_M(\overline{x}_0)$$
$$- \ln \left[1 + \sum_{j=1}^{M-1} \exp \left\{ 2(M-j)\lambda_{M-j}(f^{(j)}(\overline{x}_0)) \right\} \right] - \ln \sigma^2$$

holds approximately.

As an alternative to Kullback–Leibler distance, [28] considered L_2-distance given by

$$D_M(\delta, \mathbf{x}) = \int_{-\infty}^{+\infty} \{\pi_M(y|\mathbf{x} + \boldsymbol{\delta}) - \pi_M(y|\mathbf{x})\}^2 \mathrm{d}y .$$

Because of a similar argument, for small $\boldsymbol{\delta}$, $D_M(\delta, \mathbf{x})$ can be approximated by

$$D_M(\delta, \mathbf{x}) = \boldsymbol{\delta}' J_M(\mathbf{x})\boldsymbol{\delta} + o(||\boldsymbol{\delta}||^2)$$

where

$$J_M(\mathbf{x}) = \int_{-\infty}^{+\infty} \Delta \pi_M(y|\mathbf{x}) \Delta \pi_M(y|\mathbf{x})' \mathrm{d}y .$$

Note that $J_M(\mathbf{x})$ cannot be interpreted as Fisher's information but can still be used as a sensitivity measure.

Both $I_M(\mathbf{x})$ and $J_M(\mathbf{x})$ can be estimated nonparametrically. Consider the minimization problem of $\sum_{t=1}^{T-M+1} (\kappa_h(x_{t+M-1} - y) - \beta_{M,0} - \beta'_{M,1}(\mathsf{X}_{t-1} - \mathbf{x}))^2 K_H(\mathsf{X}_{t-1} - \mathbf{x})$ where $\kappa_h(u) = \kappa(u/h)/h$, h is the bandwidth and κ is a univariate kernel function, instead of minimizing $\sum_{t=1}^{T-M+1} (x_{t+M-1} - \beta_{M,0} - \beta'_{M,1}(\mathsf{X}_{t-1} - \mathbf{x}))^2 K_H(\mathsf{X}_{t-1} - \mathbf{x})$. Then, $\widehat{\beta}_{M,0}(\mathbf{x}, y) = \widehat{\pi}_M(y|\mathbf{x})$ and $\widehat{\beta}_{M,1}(\mathbf{x}, y) = \Delta\widehat{\pi}_M(y|\mathbf{x})$, where $\widehat{\beta}_M(\mathbf{x}, y) = (\widehat{\beta}_{M,0}(\mathbf{x}, y), \widehat{\beta}_{M,1}(\mathbf{x}, y)')' = (\mathsf{X}'_x \mathbf{W}_x \mathsf{X}_x)^{-1} \mathsf{X}'_x \mathbf{W}_x \mathbf{Y}_y$,

$$\mathsf{X}_x = \begin{bmatrix} 1 & (\mathsf{X}_0 - \mathbf{x})' \\ \vdots & \vdots \\ 1 & (\mathsf{X}_{T-M} - \mathbf{x})' \end{bmatrix} ,$$

$\mathbf{Y}_y = \{\kappa_h(x_M - y), \ldots, \kappa_h(x_T - y)\}'$ and $W_x = \text{diag}\{K_H(\mathsf{X}_0 - \mathbf{x}), \ldots, K_H(\mathsf{X}_{T-M} - \mathbf{x})\}$. Then the estimators of $I_M(\mathbf{x})$ and $J_M(\mathbf{x})$ are given by

$$\widehat{I}_M(\mathbf{x}) = \int_{-\infty}^{+\infty} \Delta\widehat{\pi}_M(y|\mathbf{x})\Delta\widehat{\pi}_M(y|\mathbf{x})'/\widehat{\pi}_M(y|\mathbf{x})\mathrm{d}y$$

and

$$\widehat{J}_M(\mathbf{x}) = \int_{-\infty}^{+\infty} \Delta\widehat{\pi}_M(y|\mathbf{x})\Delta\widehat{\pi}_M(y|\mathbf{x})'\mathrm{d}y\,,$$

respectively.

Forecasting Financial Asset Returns and Sensitive Dependence

Nonlinear Forecasting of Asset Returns

In this subsection, a quick review of the general issues of forecasting financial asset returns is first provided, then the empirical results on the nonlinear forecasting based on nonparametric methods are summarized.

In the past, the random walk model was considered as the most appropriate model to describe the dynamics of asset prices in practice (see [24]). However, after decades of investigation, more evidence on some predictable components of asset returns has been documented in the literature. Although the evidence is often not very strong, several studies report the positive serial dependence for relatively short horizon stock returns. For example, [47] show that first-order autocorrelation of weekly returns on the Center for Research in Security Prices (CRSP) index is as high as 30 percent and significant when an equal-weighted index is used, but is somewhat less when a value-weighted index is used ([12] provide similar evidence for the daily return). The conditional mean of stock returns may not depend only on the past returns but also on other economic variables, including dividend yields, price earnings ratio, short and long interest rates, industrial production and inflation rate. A comprehensive statistical analysis to evaluate the 1-month-ahead out-of-sample forecast of 1 month excess returns by these predictors is conducted by [55]. Some results on the long-horizon predictability in stock returns, based on lagged returns (e. g., [25] and [57]) and other economic variables such as dividend yields or dividend-price ratios (e. g., [26] and [13]) are also available. This evidence on long-horizon forecasts, however, is still controversial because the standard statistical inference procedure may not be reliable in case when the correlation coefficient is computed from a small number of nonoverlapping observations [62] or when the predictor is very persistent in the forecasting regression [73].

The question is whether the introduction of nonlinear structure helps improve the forecasting performance of future asset returns. When the nonlinear condition mean function is unspecified, the neural network method has often been employed as a reliable nonparametric method in predicting the returns. For IBM daily stock returns, [75] found no improvement in out-of-sample predictability based on the neural network model. For daily returns of the Dow Jones Industrial Average (DJIA) index, [33] estimated a nonlinear AR model using the same method. He, in contrast, showed that MSFE reduction over a benchmark linear AR model could be as large as 12.3 percent for the 10-day-ahead out-of-sample forecast. The role of economic fundamentals as predictors can be also investigated under the nonlinear framework. Using a model selection procedure similar to the one employed by [55], some evidence of MSFE improvement from neural network-based forecast of excess returns was provided in [60] and [59] but no encouraging evidence was found in similar studies by [61] and [49]. In practice, 'noise traders' or 'chartists' may predict prices using some technical trading rules (TTRs) rather than using economic fundamentals. For example, a simple TTR based on the moving average can generate a buy signal when the current asset price level P_t is above $n^{-1}\sum_{i=1}^{n} P_{t-i+1}$ for some positive integer n and a sell signal when it is below. [11] found some evidence on the nonlinearity in the conditional mean of DJIA returns conditional on buy-sell signals. [33] further considered including past buy-sell signals as predictors in the neural network model and found that an improvement in MSFE over the linear AR model was even larger than the case when only lagged returns are used as a predictor in the neural network model. One useful nonlinear model is the functional coefficient AR model where the AR coefficient can depend on time or some variables. For example, as in [39], the AR coefficient can be a function of buy-sell signals. [44] claimed that a functional coefficient AR model with a coefficient as a function of the moving average of squared returns well described the serial correlation feature of stock returns.

This moderate but increasing evidence of nonlinear forecastability applies not only to the stock market but also to the foreign exchange market. In the past, [52] could not find any reasonable linear model that could out-perform the random walk model in an out-of-sample forecast of foreign exchange rates. The nonlinear AR model was estimated nonparametrically by [19] but no forecasting improvement over the random walk model could be found in their analysis. However, many follow-up studies, including [34,39,43,76], provided some evidence on forecastability with nonlinear AR models es-

timated using neural networks or other nonparametric methods.

One important and robust empirical fact is that much higher positive serial correlation is typically observed for the volatility measures such as the absolute returns, $|x_t|$, and their power transformation, $|x_t|^\alpha$ for $\alpha > 0$, than for the returns, x_t ([20,69]). This observation is often referred to as a volatility clustering. As a result, forecasting volatility has been much more successful than forecasting returns themselves. The most commonly used approach in forecasting volatility is to describe the conditional variance of asset returns using the class of ARCH and GARCH models ([8,23]). The volatility of stock returns is also known to respond more strongly to negative shocks in returns than positive ones. This 'leverage effect' often motivates the introduction of nonlinear structure in volatility modeling such as the EGARCH model of [53]. Instead of estimating the unknown parameter in a specified ARCH model, the nonparametric method can be also employed to estimate the possibly nonlinear ARCH model in forecasting (see [46]). The better forecastability of market direction (or market timing), sign (x_t), than that of returns, has also been documented in the literature. Examples are [55] for the stock market and [39,43], and [16] for the foreign exchange market. Since the return, x_t, can be decomposed into a product of the two components, $|x_t| \times \text{sign}(x_t)$, one may think the strong linear or nonlinear forecastability of the volatility and the sign of returns should lead to forecastability of the returns as well. Interestingly, however, [17] theoretically showed that the serial dependence of asset return volatilities and that of return signs did not necessarily imply the serial dependence of returns.

In summary, a growing number of recent studies show some evidence of linear and nonlinear forecastability of asset returns, and stronger evidence of forecastability of their nonlinear transformations, such as the squared returns, absolute returns and the sign of returns. In this sense, the nonlinearity seems to be playing a non-negligible role in explaining the dynamic behavior of asset prices.

Initial Value Sensitivity in Financial Data

Theoretically, when investors have heterogeneous expectations about the future prices, asset price dynamics can be chaotic with a positive Lyapunov exponent [10]. A comprehensive list on earlier empirical work related to the sensitive dependence and chaos in financial data is provided in [2,6]. Many early studies employed either the BDS test or a dimension estimator and provided the indirect evidence on sensitive dependence and chaos. For

example, [64] applied the BDS test to weekly returns on the value-weighted CRSP portfolio and rejected iid randomness. [41] further examined weekly value-weighted and equally weighted CRSP portfolio returns, as well as Standard & Poor 500 (S&P 500) index returns for various frequencies, and found strong evidence against iid. Similar findings are also reported for the daily foreign exchange rate returns in [40]. For financial variables, high-frequency data or tick data is often available to researchers. Earlier examples of studies on chaos using high-frequency data include [50], who found some evidence of low-dimensional chaos based on the correlation dimension and K_2 entropy of 20-second S&P 500 index returns, with a number of observations as large as 19,027. Estimation results on Lyapunov exponents for high-frequency stock returns are also available. In addition to the BDS test, [1] and [2] employed the neural network method and found negative Lyapunov exponents in 1- and 5-minute returns of cash series of S&P 500, UK Financial Times Stock Exchange-100 (FTSE-100) index, Deutscher Aktienindex (DAX), the Nikkei 225 Stock Average, and of futures series of S&P 500 and FTSE-100. Using the resampling procedure of [32], [65] obtained significantly negative Lyapunov exponents for daily stock returns of the Austrian Traded Index (ATX). For the foreign exchange market, [18] estimated Lyapunov exponents of the Canadian, German, Italian and Japanese monthly spot exchange rates using neural nets and found some mixed result regarding their sign.

By using the absolute returns or their power transformation instead of using returns themselves, sensitive dependence of volatility on initial conditions may be examined nonparametrically. [68] used neural nets and estimated Lyapunov exponents of higher order daily returns of the DJIA index. Figure 5 shows their global Lyapunov exponent estimates for simple returns, squared returns and absolute returns. For all cases, Lyapunov exponents are significantly negative but the values of absolute returns are always larger than that of simple returns. While some estimates are close to zero, the observation of the monotonically increasing Lyapunov exponent with increasing p, for daily and absolute returns, resembles the simulation results of the previous section implying the upward bias when the true Lyapunov exponent is negative.

For the exchange rate market, [29] applied [63]'s method to absolute changes and their power transformation of Canadian and German nominal exchange rates and did not reject the null hypothesis of chaos.

For a local measure of initial value sensitivity, [68] also reported the median values of 145 estimates of local Lyapunov exponents for DJIA returns, in addition
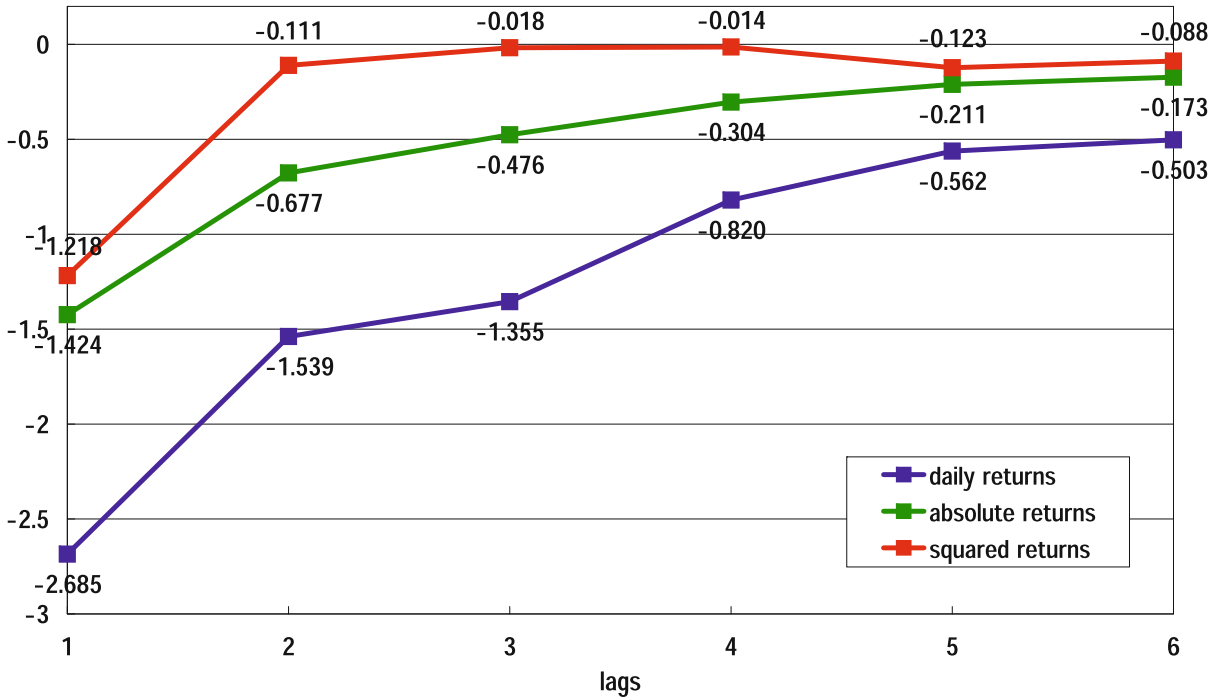
Financial Forecasting, Sensitive Dependence, Figure 5
Global Lyapunov exponents of stock returns

to the global Lyapunov exponents. [45] reported the nonlinear impulse response functions of yen/dollar and deutschemark/dollar exchange rate returns based on parametrically estimated GARCH model. [14] reported a Lyapunov-like index of [80] for the simple returns and absolute returns of CRSP data used in [64]. From Fig. 6, they concluded that (i) the first half of the CRSP series is more volatile than its second half, suggesting that the market
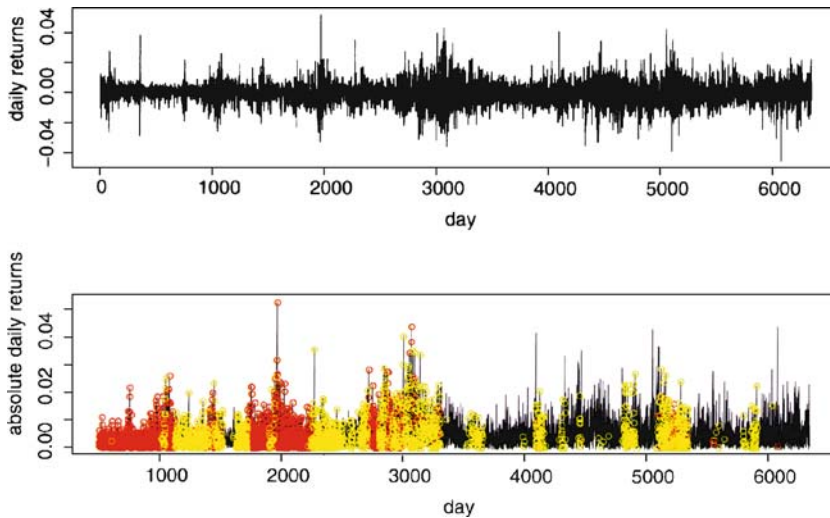
Financial Forecasting, Sensitive Dependence, Figure 6
Upper panel displays the time series plot of the CRSP daily returns. *Lower* panel shows the absolute CRSP daily returns with data coloured *red* whenever their Lyapunov-like indices are above the third quartile of the indices, and data coloured *yellow* if their indices are between the median and the third quartile

becomes more mature with time, and (ii) volatile periods tend to form clusters. [65] reported the information matrix measure of local sensitive dependence computed from ATX data based on the parametric estimation of ARCH and GARCH models, in addition to nonparametric estimates of the global Lyapunov exponent.

On the whole, empirical studies on global and local sensitivity measures suggested less sensitive dependence than the chaotic model would predict, but some sensitivity of short-term forecastability on initial conditions.

Future Directions

Some of the possible directions of future research topics are in order.

The first direction is to search for the economic theory behind the initial value sensitivity if detected in the data. The statistical procedures introduced here are basically data description and the empirical results obtained by this approach are not directly connected to underlying economic or finance theory. Theories, such as the one developed by [10], can predict complex behavior of asset prices but direct estimation of the model are typically not possible. Thus, for most cases, the model is evaluated by matching the actual data with the one generated from the model in simulation. Thus direct implication to the sensitive dependence measure would provide a more convincing argument for the importance of knowing the structure. [38] may be considered as one attempt in this direction.

The second direction is to develop better procedures in estimating the initial value sensitivity with the improved accuracy in the environment of a relatively small sample size. In the Jacobian method of estimating the Lyapunov exponent, the conditional mean function has been estimated either parametrically and nonparametrically. A fully nonparametric approach, however, is known to suffer from a high dimensionality problem. A semiparametric approach, such as the one for an additive AR model, is likely to be useful in this context but has not been used in the initial value sensitivity estimation.

The third direction is towards further analysis based on high-frequency data, which has become more commonly available in empirical finance. Much progress has been made in the statistical theory on the realized volatility computed from such data, and forecasting volatility of asset returns based on the realized volatility has been empirically successful (see, e. g., [3]). However, so far, this approach has not been used in detecting the initial value sensitivity in volatility. In addition, realized volatility is known to suffer from market microstructure noise

when sampling frequency increases. Given the fact that the initial value sensitivity measures can be considered in the framework of the nonlinear AR models, namely, the stochastic environment in the presence of noise, it is of interest in investigating the robustness of the procedure to the market microstructure noise when applied to high-frequency returns.

Bibliography

1. Abhyankar A, Copeland LS, Wong W (1995) Nonlinear dynamics in real-time equity market indices: evidence from the United Kingdom. Econ J 105:864–880
2. Abhyankar A, Copeland LS, Wong W (1997) Uncovering nonlinear structure in real-time stock-market indexes: The S&P 500, the DAX, the Nikkei 225, and the FTSE-100. J Bus Econ Stat 15:1–14
3. Andersen TB, Bollerslev T, Diebold FX, Labys P (2003) Modeling and forecasting realized volatility. Econometrica 71:579–625
4. Bailey BA, Ellner S, Nychka DW (1997) Chaos with confidence: Asymptotics and applications of local Lyapunov exponents. In: Cutler CD, Kaplan DT (eds) Fields Institute Communications, vol 11. American Mathematical Society, Providence, pp 115–133
5. Barnett WA, Gallant AR, Hinich MJ, Jungeilges J, Kaplan D, Jensen MJ (1995) Robustness of nonlinearity and chaos tests to measurement error, inference method, and sample size. J Econ Behav Organ 27:301–320
6. Barnett WA, Serletis A (2000) Martingales, nonlinearity, and chaos. J Econ Dyn Control 24:703–724
7. Bask M, de Luna X (2002) Characterizing the degree of stability of non-linear dynamic models. Stud Nonlinear Dyn Econom 6:3
8. Bollerslev T (1986) Generalized autoregressive conditional heteroskedasticity. J Econ 31:307–327
9. Brock WA, Dechert WD, Scheinkman JA, LeBaron B (1996) A test for independence based on the correlation dimension. Econ Rev 15(3):197–235
10. Brock WA, Hommes CH (1998) Heterogeneous beliefs and routes to chaos in a simple asset pricing model. J Econ Dyn Control 22:1235–1274
11. Brock WA, Lakonishok J, LeBaron B (1992) Simple technical trading rules and the stochastic properties of stock returns. J Financ 47:1731–1764
12. Campbell JY, Lo AW, MacKinlay AC (1997) The Econometrics of Financial Markets. Princeton University Press, Princeton
13. Campbell JY, Shiller R (1988) The dividend-price ratio and expectations of future dividends and discount factors. Rev Financ Stud 1:195–228
14. Chan KS, Tong H (2001) Chaos: A Statistical Perspective. Springer, New York
15. Cheng B, Tong H (1992) On consistent nonparametric order determination and chaos. J Royal Stat Soc B 54:427–449
16. Cheung YW, Chinn MD, Pascual AG (2005) Empirical exchange rate models of the nineties: Are any fit to survive? J Int Money Financ 24:1150–1175
17. Christoffersen PF, Diebold FX (2006) Financial asset returns, direction-of-change forecasting, and volatility dynamics. Management Sci 52:1273–1287

18. Dechert WD, Gençay R (1992) Lyapunov exponents as a non-parametric diagnostic for stability analysis. J Appl Econom 7:S41–S60

19. Diebold FX, Nason JA (1990) Nonparametric exchange rate prediction? J Int Econ 28:315–332

20. Ding Z, Granger CWJ, Engle RF (1993) A long memory property of stock market returns and a new model. J Empir Financ 1:83–106

21. Eckmann JP, Kamphorst SO, Ruelle D, Ciliberto S (1986) Liapunov exponents from time series. Phys Rev A 34:4971–4979

22. Eckmann JP, Ruelle D (1985) Ergodic theory of chaos and strange attractors. Rev Mod Phys 57:617–656

23. Engle RF (1982) Autoregressive conditional heteroskedasticity with estimates of the variance of UK inflation. Econometrica 50:987–1008

24. Fama E (1970) Efficient capital markets: Review of theory and empirical work. J Financ 25:383–417

25. Fama E, French K (1988) Permanent and temporary components of stock prices. J Political Econ 96:246–273

26. Fama E, French K (1988) Dividend yields and expected stock returns. J Financ Econ 22:3–5

27. Fan J, Yao Q (2003) Nonlinear Time Series: Nonparametric and Parametric Methods. Springer, New York

28. Fan J, Yao Q, Tong H (1996) Estimation of conditional densities and sensitivity measures in nonlinear dynamical systems. Biometrika 83:189–206

29. Fernándes-Rodríguez F, Sosvilla-Rivero S, Andrada-Félix J (2005) Testing chaotic dynamics via Lyapunov exponents. J Appl Econom 20:911–930

30. Frank M, Stengos T (1989) Measuring the strangeness of gold and silver rates of return. Rev Econ Stud 56:553–567

31. Gallant AR, Rossi PE, Tauchen G (1993) Nonlinear dynamic structures. Econometrica 61:871–907

32. Gençay R (1996) A statistical framework for testing chaotic dynamics via Lyapunov exponents. Physica D 89:261–266

33. Gençay R (1998) The predictability of security returns with simple technical trading rules. J Empir Financ 5:347–359

34. Gençay R (1999) Linear, non-linear and essential foreign exchange rate prediction with simple technical trading rules. J Int Econ 47:91–107

35. Giannerini S, Rosa R (2001) New resampling method to assess the accuracy of the maximal Lyapunov exponent estimation. Physica D 155:101–111

36. Grassberger P, Procaccia I (1983) Estimation of the Kolmogorov entropy from a chaotic signal. Phys Rev A 28:2591–2593

37. Hall P, Wolff RCL (1995) Properties of invariant distributions and Lyapunov exponents for chaotic logistic maps. J Royal Stat Soc B 57:439–452

38. Hommes CH, Manzan S (2006) Comments on Testing for nonlinear structure and chaos in economic time series. J Macroecon 28:169–174

39. Hong Y, Lee TH (2003) Inference on predictability of foreign exchange rates via generalized spectrum and nonlinear time series models. Rev Econ Stat 85:1048–1062

40. Hsieh DA (1989) Testing for nonlinear dependence in daily foreign exchange rates. J Bus 62:339–368

41. Hsieh DA (1991) Chaos and nonlinear dynamics: application to financial markets. J Financ 46:1839–1877

42. Koop G, Pesaran MH, Potter SM (1996) Impulse response analysis in nonlinear multivariate models. J Econom 74:119–147

43. Kuan CM, Liu T (1995) Forecasting exchange rates using feedforward and recurrent neural networks. J Appl Econom 10:347–364

44. LeBaron B (1992) Some relation between the volatility and serial correlations in stock market returns. J Bus 65:199–219

45. Lin WL (1997) Impulse response function for conditional volatility in GARCH models. J Bus Econ Stat 15:15–25

46. Linton OB (2008) Semiparametric and nonparametric ARCH modelling. In: Anderson TG, Davis RA, Kreiss JP, Mikosch T (ed) Handbook of Financial Time Series. Springer, Berlin

47. Lo AW, MacKinlay AC (1988) Stock market prices do not follow random walks: evidence from a simple specification test. Rev Financ Stud 1:41–66

48. Lu ZQ, Smith RL (1997) Estimating local Lyapunov exponents. In: Cutler CD, Kaplan DT (eds) Fields Institute Communications, vol 11. American Mathematical Society, Providence, pp 135–151

49. Maasoumi E, Racine J (2002) Entropy and predictability of stock market returns. J Econom 107:291–312

50. Mayfield ES, Mizrach B (1992) On determining the dimension of real time stock price data. J Bus Econ Stat 10:367–374

51. McCaffrey DF, Ellner S, Gallant AR, Nychka DW (1992) Estimating the Lyapunov exponent of a chaotic system with nonparametric regression. J Am Stat Assoc 87:682–695

52. Meese R, Rogoff K (1983) Exchange rate models of the seventies. Do they fit out of sample? J Int Econ 14:3–24

53. Nelson DB (1990) Conditional heteroskedasticity in asset returns: A new approach. Econometrica 59:347–370

54. Nychka D, Ellner S, Gallant AR, McCaffrey D (1992) Finding chaos in noisy system. J Royal Stat Soc B 54:399–426

55. Pesaran MH, Timmermann A (1995) Predictability of stock returns: robustness and economic significance. J Financ 50:1201–1228

56. Pesin JB (1977) Characteristic Liapunov exponents and smooth ergodic theory. Russ Math Surv 32:55–114

57. Poterba JM, Summers LH (1988) Mean reversion in stock prices: evidence and implications. J Financ Econ 22:27–59

58. Potter SM (2000) Nonlinear impulse response functions. J Econ Dyn Control 24:1425–1446

59. Qi M (1999) Nonlinear predictability of stock returns using financial and economic variables. J Bus Econ Stat 17:419–429

60. Qi M, Maddala GS (1999) Economic factors and the stock market: a new perspective. J Forecast 18:151–166

61. Racine J (2001) On the nonlinear predictability of stock returns using financial and economic variables. J Bus Econ Stat 19:380–382

62. Richardson M, Stock JH (1989) Drawing inferences from statistics based on multiyear asset returns. J Financ Econ 25:323–348

63. Rosenstein MT, Collins JJ, De Luca CJ (1993) A practical method for calculating largest Lyapunov exponents from small data sets. Physica D 65:117–134

64. Scheinkman JA, LeBaron B (1989) Nonlinear dynamics and stock returns. J Bus 62:311–337

65. Schittenkopf C, Dorffner G, Dockner EJ (2000) On nonlinear, stochastic dynamics in economic and financial time series. Stud Nonlinear Dyn Econom 4:101–121

66. Shintani M (2006) A nonparametric measure of convergence towards purchasing power parity. J Appl Econom 21:589–604

67. Shintani M, Linton O (2003) Is there chaos in the world economy? A nonparametric test using consistent standard errors. Int Econ Rev 44:331–358

68. Shintani M, Linton O (2004) Nonparametric neural network estimation of Lyapunov exponents and a direct test for chaos. J Econom 120:1–33
69. Taylor SJ (1986) Modelling Financial Time Series. Wiley, New York
70. Tjostheim D, Auestad BH (1994) Nonparametric identification of nonlinear time series: Selecting significant lags. J Am Stat Assoc 89:1410–1419
71. Tschernig R, Yang L (2000) Nonparametric lag selection for time series. J Time Ser Analysis 21:457–487
72. Tschernig R, Yang L (2000) Nonparametric estimation of generalized impulse response functions. Michigan State University, unpublished
73. Valkanov R (2003) Long-horizon regressions: theoretical results and applications. J Financ Econom 68:201–232
74. Whang YJ, Linton O (1999) The asymptotic distribution of nonparametric estimates of the Lyapunov exponent for stochastic time series. J Econom 91:1–42
75. White H (1988) Economic prediction using neural networks: the case of IBM stock returns. Proceedings of the IEEE International Conference on Neural Networks 2. The Institute of Electrical and Electronics Engineers, San Diego, pp 451–458
76. White H, Racine J (2001) Statistical inference, the bootstrap, and neural-network modeling with application to foreign exchange rates. IEEE Trans Neural Netw 12:657–673
77. Wolf A, Swift JB, Swinney HL, Vastano JA (1985) Determining Lyapunov exponents from a time series. Physica D 16:285–317
78. Wolff RCL (1992) Local Lyapunov exponent: Looking closely at chaos. J Royal Stat Soc B 54:353–371
79. Wolff R, Yao Q, Tong H (2004) Statistical tests for Lyapunov exponents of deterministic systems. Stud Nonlinear Dyn Econom 8:10
80. Yao Q, Tong H (1994) Quantifying the influence of initial values on non-linear prediction. J Royal Stat Soc Ser B 56:701–725
81. Yao Q, Tong H (1994) On prediction and chaos in stochastic systems. Philos Trans Royal Soc Lond A 348:357–369

Finite Dimensional Controllability

LIONEL ROSIER
Institut Elie Cartan, Vandoeuvre-lès-Nancy, France

Article Outline

Glossary

Control system A **control system** is a dynamical system incorporating a **control input** designed to achieve a control objective. It is **finite dimensional** if the phase space (e. g. a vector space or a manifold) is of finite dimension. A **continuous-time** control system takes the form $\mathrm{d}x/\mathrm{d}t = f(x, u), x \in X, u \in U$ and $t \in \mathbb{R}$ denoting respectively the **state**, the **input**, and the **continuous time**. A **discrete-time** system assumes the form $x_{k+1} = f(x_k, u_k)$, where $k \in \mathbb{Z}$ is the **discrete time**.

Open/closed loop A control system is said to be in **open loop** form when the input u is any function of time, and in **closed loop** form when the input u is a function of the state only, i. e., it takes the more restrictive form $u = h(x(t))$, where $h: X \rightarrow U$ is a given function called a **feedback law**.

Controllability A control system is **controllable** if any pair of states may be connected by a trajectory of the system corresponding to an appropriate choice of the control input.

Stabilizability A control system is **asymptotically stabilizable** around an equilibrium point if there exists a feedback law such that the corresponding closed loop system is asymptotically stable at the equilibrium point.

Output function An **output function** is any function of the state.

Observability A control system given together with an output function is said to be **observable** if two different states give rise to two different outputs for a convenient choice of the input function.

Flatness An output function is said to be **flat** if the state and the input can be expressed as functions of the output and of a finite number of its derivatives.

Definition of the Subject

A control system is controllable if any state can be steered to another one in the phase space by an appropriate choice of the control input. While the stabilization issue has been addressed since the XIXth century (Watt's steam engine governor providing a famous instance of a stabilization mechanism at the beginning of the English Industrial Revolution), the controllability issue has been addressed for the first time by Kalman in the 1960s [20]. The controllability is a basic mathematical property which characterizes the degrees of freedom available when we try to control a system. It is strongly connected to other control concepts: optimal control, observability, and stabilizability. While the controllability of linear finite-dimensional systems is well understood since Kalman's seminal papers,

the situation is more tricky for nonlinear systems. For the later, concepts borrowed from differential geometry (e. g. Lie brackets, holonomy, ...) come into play, and the study of their controllability is still a field of active research.

The controllability of finite dimensional systems is a basic concept in control theory, as well as a notion involved in many applications, such as spatial dynamics (with e. g. spatial rendezvous), airplane autopilot, industrial robots, quantic chemistry.

Introduction

A very familiar example of a controllable finite dimensional system is given by a car that one attempts to park at some place in a parking. The phase space is roughly the three dimensional Euclidean space $\mathbb{R}^3$, a state being composed of the two coordinates of the center of mass together with the angle formed by some axis linked to the car with the (fixed) abscissa axis. The driver may act on the angle of the wheels and on their velocity, which may thus be taken as control inputs. In general, the presence of obstacles (e. g. other cars) impose to change the phase space $\mathbb{R}^3$ to a subset of it. The controllability issue is roughly how to combine changes of direction and of velocity to drive the car from a position to another one. Note that the system is controllable, even if the number of control inputs (2) is less than the number of independent coordinates (3). This is an important property resting upon the many connections between the coordinates of the state. While in nature each motion is generally controlled by an input (think of the muscles in an arm), the control theory focuses on the study of systems in which an input living in a space of low dimension (typically, one) is sufficient to control the coordinates of a state living in a space of high dimension.

The article is outlined as follows. In Sect. "Control Systems", we introduce the basic concepts (controllability, stabilizability) used thereafter. In the next section, we review the linear theory, recalling the Kalman and Hautus tests for the controllability of a time invariant system, and the Gramian test for a time dependent system. Section "Linearization Principle" is devoted to the linearization principle, which allows to deduce the controllability of a nonlinear system from the controllability of its linearization along a trajectory. The focus in Sect. "High Order Tests" is on nonlinear systems for which the linear test fails, i. e., the linearized system fails to be controllable. High order conditions based upon Lie brackets ensuring controllability will be given, first for systems without drift, and next for systems with a drift. Section "Controllability and Observability" explores the connections between controllability and observability, while Sect. "Controllability

and Stabilizability" shows how to derive stabilization results from the controllability property. A final section on the flatness, a new theory used in many applications to design explicit control inputs, is followed by some thoughts on future directions.

Control Systems

A **finite dimensional (continuous-time) control system** is a differential equation of the form

$$\dot{x} = f(x, u) \tag{1}$$

where $x \in X$ is the **state**, $u \in U$ is the **input**, $f: X \times U \to U$ is a smooth (typically real analytic) nonlinear function, $\dot{x} = \mathrm{d}x/\mathrm{d}t$, and X and U denote finite dimensional manifolds. For the sake of simplicity, we shall assume here that $X \subset \mathbb{R}^n$ and $U \subset \mathbb{R}^m$ are open sets. Sometimes, we impose U to be bounded (or to be a compact set) to force the control input to be bounded. Given some control input $u \in L^\infty(I; U)$, i. e. a measurable essentially bounded function $u: I \to U$, a **solution** of (1) is a locally Lipschitz continuous function $x(\cdot): J \to X$, where $J \subset I$, such that

$$\dot{x}(t) = f(x(t), u(t)) \qquad \text{for almost every } t \in J. \tag{2}$$

Note that $J \subset I$ only, that is, x needs not exist on all I, as it may escape to infinity in finite time. In general, u is piecewise smooth, so that (2) holds actually for all t except for finitely many values. The basic problem of the controllability is the issue whether, given an initial state $x_0 \in X$, a terminal state $x_T \in X$, and a control time $T > 0$, one may design a control input $u \in L^\infty([0, T]; U)$ such that the solution of the system

$$\begin{cases} \dot{x}(t) = f(x(t), u(t)) \\ x(0) = x_0 \end{cases} \tag{3}$$

satisfies $x(T) = x_T$.

An **equilibrium position** for (1) is a point $\overline{x} \in X$ such that there exists a value $\overline{u} \in U$ (typically, 0) such that $f(\overline{x}, \overline{u}) = 0$.

An **asymptotically stabilizing feedback law** is a function $k: X \to U$ with $k(\overline{x}) = \overline{u}$, such that the closed loop system

$$\dot{x} = f(x, k(x)) \tag{4}$$

obtained by plugging the control input

$$u(t) := k(x(t)) \tag{5}$$

into (1), is **locally asymptotically stable** at $\overline{x}$. Recall that it means that (i) the equilibrium point is **stable**: for any $\varepsilon > 0$, there exists some $\delta > 0$ such that any solution of (4) starting from a point x_0 such that $|x_0 - \overline{x}| < \delta$ at $t = 0$, is defined on $\mathbb{R}^+$ and satisfies $|x(t)| \leq \varepsilon$ for all $t \geq 0$; (ii) the equilibrium point is **attractive**. For some $\delta > 0$ as in (i), we have also that $x(t) \to 0$ as $t \to \infty$ whenever $|x_0| < \delta$.

The feedback laws considered here will be **continuous**, and we shall mean by a solution of (4) any function $x(t)$ satisfying (4) for all t. Notice that the solutions of the Cauchy problems exist (locally in time) by virtue of Peano theorem.

A control system is **asymptotically stabilizable** around an equilibrium position if an asymptotically stabilizing feedback law as above does exist. In the following, we shall also consider time-varying systems, i. e. systems of the form

$$\dot{x} = f(x, t, u) \tag{6}$$

where $f \colon X \times I \times U \to X$ is smooth and $I \subset \mathbb{R}$ denotes some interval. The controllability and stabilizability concepts extend in a natural way to that setting. Time-varying feedback laws

$$u(t) = k(x(t), t) \tag{7}$$

where $k \colon X \times \mathbb{R} \to U$ is a smooth (generally time periodic) function, prove to be useful in situations where the classical static stabilization defined above fails.

Linear Systems

Time Invariant Linear Systems

A **time invariant linear system** is a system of the form

$$\dot{x} = Ax + Bu \tag{8}$$

where $A \in \mathbb{R}^{n \times n}$, $B \in \mathbb{R}^{n \times m}$ denote some time invariant matrices. This corresponds to the situation where the function f in (1) is linear in both the state and the input. Here, $x \in \mathbb{R}^n$ and $u \in \mathbb{R}^m$, and we consider square integrable inputs $u \in L^2([0, T]; \mathbb{R}^m)$. We say that (8) is **controllable in time** T if for any pair of states $x_0, x_T \in \mathbb{R}^n$, one may construct an input $u \in L^2([0, T]; \mathbb{R}^m)$ that steers (8) from x_0 to x_T. Recall that the solution of (8) emanating from x_0 at $t = 0$ is given by Duhamel formula

$$x(t) = e^{tA} x_0 + \int_0^t e^{(t-s)A} Bu(s) \, ds \,.$$

Let us introduce the $n \times nm$ matrix $\mathcal{R}(A, B) = (B|AB|A^2B| \cdots |A^{n-1}B)$ obtained by gathering together the matrices

$B, AB, A^2 B, \ldots, A^{n-1} B$. Then we have the following rank condition due to Kalman [20] for the controllability of a linear system.

Theorem 1 *(Kalman) The linear system (8) is controllable in time T if and only if* $\operatorname{rank} \mathcal{R}(A, B) = n$.

We notice that the controllability of a linear system does not depend of the control time T, and that the control input may actually be chosen very smooth (e. g. in $C^\infty([0, T]; \mathbb{R}^n)$).

Example 1 Consider a pendulum to which is applied a torque as a control input. A simplified model is then given by the following linear system

$$\dot{x}_1 = x_2 \tag{9}$$

$$\dot{x}_2 = -x_1 + u \tag{10}$$

where x_1, x_2 and u stand respectively for the angle with the vertical, the angular velocity, and the torque. Here, $n = 2$, $m = 1$, and

$$A = \begin{pmatrix} 0 & 1 \\ -1 & 0 \end{pmatrix} \text{ and } B = \begin{pmatrix} 0 \\ 1 \end{pmatrix} . \tag{11}$$

As $\operatorname{rank}(B, AB) = 2$, we infer from Kalman rank test that (9)–(10) is controllable.

When (8) fails to be controllable, it may be important (e. g. when studying the stabilizability) to identify the uncontrollable part of (8). Assume that (8) is not controllable, and let $r = \operatorname{rank} \mathcal{R}(A, B) < n$. It may be seen that the reachable space from the origin, that is the set

$$\mathcal{R} = \left\{ x_T \in \mathbb{R}^n; \exists T > 0, \exists u \in L^2\left([0, T]; \mathbb{R}^n\right), \right.$$
$$\left. \int_0^T e^{(T-s)A} Bu(s) \, ds = x_T \right\}, \tag{12}$$

coincides with the space spanned by the columns of the matrix $\mathcal{R}(A, B)$:

$$\mathcal{R} = \{ \mathcal{R}(A, B) V ; V \in \mathbb{R}^{nm} \} \tag{13}$$

In particular, $\dim \mathcal{R} = \operatorname{rank} \mathcal{R}(A, B) = r < n$. Let $\mathbf{e} = \{e_1, \ldots, e_n\}$ be the canonical basis of $\mathbb{R}^n$, and let $\{f_1, \ldots, f_r\}$ be a basis of $\mathcal{R}$, that we complete in a basis $\mathbf{f} = \{f_1, \ldots, f_n\}$ of $\mathbb{R}^n$. If x (resp. $\tilde{x}$) denotes the vector of the coordinates of a point in the basis $\mathbf{e}$ (resp. $\mathbf{f}$), then $x = T\tilde{x}$ where $T = (f_1, f_2, \ldots, f_n) \in \mathbb{R}^{n \times n}$. In the new coordinates $\tilde{x}$, (8) may be written

$$\dot{\tilde{x}} = \tilde{A}\tilde{x} + \tilde{B}u \tag{14}$$

where $\tilde{A} := T^{-1}AT$ and $\tilde{B} := T^{-1}B$ read

$$\tilde{A} = \begin{pmatrix} \tilde{A}_1 & \tilde{A}_2 \\ 0 & \tilde{A}_3 \end{pmatrix} \quad \tilde{B} = \begin{pmatrix} \tilde{B}_1 \\ 0 \end{pmatrix} \qquad (15)$$

with $\tilde{A}_1 \in \mathbb{R}^{r \times r}$, $\tilde{B}_1 \in \mathbb{R}^{r \times m}$. Writing $\tilde{x} = (\tilde{x}_1, \tilde{x}_2) \in \mathbb{R}^r \times \mathbb{R}^{n-r}$, we have

$$\dot{\tilde{x}}_1 = \tilde{A}_1 \tilde{x}_1 + \tilde{A}_2 \tilde{x}_2 + \tilde{B}_1 u \qquad (16)$$

$$\dot{\tilde{x}}_2 = \tilde{A}_3 \tilde{x}_2 \qquad (17)$$

and it may be proved that

$$\operatorname{rank} \mathcal{R}(\tilde{A}_1, \tilde{B}_1) = r . \qquad (18)$$

This is the **Kalman controllability decomposition**. By (18), the dynamics of $\tilde{x}_1$ is well controlled. Actually, a solution of (18) evaluated at $t = T$ assumes the form

$$\tilde{x}_1(T) = e^{T\tilde{A}_1} \tilde{x}_1(0) + \int_0^T e^{(T-s)\tilde{A}_1} \tilde{A}_2 \tilde{x}_2(s) \, ds$$
$$+ \int_0^T e^{(T-s)\tilde{A}_1} \tilde{B}_1 u(s) ds$$
$$= e^{T\tilde{A}_1} \overline{x}_1 + \int_0^T e^{(T-s)\tilde{A}_1} \tilde{B}_1 u(s) ds$$

if we set $\overline{x}_1 = \tilde{x}_1(0) + \int_0^T e^{-s\tilde{A}_1} \tilde{A}_2 \tilde{x}_2(s) ds$. Hence $\tilde{x}_1(T)$ may be given any value in $\mathbb{R}^r$. On the other hand, no control input is present in (17). Thus $\tilde{x}_2$ stands for the uncontrolled part of the dynamics of (8).

Another test based upon a spectral analysis has been furnished by Hautus in [16].

Theorem 2 *(Hautus) The control system (8) is controllable in time T if and only if* $\operatorname{rank}(\lambda I - A, B) = n$ *for all* $\lambda \in \mathbb{C}$.

Notice that in Hautus test we may restrict ourselves to the complex numbers λ which are eigenvalues of A, for otherwise $\operatorname{rank}(\lambda I - A) = n$.

Time-Varying Linear Systems

Let us now turn to the controllability issue for a time-varying linear system

$$\dot{x}(t) = A(t)x(t) + B(t)u(t) \qquad (19)$$

where $A \in L^\infty([0, T] ; \mathbb{R}^{n \times n})$, $B \in L^\infty([0, T] ; \mathbb{R}^{n \times m})$ denote time-varying matrices. Such a system arises in a natural way when linearizing a control system (1) along a trajectory. The input $u(t)$ is any function in $L^\infty([0, T] ; \mathbb{R}^m)$ and a solution of (19) is any locally Lipschitz continuous function satisfying (19) almost everywhere.

We define the **fundamental solution** ϕ associated with A as follows. Pick any $s \in [0, T]$, and let $M : [0, T] \to \mathbb{R}^{n \times n}$ denote the solution of the system

$$\dot{M}(t) = A(t)M(t)$$
$$M(s) = I \qquad (20)$$

Then $\phi(t, s) := M(t)$. Notice that $\phi(t, s) = e^{(t-s)A}$ when A is constant.

The solution x of (20) starting from x_0 at time t_0 reads then

$$x(t) = \phi(t, t_0)x_0 + \int_{t_0}^t \phi(t, s)B(s)u(s)ds .$$

The **controllability Gramian** of (20) is the matrix

$$G = \int_0^T \phi(T, t)B(t)B^*(t)\phi^*(T, t)dt$$

where * denotes transpose. Note that $G \in \mathbb{R}^{n \times n}$, and that G is a nonnegative symmetric matrix. Then the following result holds.

Theorem 3 *(Gramian test) The system (19) is controllable on $[0, T]$ if and only if the Gramian G is invertible.*

Note that the Gramian test provides a third criterion to test whether a time invariant linear system (8) is controllable or not.

Corollary 4 *(8) is controllable in time T if and only if the Gramian*

$$G = \int_0^T e^{(T-t)A} BB^* e^{(T-t)A^*} dt$$

is invertible.

As the value of the control time T plays no role according to Kalman test, it follows that the Gramian G is invertible for *all* $T > 0$ whenever it is invertible for *one* $T > 0$.

If (19) is controllable, then an explicit control input steering (19) from x_0 to x_T is given by

$$\overline{u}(t) = B^*(t)\phi^*(T, t)G^{-1}(x_T - \phi(T, 0)x_0) . \qquad (21)$$

A remarkable property of the control input $\overline{u}$ is that $\overline{u}$ minimizes the control cost

$$E(u) = \int_0^T |u(t)|^2 dt$$

among all the control inputs $u \in L^\infty([0, T] ; \mathbb{R}^m)$ (or $u \in L^2([0, T] ; \mathbb{R}^m)$) steering (19) from x_0 to x_T. Actually a little more can be said.

Proposition 5 *If $u \in L^2([0, T]; \mathbb{R}^m)$ is such that the solution x of (19) emanating from x_0 at $t = 0$ reaches x_T at time T, and if $u \neq \overline{u}$, then*

$$E(\overline{u}) < E(u).$$

The main drawback of the Gramian test is that the knowledge of the fundamental solution $\phi(t, s)$ and the computation of an integral term are both required. In the situation where $A(t)$ and $B(t)$ are smooth functions of time, a criterion based only upon derivatives in time is also available. Assume that $A \in C^\infty([0, T]; \mathbb{R}^{n \times n})$ and that $B \in C^\infty([0, T]; \mathbb{R}^{n \times m})$, and define a sequence of functions $B_i \in C^\infty([0, T]; \mathbb{R}^{n \times m})$ by induction on i by

$$
\begin{aligned}
B_0 &= B \\
B_i &= AB_{i-1} - \frac{\mathrm{d}B_{i-1}}{\mathrm{d}t} \,.
\end{aligned}
\tag{22}
$$

Then the following result holds (see, e. g., [11]).

Theorem 6 *Assume that there exists a time $\overline{t} \in [0, T]$ such that*

$$\mathrm{span}(B_i(\overline{t})v \,; v \in \mathbb{R}^m, i \geq 0) = \mathbb{R}^n \,. \tag{23}$$

Then (19) is controllable.

The converse of Theorem 6 is true when A and B are real analytic functions of time. More precisely, we have the following result.

Theorem 7 *If A and B are real analytic on $[0, T]$, then (20) is controllable if and only if for all $\overline{t} \in [0, T]$*

$$\mathrm{span}(B_i(\overline{t})v \,; v \in \mathbb{R}^m, i \geq 0) = \mathbb{R}^n \,.$$

Clearly, Theorem 7 is not valid when A and B are merely of class C^∞. (Take $n = m = 1$, $A(t) = 0$, $B(t) = \exp(-t^{-1})$ and $\overline{t} = 0$.)

Linear Control Systems in Infinite Dimension

Let us end this section with some comments concerning the extensions of the above controllability tests to control systems in infinite dimension (see [30] for more details). Let us consider a control system of the form

$$\dot{x} = Ax + Bu \tag{24}$$

where $A: D(A) \subset X \to X$ is an (unbounded) operator generating a strongly continuous semigroup $(S(t))_{t \geq 0}$ on a (complex) Hilbert space X, and $B: U \to X$ is a bounded operator, U denoting another Hilbert space.

Definition 8 We shall say that (24) is

- **Exactly controllable in time** T if for any $x_0, x_T \in X$ there exists $u \in L^2([0, T]; U)$ such that the solution x of (24) emanating from x_0 at $t = 0$ satisfies $x(T) = x_T$;
- **Null controllable in time** T if for any $x_0 \in X$ there exists $u \in L^2([0, T]; U)$ such that the solution x of (24) emanating from x_0 at $t = 0$ satisfies $x(T) = 0$;
- **Approximatively controllable in time** T if for any $x_0, x_T \in X$ and any $\varepsilon > 0$ there exists $u \in L^2([0, T]; U)$ such that the solution x of (24) emanating from x_0 at $t = 0$ satisfies $||x(T) - x_T|| < \varepsilon$ ($|| \cdot ||$ denoting the norm in X).

This setting is convenient for a partial differential equation with an internal control $Bu := gu(t)$, where $g = g(x)$ is such that $gU \subset X$. Let us review the above controllability tests.

- Kalman rank test, which is based on a computation of the dimension of the reachable space, possesses some extension giving the approximate (not exact!) controllability of (24) (See [13], Theorem 3.16). More interestingly, a Kalman-type condition has been introduced in [23] to investigate the null controllability of a system of coupled parabolic equations.
- Hautus test admits the following extension due to Liu [25]: (24) is (exactly) controllable in time T if and only if there exists some constant $\delta > 0$ such that

$$
\begin{aligned}
||(A^* - \lambda I)z||^2 + ||B^*z||^2 &\geq \delta ||z||^2 \\
&\forall z \in D(A^*), \forall \lambda \in \mathbb{C}.
\end{aligned}
\tag{25}
$$

In (25), A^* (resp. B^*) denotes the adjoint of the operator A (resp. B), and $|| \cdot ||$ denotes the norm in X.

- The Gramian test admits the following extension, due to Dolecky–Russell [14] and J.L. Lions [24]: (24) is exactly controllable in time T if and only if there exists some constant $\delta > 0$ such that

$$\int_0^T ||B^*S^*(t)x_0||^2 \mathrm{d}t \geq \delta ||x_0||^2 \quad \forall x_0 \in X \,.$$

Linearization Principle

Assume given a smooth nonlinear control system

$$\dot{x} = f(x, u) \tag{26}$$

where $f: X \times U \to \mathbb{R}^n$ is a smooth map (i. e. of class C^∞) and $X \subset \mathbb{R}^n$, $U \subset \mathbb{R}^m$ denote some open sets. Assume also given a reference trajectory $(\overline{x}, \overline{u}): [0, T] \to X \times U$

where $\overline{u} \in L^{\infty}([0, T] ; U)$ is the control input and $\overline{x}$ solves (26) for $u \equiv \overline{u}$. We introduce the linearized system along the reference trajectory defined as

$$\dot{y} = A(t)y + B(t)v \tag{27}$$

where

$$
\begin{aligned}
A(t) &= \frac{\partial f}{\partial x}(\overline{x}(t), \overline{u}(t)) \in \mathbb{R}^{n \times n} , \\
B(t) &= \frac{\partial f}{\partial u}(\overline{x}(t), \overline{u}(t)) \in \mathbb{R}^{n \times m}
\end{aligned}
\tag{28}
$$

and $y \in \mathbb{R}^n, v \in \mathbb{R}^m$. (27) is formally derived from (26) by letting $x = \overline{x} + y, u = \overline{u} + v$ and observing that

$$
\begin{aligned}
\dot{y} &= \dot{x} - \dot{\overline{x}} = f(\overline{x} + y, \overline{u} + v) - f(\overline{x}, \overline{u}) \\
&\approx \frac{\partial f}{\partial x}(\overline{x}, \overline{u})y + \frac{\partial f}{\partial u}(\overline{x}, \overline{u})v .
\end{aligned}
$$

Notice that if $(\overline{x}_0, \overline{u}_0)$ is an equilibrium point of f (i. e. $f(\overline{x}_0, \overline{u}_0) = 0$) and $\overline{x}(t) \equiv x_0, \overline{u}(t) \equiv u_0$, then $A(t) = A = \frac{\partial f}{\partial x}(\overline{x}_0, \overline{u}_0)$ and $B(t) = B = \frac{\partial f}{\partial u}(\overline{x}_0, \overline{u}_0)$ take constant values. Equation (27) is in that case the time invariant linear system

$$\dot{y} = Ay + Bv . \tag{29}$$

Let $x_0, x_T \in X$. We seek for a trajectory x of (26) connecting x_0 to x_T when x_0 (resp. x_T) is close to $\overline{x}(0)$ (resp. $\overline{x}(T)$). In addition, we will impose that the trajectory (x, u) be uniformly close to the reference trajectory $(\overline{x}, \overline{u})$. We are led to the following

Definition 9 The system (26) is said to be **controllable along** $(\overline{x}, \overline{u})$ if for each $\varepsilon > 0$, there exists some $\delta > 0$ such that for each $x_0, x_T \in X$ with $||x_0 - \overline{x}(0)|| < \delta, ||x_T - \overline{x}(T)|| < \delta$, there exists a control input $u \in L^{\infty}([0, T] ; U)$ such that the solution of (26) starting from x_0 at $t = 0$ satisfies $x(T) = x_T$ and

$$\sup_{t \in [0, T]} \left(|x(t) - \overline{x}(t)| + |u(t) - \overline{u}(t)| \right) \leq \varepsilon .$$

We are in a position to state the linearization principle.

Theorem 10 Let $(\overline{x}, \overline{u})$ be a trajectory of (26). If the linearized system (27) along $(\overline{x}, \overline{u})$ is controllable, then the system (26) is controllable along the trajectory $(\overline{x}, \overline{u})$.

When the reference trajectory is stationary, we obtain the following result.

Corollary 11 Let (x_0, u_0) be such that $f(x_0, u_0) = 0$. If the linearized system (29) is controllable, then the system (26) is controllable along the stationary trajectory $(\overline{x}, \overline{u}) = (x_0, u_0)$.

Notice that the converse of Corollary 11 is not true. (Consider the system $\dot{x} = u^3$ with $n = m = 1$, $x_0 = u_0 = 0$, which is controllable at the origin as it may be seen by performing the change of inputs $v = u^3$.) Often, the nonlinear part of f plays a crucial role in the controllability of (26). This will be explained in the next section using Lie algebraic techniques. Another way to use the nonlinear contribution in f is to consider a linearization of (26) along a convenient (not stationary) trajectory. We consider a system introduced by Brockett in [6] to exhibit an obstruction to stabilizability.

Example 2 The Brockett's system reads

$$\dot{x}_1 = u_1 \tag{30}$$

$$\dot{x}_2 = u_2 \tag{31}$$

$$\dot{x}_3 = x_1 u_2 - x_2 u_1 . \tag{32}$$

Its linearization along $(x_0, u_0) = (0, 0)$, which reads

$$
\begin{aligned}
\dot{y}_1 &= v_1 \\
\dot{y}_2 &= v_2 \\
\dot{y}_3 &= 0 ,
\end{aligned}
$$

is not controllable, by virtue of the Kalman rank condition. We may however construct a smooth closed trajectory such that the linearization of (30)–(32) along it is controllable. Pick any time $T > 0$ and let $\overline{u}_1(t) = \cos(\pi t/T)$, $\overline{u}_2(t) = 0$ for $t \in [0, T]$ and $x_0 = 0$. Let $\overline{x}$ denote the corresponding solution of (30)–(32). Notice that $\overline{x}_1(t) = (T/\pi) \sin(\pi t/T)$ (hence $\overline{x}_1(T) = 0$), and $\overline{x}_2 = \overline{x}_3 \equiv 0$. The linearization of (30)–(32) along $(\overline{x}, \overline{u})$ is (27) with

$$
\begin{aligned}
A(t) &= \begin{pmatrix} 0 & 0 & 0 \\ 0 & 0 & 0 \\ 0 & -\cos(\pi t/T) & 0 \end{pmatrix} , \\
B(t) &= \begin{pmatrix} 1 & 0 \\ 0 & 1 \\ 0 & (T/\pi)\sin(\pi t/T) \end{pmatrix} .
\end{aligned}
\tag{33}
$$

Notice that A and B are real analytic, so that we may apply Theorem 8 to check whether (27) is controllable or not on $[0, T]$. Simple computations give

$$
B_1(t) = \begin{pmatrix} 0 & 0 \\ 0 & 0 \\ 0 & -2\cos(\pi t/T) \end{pmatrix} .
$$

Clearly,

$$
\text{span}\left(B_0(0)\begin{pmatrix} 1 \\ 0 \end{pmatrix}, B_0(0)\begin{pmatrix} 0 \\ 1 \end{pmatrix}, B_1(0)\begin{pmatrix} 0 \\ 1 \end{pmatrix} \right) = \mathbb{R}^3
$$

hence (27) is controllable. We infer from Theorem 10 that (30)–(32) is controllable along $(\overline{x}, \overline{u})$.

Notice that if we want to prove the controllability around an equilibrium point as above for Brockett's system, we have to design a reference control input $\overline{u}$ so that

$$x_0 = \overline{x}(0) = \overline{x}(T) . \tag{34}$$

When f is odd with respect to the control, i. e.

$$f(x, -u) = -f(x, u) , \tag{35}$$

then (34) is automatically satisfied whenever $\overline{u}$ fulfills

$$\overline{u}(t) = -\overline{u}(T - t) \quad \forall t \in [0, T] . \tag{36}$$

Indeed, it follows from (35)–(36) that the solution $\overline{x}$ to (26) starting from x_0 at $t = 0$ satisfies

$$\overline{x}(t) = \overline{x}(T - t) \quad \forall t \in [0, T] . \tag{37}$$

In particular, $\overline{x}(T) = \overline{x}(0) = x_0$. Of course, the control inputs of interest are those for which the linearized system (27) is controllable, and the latter property is "generically" satisfied for a controllable system. The above construction of the reference trajectory is due to J.-M. Coron, and is referred to as the **return method**. A precursor to that method is [34]. Beside giving interesting results for the stabilization of finite dimensional systems (see below Sect. "Controllability and Stabilizability"), the return method has also been successfully applied for the control of some important partial differential equations arising in Fluid Mechanics (see [11]).

High Order Tests

In this section, we shall derive new controllability tests for systems for which the linearization principle is inconclusive; that is, the linearization at an equilibrium point fails to be controllable. To simplify the exposition, we shall limit ourselves to systems **affine in the control**, i. e. systems of the form

$$\dot{x} = f_0(x) + u_1 f_1(x) + \cdots + u_m f_m(x), \quad x \in \mathbb{R}^n, |u|_\infty \leq \delta \tag{38}$$

where $|u|_\infty := \sup_{1 \leq i \leq m} |u_i|$ and $\delta > 0$ denotes a fixed number. We assume that $f_0(0) = 0$ and that $f_i \in C^\infty(\mathbb{R}^n, \mathbb{R}^n)$ for each $i \in \{0, \ldots, m\}$. To state the results we need to introduce a few notations. Let $v = (v_1, \ldots, v_n)$ and $w = (w_1, \ldots, w_n)$ be two vector fields of class C^∞ on

$\mathbb{R}^n$. The **Lie bracket** of v and w, denoted by $[v, w]$, is the vector field

$$[v, w] = \frac{\partial w}{\partial x} v - \frac{\partial v}{\partial x} w$$

where $\partial w / \partial x$ is the Jacobian matrix $(\partial w_i / \partial x_j)_{i,j=1,\ldots,n}$, and the vector $v(x) = (v_1(x), \ldots, v_n(x))$ is identified to the column $\begin{pmatrix} v_1(x) \\ \vdots \\ v_n(x) \end{pmatrix}$. As $[v, w]$ is still a smooth vector field, we may bracket it with v, or w, etc. Vector fields like $[v, [v, w]], [[v, w], [v, [v, w]]]$, etc. are termed **iterated Lie brackets** of v, w.

The Lie bracketing of vector fields is an operation satisfying the two following properties (easy to check)

1. **Anticommutativity:** $[w, v] = -[v, w]$;
2. **Jacobi identity:** $[f, [g, h]] + [g, [h, f]] + [h, [f, g]] = 0$.

The **Lie algebra** generated by $f_1, \ldots, f_m$, denoted $\text{Lie}(f_1, \ldots, f_m)$, is the smallest vector subspace v of $C^\infty(\mathbb{R}^n, \mathbb{R}^n)$ which contains $f_1, \ldots, f_m$ and which is closed under Lie bracketing (i. e. $v, w \in V \Rightarrow [v, w] \in V$). It is easily to see that $\text{Lie}(f_1, \ldots, f_m)$ is the vector space spanned by all the iterated Lie brackets of $f_1, \ldots, f_m$. (Actually, using the anticommutativity and Jacobi identity, we may restrict ourselves to iterated Lie brackets of the form $[f_{i_1}, [f_{i_2}, [\ldots [f_{i_{p-1}}, f_{i_p}] \ldots]]]$ to span $\text{Lie}(f_1, \ldots, f_m)$.) For any $x \in \mathbb{R}^n$, we set

$$\text{Lie}(f_1, \ldots, f_m)(x) = \{g(x); g \in \text{Lie}(f_1, \ldots, f_m)\} \subset \mathbb{R}^n .$$

Example 3

Let us consider the system

$$\dot{x} = u_1 f_1(x) + u_2 f_2(x) = u_1 \begin{pmatrix} 1 \\ 0 \\ 0 \end{pmatrix} + u_2 \begin{pmatrix} 0 \\ 1 \\ x_1 \end{pmatrix} \tag{39}$$

where $x = (x_1, x_2, x_3) \in \mathbb{R}^3$. Clearly, the linearized system at $(x_0, u_0) = (0, 0)$, which reduces to $\dot{x} = u_1 f_1(0) + u_2 f_2(0)$, is not controllable. However, we shall see later that the system (39) is controllable. First, each point $(\pm \varepsilon, 0, 0)$ (resp. $(0, \pm \varepsilon, 0)$) may be reached from the origin by letting $u(t) = (\pm 1, 0)$ (resp. $u(t) = (0, \pm 1)$) on the time interval $[0, \varepsilon]$. More interestingly, any point $(0, 0, \varepsilon^2)$ may also be reached from the origin in a time $T = O(\varepsilon)$, even though $(0, 0, 1) \notin \text{span}(f_1(0), f_2(0))$. To prove the last claim, let us introduce for $i = 1, 2$ the flow map ϕ_i^t defined by $\phi_i^t(x_0) = x(t)$, where $x(\cdot)$ solves

$$\dot{x} = f_i(x) , \quad x(0) = x_0 .$$

Then, it is easy to see that

$$\phi_2^{-\varepsilon} \phi_1^{-\varepsilon} \phi_2^{\varepsilon} \phi_1^{\varepsilon}(0) = (0, 0, \varepsilon^2) .$$
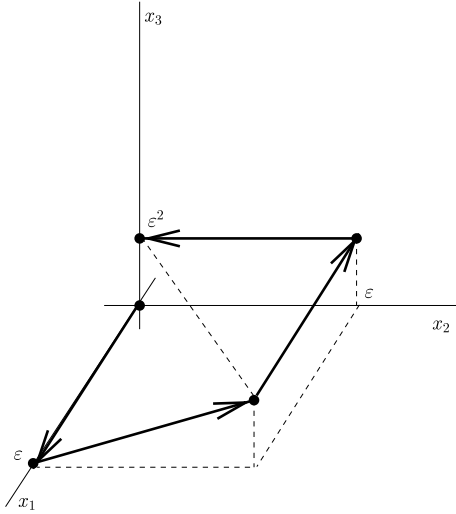
Finite Dimensional Controllability, Figure 1
Trajectory from $x_0 = (0, 0, 0)$ to $x_T = (0, 0, \varepsilon^2)$

It means that the control

$$u(t) = \begin{cases} (1, 0) & \text{if } 0 \leq t < \varepsilon \\ (0, 1) & \text{if } \varepsilon \leq t < 2\varepsilon \\ (-1, 0) & \text{if } 2\varepsilon \leq t < 3\varepsilon \\ (0, -1) & \text{if } 3\varepsilon \leq t < 4\varepsilon \end{cases}$$

steers the solution of (39) from $x_0 = (0, 0, 0)$ at $t = 0$ to $x_T = (0, 0, \varepsilon^2)$ at $T = 4\varepsilon$. (See Fig. 1.) More generally, if f_1 and f_2 denote arbitrary smooth vector fields and ϕ_1^t, ϕ_2^t denote the corresponding flow maps, we have that

$$\phi_2^{-\varepsilon}\phi_1^{-\varepsilon}\phi_2^{\varepsilon}\phi_1^{\varepsilon}(0) = \varepsilon^2[f_1, f_2](0) + O(\varepsilon^3) . \tag{40}$$

Thus, it is possible to reach points in the direction of the Lie bracket $[f_1, f_2](0)$. However, the process is not very efficient: in the direction of the Lie bracket $[f_1, f_2](0)$, only points located at a distance from the origin of order ε^2 may be reached in a time of order ε.

Let us proceed to the controllability properties of affine systems. We first review the results for systems without drift (i. e. $f_0 = 0$), and next consider the general (only partially understood) situation where a drift exists.

Affine Systems Without Drift

We assume that $f_0 = 0$, so that (38) takes the form

$$\dot{x} = u_1 f_1(x) + \cdots + u_m f_m(x) . \tag{41}$$

When $m < n$, the linearized system at any stationary trajectory $(\overline{x}, 0)$, which reads

$$\dot{y} = (f_1(\overline{x}) \cdots f_m(\overline{x}))v , \tag{42}$$

cannot be controllable. High order tests are therefore very useful in this setting. We adopt the following

Definition 12 The system (41) is said to be **controllable** if for any $T > 0$ and any $x_0, x_T \in \mathbb{R}^n$, there exists a control $u \in L^\infty([0, T]; \mathbb{R}^m)$ such that the solution $x(t)$ of (41) emanating from x_0 at $t = 0$ reaches x_T at time $t = T$.

Note that the controllability may be supplemented with the words: local, with small controls, in small time, in large time, etc. An important instance, the small-time local controllability, is introduced below in Definition 14.

The following result has been obtained by Rashevski [28] and Chow [8].

Theorem 13 *(Rashevski–Chow) Assume that $f_i \in C^\infty$ $(\mathbb{R}^n, \mathbb{R}^n)$ for all $i \in \{1, \ldots, m\}$. If*

$$\text{Lie}(f_1, \ldots, f_m)(x) = \mathbb{R}^n \qquad \forall x \in \mathbb{R}^n , \tag{43}$$

then (41) is controllable.

Example 3 continued (39) is controllable, since $[f_1, f_2] = (0, 0, 1)$ gives

$$\text{span}(f_1(x), f_2(x), [f_1, f_2](x)) = \mathbb{R}^3 \quad \forall x \in \mathbb{R}^3 . \tag{44}$$

Example 2 continued Brockett's system is also controllable, since $[f_1, f_2] = (0, 0, 2)$ gives (43).

Notice that Theorem 13 is almost sharp. Indeed, it has been proved by Sussmann–Jurdjevic in [36] that (43) has to be satisfied for a controllable system (41) with real analytic vector fields (i. e. $f_i \in C^\omega(\mathbb{R}^n, \mathbb{R}^n)$ for each i).

Affine Systems with Drifts

We consider now a control system with a drift

$$\dot{x} = f_0(x) + u_1 f_1(x) + \cdots + u_m f_m(x) \tag{45}$$

where $f_i \in C^\infty(\mathbb{R}^n, \mathbb{R}^n)$ for any $i \in \{0, \ldots, m\}$, and $f_0(0) = 0$ but $f_0 \not\equiv 0$. As no control is associated with the vector field f_0, one expects that the latter will occupy a singular place in the controllability tests.

Let $\rho > 0$ be a given number. We assume that the control input u is a measurable function taking its values in the compact set $U_\rho = [-\rho, \rho]^m$. To emphasize the dependence in u, we shall sometimes denote by $x(t, u)$ the solution of (45) such that $x(0, u) = 0$. The **attainable set at time** $T > 0$ is the set

$$\mathcal{A}_\rho(T) = \{x(T, u); u(t) \in U_\rho \forall t \in [0, T]\} .$$

Definition 14 We shall say that the control system (45) is

- **accessible** from the origin if the interior of $\mathcal{A}_\rho(T)$ is nonempty for any $T > 0$;
- **small time locally controllable (STLC)** at the origin if for any $T > 0$ there exists a number $\delta > 0$ such that for any pair of states (x_0, x_T) with $|x_0| < \delta$, $|x_T| < \delta$, there exists a control input $u \in L^\infty([0, T]; U_\rho)$ such that the solution x of

$$\dot{x} = f_0(x) + u_1 f_1(x) + \cdots + u_m f_m(x), \quad x(0) = x_0 \quad (46)$$

satisfies $x(T) = x_T$.

(Notice that in the definition of the small time local controllability, certain authors assume $x_0 = 0$ in above definition, or require the existence of $\delta > 0$ for any $T > 0$ and any $\rho > 0$.)

The accessibility property turns out to be easy to characterize. As a STLC system is clearly accessible, the accessibility property is often considered as the first property to test before investigating the controllability of a given affine system.

The following result provides an accessibility test based upon the rank at the origin of the Lie algebra generated by all the vectors fields involved in the control system.

Theorem 15 *(Hermann–Nagano) If*

$$\dim \mathrm{Lie}(f_0, f_1, \ldots, f_m)(0) = n \quad (47)$$

then the system (45) is accessible from the origin. Conversely, if (45) is accessible from the origin and the vector fields f_0, f_1, …, f_m are real analytic, then (47) has to be satisfied.

Example 4 Pick any $k \in \mathbb{N}^*$ and consider as in [22] the system

$$\begin{cases} \dot{x}_1 = u, \quad |u| \leq 1, \\ \dot{x}_2 = (x_1)^k \end{cases} \quad (48)$$

so that $n = 2$, $m = 1$, and $f_0(x) = (0, (x_1)^k)$, $f_1(x) = (1, 0)$. Setting

$$(\mathrm{ad}\, f_1, f_0) = [f_1, f_0]$$
$$(\mathrm{ad}^{i+1} f_1, f_0) = [f_1, (\mathrm{ad}^i f_1, f_0)] \quad \forall i \geq 1$$

we obtain at once that

$$(\mathrm{ad}^k f_1, f_0)(x) = (0, k!)$$

hence

$$\mathrm{Lie}(f_0, f_1)(0) = \mathbb{R}^2 .$$

It follows from Theorem 15 that (48) is accessible from the origin. On the other hand, it is clear that (48) is not STLC at the origin when k is even, since $\dot{x}_2 = (x_1)^k \geq 0$, hence x_2 is nondecreasing. Using a controllability test due to Hermes [17], it may be shown that (48) is STLC at the origin if and only if k is odd. Note that the linearized system at the origin fails to be controllable whenever $k \geq 2$.

Let us consider some affine system (45) which is accessible from the origin. We exclude the trivial situation when the linearization principle may be applied, and seek for a Lie algebraic condition ensuring that (45) is STLC. If h is a given iterated Lie bracket of $f_0, f_1, \ldots, f_m$, we let $\delta_i(h)$ denote the number of occurrences of f_i in the definition of h. For instance, if $m = 3$ and

$$h = [[f_1, [f_1, f_0]], [f_1, f_2]] \quad (49)$$

then

$$\delta_0(h) = 1, \delta_1(h) = 3, \delta_2(h) = 1, \delta_3(h) = 0 .$$

Notice that the fields $f_0, f_1, \ldots, f_m$ are considered as indeterminates when computing the $\delta_i(h)$'s, their effective values as vector fields being ignored.

Let S_m denote the usual symmetric group, i. e. S_m is the group of all the permutations of the set $\{1, \ldots, m\}$. If $\psi \in S_m$ and h is an iterated Lie bracket of $f_0, f_1, \ldots, f_m$, we denote by h^ψ the iterated Lie bracket obtained by replacing, in the definition of h, f_i by $f_{\psi(i)}$ for each $i \in \{1, \ldots, m\}$. For instance, if $\psi = (1\,2\,3)$ and h is as in (49), then

$$h^\psi = [[f_2, [f_2, f_0]], [f_2, f_3]] .$$

Finally, we set

$$\sigma(h) = \sum_{\psi \in S_m} h^\psi .$$

With h as in (49), and $S_3 = \{\mathrm{id}_{\{1,2,3\}}, (1\,2\,3), (1\,3\,2), (1\,2), (1\,3), (2\,3)\}$, we obtain

$$\begin{aligned} \sigma(h) = {} & [[f_1, [f_1, f_0]], [f_1, f_2]] + [[f_2, [f_2, f_0]], [f_2, f_3]] \\ & + [[f_3, [f_3, f_0]], [f_3, f_1]] + [[f_2, [f_2, f_0]], [f_2, f_1]] \\ & + [[f_3, [f_3, f_0]], [f_3, f_2]] + [[f_1, [f_1, f_0]], [f_1, f_3]] . \end{aligned}$$

We need the following

Definition 16 Let $\theta \in [0, 1]$. We say that the system (45) satisfies the condition $S(\theta)$ if, for any iterated Lie bracket h with $\delta_0(h)$ odd and $\delta_i(h)$ even for all $i \in \{1, \ldots, m\}$, the

vector $\sigma(h)(0)$ belongs to the vector space spanned by the vectors $g(0)$ where g is an iterated Lie bracket satisfying

$$\theta\delta_0(g) + \sum_{i=1}^{m}\delta_i(g) < \theta\delta_0(h) + \sum_{i=1}^{m}\delta_i(h) .$$

Then we have the following result proved in [35].

Theorem 17 *(Sussmann) If the condition $S(\theta)$ is satisfied for some $\theta \in [0, 1]$, then the system (45) is small time locally controllable at the origin.*

When $\theta = 0$, Sussmann's theorem is nothing else than Hermes' theorem. Sussmann's theorem, which is in itself very useful in Engineering (see, e. g., [5]), has been extended in [1,4,18].

Controllability and Observability

Assume given a control system

$$\dot{x} = f(x, u) \tag{50}$$

together with an output function

$$y = h(x) \tag{51}$$

where $f: X \times U \to \mathbb{R}^n$ and $h: X \to Y$ are smooth functions, and $X \subset \mathbb{R}^n$, $U \subset \mathbb{R}^m$ and $Y \subset \mathbb{R}^p$ denote some open sets. y typically stands for a (partial) measurement of the state, e. g. the p first coordinates, where $p < n$. Often, only y is available, and for the stabilization of (50) we should consider an output feedback law of the form $u = k(y)$. We shall say that (50)–(51) is **observable** on the interval $[0, T]$ if, for any pair $x_0, \tilde{x}_0$ of points in X, one may find a control input $u \in L^\infty([0, T]; U)$ such that if x (resp. $\tilde{x}$) denotes the solution of (50) emanating from x_0 (resp. $\tilde{x}_0$) at time $t = 0$, and y (resp. $\tilde{y}$) denotes the corresponding output function, we have

$$y(t) \neq \tilde{y}(t) \quad \text{for some } t \in [0, T] . \tag{52}$$

For a time invariant linear control system and a linear output function

$$\dot{x} = Ax + Bu \tag{53}$$

$$y = Cx \tag{54}$$

with $A \in \mathbb{R}^{n \times n}$, $B \in \mathbb{R}^{n \times m}$, $C \in \mathbb{R}^{p \times n}$, the output is found to be

$$y(t) = Cx(t) = Ce^{tA}x_0 + C\int_0^t e^{(t-s)A}Bu(s)ds . \tag{55}$$

In particular,

$$y(t) - \tilde{y}(t) = Ce^{tA}(x_0 - \tilde{x}_0) \tag{56}$$

and B does not play any role. Therefore, (53)–(54) is observable in time T if and only if the only state $\overline{x} \in \mathbb{R}^n$ such that $Ce^{tA}\overline{x} = 0$ for any $t \in [0, T]$ is $\overline{x} = 0$. Differentiating in time and applying Cayley–Hamilton theorem, we obtain the following result.

Theorem 18 *System (53)–(54) is observable in time T if and only if* rank $\mathcal{O}(A, C) = n$, *where the* **observability matrix** $\mathcal{O}(A, C) \in \mathbb{R}^{np \times n}$ *is defined by*

$$\mathcal{O}(A, C) = \begin{pmatrix} C \\ CA \\ \vdots \\ CA^{n-1} \end{pmatrix}.$$

Noticing that

$$\mathcal{R}(A, B)^* = \mathcal{O}(A^*, B^*)$$

and introducing the adjoint system

$$\dot{x} = A^*\tilde{x} \tag{57}$$

$$\tilde{y} = B^*\tilde{x} \tag{58}$$

we arrive to the following duality principle.

Theorem 19 *A time invariant linear system is controllable if and only if its adjoint system is observable.*

Notice that the observability of the adjoint system is easily shown to be equivalent to the existence of a constant $\delta > 0$ such that

$$\int_0^T ||B^*e^{tA^*}x_0||^2 dt \geq \delta||x_0||^2 \quad \forall x_0 \in \mathbb{R}^n . \tag{59}$$

As it has been pointed out in Sect. "Linear Systems", the equivalence between the controllability of a system and the observability of its adjoint, expressed in the form (59), is still true in infinite dimension, and provides a very useful way to investigate the controllability of a partial differential equation.

Finally, using the duality principle, we see that any controllability test gives rise to an observability test, and vice-versa.

Controllability and Stabilizability

In this section we shall explore the connections between the controllability and the stabilizability of a system. Let us begin with a linear control system

$$\dot{x} = Ax + Bu .\tag{60}$$

Performing a linear change of coordinates if needed, we may assume that (60) has the block structure given by the Kalman controllability decomposition

$$\begin{pmatrix} \dot{x}_1 \\ \dot{x}_2 \end{pmatrix} = \begin{pmatrix} A_1 & A_2 \\ 0 & A_3 \end{pmatrix} \begin{pmatrix} x_1 \\ x_2 \end{pmatrix} + \begin{pmatrix} B_1 \\ 0 \end{pmatrix} u \tag{61}$$

where $A_1 \in \mathbb{R}^{r \times r}$, $B_1 \in \mathbb{R}^{r \times m}$, and where $\operatorname{rank} \mathcal{R}$ $(A_1, B_1) = r$. For any square matrix $M \in \mathbb{C}^{n \times n}$ we denote by $\sigma(M)$ its spectrum, i. e.

$$\sigma(M) = \{\lambda \in \mathbb{C} ; \det(M - \lambda I) = 0\} .$$

Let us note $\mathbb{C}_- := \{\lambda \in \mathbb{C} ; \operatorname{Re} \lambda < 0\}$. Then the asymptotic stabilizability of (60) may be characterized as follows.

Theorem 20 *There exists a (continuous) asymptotically stabilizing feedback law $u = k(x)$ with $k(0) = 0$ for (60) if and only if $\sigma(A_3) \subset \mathbb{C}_-$. If it is the case, then for any family $S = (\lambda_i)_{1 \le i \le r}$ of elements of $\mathbb{C}_-$ invariant by conjugation, there exists a linear asymptotically stabilizing feedback law $u(x) = Kx$, with $K \in \mathbb{R}^{m \times n}$, such that*

$$\sigma(A + BK) = \sigma(A_3) \cup S .\tag{62}$$

The property (62), which shows to what extend the spectrum of the closed loop system can be assigned, is referred to as the **pole shifting theorem**.

As a direct consequence of Theorem 20, we obtain the following result.

Corollary 21 *A time invariant linear system which is controllable is also asymptotically stabilizable.*

It is natural to ask whether Corollary 21 is still true for a nonlinear system, i. e. if a controllable system is necessarily asymptotically stabilizable. A general result cannot be obtained, as is shown by Brockett's system.

Example 2 continued Brockett's system (30)–(32) may be written $\dot{x} = u_1 f_1(x) + u_2 f_2(x) = f(x, u)$, and it follows from Theorem 13 that (30)–(32) is controllable around $(x_0, u_0) = (0, 0)$. Let us now recall a necessary condition for stabilizability due to R. Brockett [6].
Brockett third condition for stabilizability. Let $f \in C(\mathbb{R}^n \times \mathbb{R}^m ; \mathbb{R}^n)$ with $f(x_0, u_0) = 0$. If the control system $\dot{x} = f(x, u)$ can be locally asymptotically stabilized at x_0 by means of a continuous feedback law u satisfying

$u(x_0) = u_0$, then the image by f of any open neighborhood of $(x_0, u_0) \in \mathbb{R}^n \times \mathbb{R}^m$ contains an open neighborhood of $0 \in \mathbb{R}^n$.

Here, for any neighborhood V of (x_0, u_0) in $\mathbb{R}^5$, $f(V)$ does not cover an open neighborhood of 0 in $\mathbb{R}^3$, since $(0, 0, \varepsilon) \notin f(V)$ for any $\varepsilon \in \mathbb{R}$. According to Brockett's condition, the system (30)–(32) is not asymptotically stabilizable at the origin.

Thus a controllable system may fail to be asymptotically stabilizable by a continuous feedback law $u = k(x)$ due to topological obstructions. It turns out that this phenomenon does not occur when the phase space is the plane, as it has been demonstrated by Kawski in [21].

Theorem 22 *(Kawski) Let f_0, f_1 be real analytic vector fields on $\mathbb{R}^2$ with $f_0(0) = 0$ and $f_1(0) \neq 0$. Assume that the system*

$$\dot{x} = f_0(x) + u f_1(x) , \qquad u \in \mathbb{R} \tag{63}$$

is small time locally controllable at the origin. Then it is also asymptotically stabilizable at the origin by a Hölder continuous feedback law $u = k(x)$.

In larger dimension ($n \ge 3$), a way to go round the topological obstruction is to consider a time-varying feedback law $u = k(x, t)$. It has been first observed by Sontag and Sussmann in [33] that for one-dimensional state and input ($n = m = 1$), the controllability implies the asymptotic stabilizability by means of a time-varying feedback law. This kind of stabilizability was later established by Samson in [31] for Brockett's system ($n = 3$ and $m = 2$). Finally, using the return method, Coron proved that the implication
Controllability $\Rightarrow$ Asymptotic Stabilizability by Time-Varying Feedback
was a principle verified in most cases. To state precise results, we consider affine systems

$$\dot{x} = f_0(x) + \sum_{i=1}^{m} u_i f_i(x) , \qquad x \in \mathbb{R}^n$$

and distinguish again two cases: (1) $f_0 \equiv 0$ (no drift); (2) $f_0 \not\equiv 0$ (a drift).
(1) System without drift

Theorem 23 *(Coron [9]) Assume that (43) holds for the system (41). Pick any number $T > 0$. Then there exists a feedback law $u = (u_1, \ldots, u_m) \in C^\infty(\mathbb{R}^n \times \mathbb{R}, \mathbb{R}^m)$ such that*

$$u(0, t) = 0 \quad \forall t \in \mathbb{R} \tag{64}$$

$$u(x, t + T) = u(x, t) \quad \forall(x, t) \in \mathbb{R}^n \times \mathbb{R} \tag{65}$$

and such that 0 is globally asymptotically stable for the system

$$\dot{x} = \sum_{i=1}^{m} u_i(x, t) f_i(x) .$$

(2) System with a drift

Theorem 24 *(Coron* [10]*) Assume that the system* (45) *satisfies the condition* $S(\theta)$ *for some* $\theta \in [0, 1]$. *Assume also that* $n \notin \{2, 3\}$ *and that*

$$\text{Lie}(f_0, f_1, \dots, f_m)(0) = \mathbb{R}^n .$$

Pick any $T > 0$. *Then there exists a feedback law* $u = (u_1, \dots, u_m) \in C^0(\mathbb{R}^n \times \mathbb{R}, \mathbb{R}^m)$ *with* $u \in C^\infty((\mathbb{R}^n \setminus \{0\}) \times \mathbb{R}, \mathbb{R}^m)$ *such that* (64)–(65) *hold and such that 0 is locally asymptotically stable for the system*

$$\dot{x} = f_0(x) + \sum_{i=1}^{m} u_i(x, t) f_i(x) .$$

Flatness

While powerful criterions enable us to decide whether a control system is controllable or not, most of them do not provide any indication on how to design an explicit control input steering the system from a point to another one. Fortunately, there exists a large class of systems, the so-called **flat systems**, for which explicit control inputs may easily be found. The flatness theory has been introduced by Fliess, Levine, Martin, and Rouchon in [15], and since then it has attracted the interest of many researchers thanks to its numerous applications in Engineering. Here, we only sketch the main ideas, referring the interested reader to [27] for a comprehensive introduction to the subject.

Let us consider a smooth control system

$$\dot{x} = f(x, u) , \quad x \in \mathbb{R}^n , u \in \mathbb{R}^m$$

given together with an output $y \in \mathbb{R}^m$ depending on x, u, and a finite number of derivatives of u

$$y = h(x, u, \dot{u}, \dots, u^{(r)}) .$$

Following [15], we shall say that y is a **flat output** if the components of y are differentially independent, and both x and u may be expressed as functions of y and of a finite number of its derivatives

$$x = k(y, \dot{y}, \dots, y^{(p)}) \tag{66}$$

$$u = l(y, \dot{y}, \dots, y^{(q)}) . \tag{67}$$

In (66)–(67), p and q denote some nonnegative integers, and k and l denote some smooth functions. Since the

state x and the input u are parameterized by the flat output y, to solve the controllability problem

$$\dot{x} = f(x, u) \tag{68}$$

$$x(0) = x_0 , \quad x(T) = x_T \tag{69}$$

it is sufficient to pick any function $y \in C^{\max(p,q)}([0, T]; \mathbb{R}^m)$ such that

$$k(y, \dot{y}, \dots, y^{(p)})(0) = x(0) = x_0 \tag{70}$$

$$k(y, \dot{y}, \dots, y^{(p)})(T) = x(T) = x_T . \tag{71}$$

The constraints (70)–(71) are generally very easy to satisfy. The desired control input u is then given by (67). Let us show how this program may be carried out on two simple examples.

Example 5 For the simple integrator

$$\dot{x}_1 = x_2$$
$$\dot{x}_2 = u$$

the output $y = x_1$ is flat, for $x_2 = \dot{y}$ and $u = \ddot{y}$. The output $z = x_2$ is not flat, as $x_1 = \int z(t) \, dt$. To steer the system from $x_0 = (0, 0)$ to $x_T = (1, 0)$ in time T, we only have to pick a function $y \in C^2([0, T]; \mathbb{R})$ such that

$$y(0) = 0 , \quad \dot{y}(0) = 0 , \quad y(T) = 1 , \quad \text{and} \quad \dot{y}(T) = 0 .$$

Clearly, $y(t) = t^2(2T - t)^2/T^4$ is convenient.

Example 6 Consider now the nonlinear system

$$\dot{x}_1 = u_1 \tag{72}$$

$$\dot{x}_2 = u_2 \tag{73}$$

$$\dot{x}_3 = x_2 u_1 . \tag{74}$$

Eliminating the input u_1 in (72)–(74) yields $\dot{x}_3 = x_2 \dot{x}_1$, so that x_2 may be expressed as a function of x_1, x_3 and their derivatives. The same is true for u_2, thanks to (73). Let us pick

$$y = (y_1, y_2) = (x_1, x_3) .$$

We claim that y is a flat output. Indeed,

$$(x_1, x_2, x_3) = (y_1, \frac{\dot{y}_2}{\dot{y}_1}, y_2) , \tag{75}$$

$$(u_1, u_2) = (\dot{y}_1, \frac{\ddot{y}_2 \dot{y}_1 - \dot{y}_2 \ddot{y}_1}{(\dot{y}_1)^2}) . \tag{76}$$

Pick $x_0 = (0, 0, 0)$ and $x_T = (0, 0, 1)$. Notice that, by the mean value theorem, $\dot{y}_1$ has to vanish somewhere, say at $t = \bar{t}$. We shall construct y_1 in such a way that $\dot{y}_1$ vanishes *only* at $t = \bar{t}$. For x_2 not to be singular at $\bar{t}$, we have to impose the condition

$$\dot{y}_2(\bar{t}) = 0 \, .$$

If y_1 and y_2 are both analytic near $\bar{t}$, we notice that $x_2 = \dot{y}_2 / \dot{y}_1$ is analytic near $\bar{t}$, hence $u_2 = \dot{x}_2$ is well-defined (and analytic) near $\bar{t}$. To steer the solution of (72)–(74) from $x_0 = (0, 0, 0)$ to $x_T = (0, 0, 1)$, it is then sufficient to pick a function $y = (y_1, y_2) \in C^\omega([0, T]; \mathbb{R}^2)$ such that

$$y_1(0) = y_2(0) = 0, \quad \dot{y}_2(0) = 0, \quad \text{and } \dot{y}_1(0) \neq 0 \, ,$$
$$y_1(T) = 0, \quad y_2(T) = 1, \quad \dot{y}_2(T) = 0, \quad \text{and } \dot{y}_1(T) \neq 0 \, .$$
$$\dot{y}_1\left(\frac{T}{2}\right) = 0, \quad \dot{y}_2\left(\frac{T}{2}\right) = 0, \quad \ddot{y}_1\left(\frac{T}{2}\right) \neq 0$$

and $\dot{y}_1(t) \neq 0$ for $t \neq \dfrac{T}{2}$.

Clearly,

$$(y_1(t), y_2(t)) = \left(t(T - t), \frac{t^4}{4} - \frac{Tt^3}{2} + \frac{T^2 t^2}{4} \right)$$

is convenient.

Future Directions

Lie algebraic techniques have been used to provide powerful controllability tests for affine systems. However, there is still an important gap between the known necessary conditions (e. g. the Legendre Clebsh condition or its extensions) and the sufficient conditions (e. g. the $S(\theta)$ condition) for the small time local controllability. On the other hand, it has been noticed by Kawski in ([22], Example 6.1) that certain systems can be controlled on small time intervals $[0, T]$ only by using faster switching control variations, the number of switchings tending to infinity as T tends to zero. As for switched systems [2], it is not clear whether purely algebraic conditions be sufficient to characterize the controllability of systems with drift.

Of great interest for applications is the development of methods providing explicit control inputs, or control inputs computed in real time with the aid of some numerical schemes, both for the motion planning and the stabilization issue. The flatness theory seems to be a very promising method, and it has been successfully applied in Engineering. An active research is devoted to filling the gap between necessary and sufficient conditions for the existence of flat outputs, and to extending the theory to partial differential equations.

The control of nonlinear partial differential equations may sometimes be reduced to the control of a family of finite-dimensional systems by means of the Galerkin method [3]. On the other hand, the spatial discretization of partial differential equations by means of finite differences, finite elements, or spectral methods leads in a natural way to the control of finite dimensional systems (see, e. g., [38]). Of great importance is the *uniform boundedness* with respect to the discretization parameter of the $L^2(0, T)$-norms of the control inputs associated with the finite dimensional approximations.

Geometric ideas borrowed from the control of finite dimensional systems (e. g. the return method, the power series expansion, the quasistatic deformation) have been applied to prove the controllability of certain nonlinear partial differential equations whose linearization fails to be controllable. (See [11] for a survey of these techniques.) The Korteweg–de Vries equation provides an interesting example of a partial differential equation whose linearization fails to be controllable for certain lengths of the space domain [29]. However, for these critical lengths the reachable space proves to be of finite codimension, and it may be proved that the full equation is controllable by using the nonlinear term in order to reach the missing directions [7,12].

Bibliography

Primary Literature

1. Agrachev AA, Gamkrelidze RV (1993) Local controllability and semigroups of diffeomorphisms. Acta Appl Math 32:1–57
2. Agrachev A, Liberzon D (2001) Lie-algebraic stability criteria for switched systems. SIAM J Control Optim 40(1):253–269
3. Agrachev A, Sarychev A (2005) Navier–Stokes equations: controllability by means of low modes forcing. J Math Fluid Mech 7(1):108–152
4. Bianchini RM, Stefani G (1986) Sufficient conditions of local controllability. In: Proc. 25th IEEE–Conf. Decision & Control. Athens
5. Bonnard B (1992) Contrôle de l'altitude d'un satellite rigide. RAIRO Autom Syst Anal Control 16:85–93
6. Brockett RW (1983) Asymptotic stability and feedback stabilization. In: Brockett RW, Millman RS, Sussmann HJ (eds) Differential Geometric Control Theory. Progr. Math., vol 27. Birkhäuser, Basel, pp 181–191
7. Cerpa E, Crépeau E (2007) Boundary controllability for the nonlinear Korteweg–de Vries equation on any critical domain
8. Chow WL (1940) Über Systeme von linearen partiellen Differentialgleichungen erster Ordnung. Math Ann 117:98–105
9. Coron JM (1992) Global asymptotic stabilization for controllable systems without drift. Math Control Signals Syst 5: 295–312

10. Coron JM (1995) Stabilization in finite time of locally controllable systems by means of continuous time-varying feedback laws. SIAM J Control Optim 33:804–833

11. Coron JM (2007) Control and nonlinearity, mathematical surveys and monographs. American Mathematical Society, Providence

12. Coron JM, Crépeau E (2004) Exact boundary controllability of a nonlinear KdV equation with critical lengths. J Eur Math Soc 6(3):367–398

13. Curtain RF, Pritchard AJ (1978) Infinite Dimensional Linear Systems Theory. Springer, New York

14. Dolecky S, Russell DL (1977) A general theory of observation and control. SIAM J Control Optim 15:185–220

15. Fliess M, Lévine J, Martin PH, Rouchon P (1995) Flatness and defect of nonlinear systems: introductory theory and examples. Intern J Control 31:1327–1361

16. Hautus MLJ (1969) Controllability and observability conditions for linear autonomous systems. Ned Akad Wetenschappen Proc Ser A 72:443–448

17. Hermes H (1982) Control systems which generate decomposable Lie algebras. J Differ Equ 44:166–187

18. Hermes H, Kawski M (1986) Local controllability of a single-input, affine system. In: Proc. 7th Int. Conf. Nonlinear Analysis. Dallas

19. Hirschorn R, Lewis AD (2004) High-order variations for families of vector fields. SIAM J Control Optim 43:301–324

20. Kalman RE (1960) Contribution to the theory of optimal control. Bol Soc Mat Mex 5:102–119

21. Kawski M (1989) Stabilization of nonlinear systems in the plane. Syst Control Lett 12:169–175

22. Kawski M (1990) High-order small time local controllability. In: Sussmann HJ (ed) Nonlinear controllability and optimal control. Textbooks Pure Appl. Math., vol 113. Dekker, New York, pp 431–467

23. Khodja FA, Benabdallah A, Dupaix C, Kostin I (2005) Null-controllability of some systems of parabolic type by one control force. ESAIM Control Optim Calc Var 11(3):426–448

24. Lions JL (1988) Contrôlabilité exacte, perturbations et stabilisation de systèmes distribués. In: Recherches en Mathématiques Appliquées, Tome 1, vol 8. Masson, Paris

25. Liu K (1997) Locally distributed control and damping for the conservative systems. SIAM J Control Optim 35:1574–1590

26. Lobry C (1970) Contrôlabilité des systèmes non linéaires. SIAM J Control 8:573–605

27. Martin PH, Murray RM, Rouchon P (2002) Flat systems. Mathematical Control Theory, Part 1,2 (Trieste 2001). In: ICTP Lect. Notes, vol VIII. Abdus Salam Int Cent Theoret Phys, Trieste

28. Rashevski PK (1938) About connecting two points of complete nonholonomic space by admissible curve. Uch Zapiski ped inst Libknexta 2:83–94

29. Rosier L (1997) Exact boundary controllability for the Korteweg–de Vries equation on a bounded domain. ESAIM: Control Optim Calc Var 2:33–55

30. Rosier L (2007) A Survey of controllability and stabilization results for partial differential equations. J Eur Syst Autom (JESA) 41(3–4):365–412

31. Samson C (1991) Velocity and torque feedback control of a nonholonomic cart. In: de Witt C (ed) Proceedings of the International workshop on nonlinear and adaptive control: Issues in robotics, Grenoble, France nov. 1990. Lecture Notes in Control and Information Sciences, vol 162. Springer, Berlin, pp 125–151

32. Silverman LM, Meadows HE (1967) Controllability and observability in time-variable linear systems. SIAM J Control 5:64–73

33. Sontag ED, Sussmannn H (1980) Remarks on continuous feedbacks. In: Proc. IEEE Conf. Decision and Control, (Albuquerque 1980). pp 916–921

34. Stefani G (1985) Local controllability of nonlinear systems: An example. Syst Control Lett 6:123–125

35. Sussmann H (1987) A general theorem on local controllability. SIAM J Control Optim 25:158–194

36. Sussmann H, Jurdjevic V (1972) Controllability of nonlinear systems. J Differ Equ 12:95–116

37. Tret'yak AI (1991) On odd-order necessary conditions for optimality in a time-optimal control problem for systems linear in the control. Math USSR Sbornik 79:47–63

38. Zuazua E (2005) Propagation, observation and control of waves approximated by finite difference methods. SIAM Rev 47(2):197–243

Books and Reviews

Agrachev AA, Sachkov YL (2004) Control theory from the geometric viewpoint. In: Encyclopaedia of Mathematical Sciences, vol 87. Springer, Berlin

Bacciotti A (1992) Local stabilization of nonlinear control systems. In: Ser. Adv. Math. Appl. Sci, vol 8. World Scientific, River Edge

Bacciotti A, Rosier L (2005) Liapunov functions and stability in control theory. Springer, Berlin

Isidori A (1989) Nonlinear control systems. Springer, Berlin

Nijmeijer H, van der Schaft AJ (1990) Nonlinear dynamical control systems. Springer, New York

Sastry S (1999) Nonlinear systems. Springer, New York

Sontag E (1990) Mathematical control theory. Springer, New York

Zabczyk J (1992) Mathematical control theory: An introduction. Birkhäuser, Boston

Firing Squad Synchronization Problem in Cellular Automata

HIROSHI UMEO
Electro-Communication, University of Osaka, Osaka, Japan

Article Outline

Glossary

Cellular automaton A cellular automaton is a discrete computational model studied in mathematics, computer science, economics, biology, physics and chemistry etc. It consists of a regular array of cells, each cell is a finite state automaton. The array can be in any finite number of dimensions. Time (step) is also discrete, and the state of a cell at time t (≥ 1) is a function of the states of a finite number of cells (called its neighborhood) at time $t - 1$. Each cell has a same rule set for updating its next state, based on the states in the neighborhood. At every step the rules are applied to the whole array synchronously, yielding a new configuration.

Time-space diagram A time-space diagram is frequently used to represent signal propagations in one-dimensional cellular space. Usually, the time is drawn on the vertical axis and the space on the horizontal axis. The trajectories of individual signals in propagation are expressed in this diagram by sloping lines. The slope of the line represents the propagation speed of the signal. Time-space-diagrams that show the position of individual signals in time and in space are very useful for understanding cellular algorithms, signal propagations and crossings in the cellular space.

Definition of the Subject

The firing squad synchronization problem (**FSSP** for short) is formalized in terms of the model of cellular automata. Figure 1 shows a finite one-dimensional cellular array consisting of n cells, denoted by C_i, where $1 \leq i \leq n$. All cells (except the end cells) are identical finite state automata. The array operates in lock-step mode such that the next state of each cell (except the end cells) is determined by both its own present state and the present states of its right and left neighbors. All cells (*soldiers*), except the left end cell, are initially in the *quiescent* state at time $t = 0$ and have the property whereby the next state of a quiescent cell having quiescent neighbors is the quiescent state. At time $t = 0$ the left end cell (*general*) is in the *fire-when-ready* state, which is an initiation signal to the array.

The firing squad synchronization problem is stated as follows: Given an array of n identical cellular automata, including a *general* on the left end which is activated at time $t = 0$, one wants to give the description (state set and next-state transition function) of the automata so that, *at some future time*, all of the cells will *simultaneously* and, *for the first time*, enter a special *firing* state. The set of states and the next-state transition function must be independent of n. Without loss of generality, it is assumed
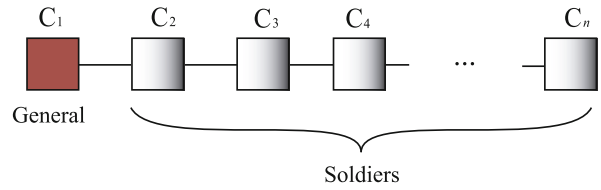
Firing Squad Synchronization Problem in Cellular Automata, Figure 1
One-dimensional cellular automaton

that $n \geq 2$. The tricky part of the problem is that the same kind of soldiers having a fixed number of states must be synchronized, regardless of the length n of the array. The problem itself is interesting as a mathematical puzzle and a good example of recursive divide-and-conquer strategy operating in parallel. It has been referred to as *achieving macro-synchronization given in micro-synchronization systems* and realizing a *global synchronization using only local information exchange* [10].

Introduction

Cellular automata are considered to be a nice model of complex systems in which an infinite one-dimensional array of finite state machines (cells) updates itself in synchronous manner according to a uniform local rule. A comprehensive study is made for a synchronization problem that gives a finite-state protocol for synchronizing a large scale of cellular automata. Synchronization of general network is a computing primitive of parallel and distributed computations. The synchronization in cellular automata has been known as *firing squad synchronization problem* since its development, in which it was originally proposed by J. Myhill in Moore [29] to synchronize all parts of self-reproducing cellular automata. The problem has been studied extensively for more than 40 years [1–80].

The present article firstly examines the state transition rule sets for the famous firing squad synchronization algorithms that give a finite-state protocol for synchronizing large-scale cellular automata, focusing on the fundamental synchronization algorithms operating in optimum steps on one-dimensional cellular arrays. The algorithms discussed herein are the Goto's first algorithm [12], the eight-state Balzer's algorithm [1], the seven-state Gerken's algorithm [9], the six-state Mazoyer's algorithm [25], the 16-state Waksman's algorithm [74] and a number of revised versions thereof. In addition, the article constructs a survey of current optimum-time synchronization algorithms and compares their transition rule sets with respect

to the number of internal states of each finite state au-
tomaton, the number of transition rules realizing the syn-
chronization, and the number of state-changes on the ar-
ray. It also presents herein a survey and a comparison of
the quantitative and qualitative aspects of the optimum-
time synchronization algorithms developed thus far for
one-dimensional cellular arrays. Then, it provides sev-
eral variants of the firing squad synchronization prob-
lems including fault-tolerant synchronization protocols,
one-bit communication protocols, non-optimum-time al-
gorithms, and partial solutions etc. Finally, a survey on
two-dimensional firing squad synchronization algorithms
is presented. Several new results and viewpoints are also
given.

Firing Squad Synchronization Problem

A Brief History of the Developments of Firing Squad Synchronization Algorithms

The problem known as the *firing squad synchronization
problem* was devised in 1957 by J. Myhill, and first ap-
peared in print in a paper by E.F. Moore [29]. This prob-
lem has been widely circulated, and has attracted much
attention. The firing squad synchronization problem first
arose in connection with the need to simultaneously turn
on all parts of a self-reproducing machine. The problem
was first solved by J. McCarthy and M. Minsky [28] who
presented a non-optimum-time synchronization scheme
that operates in $3n+O(1)$ steps for synchronizing n cells.
In 1962, the first optimum-time, i. e. $(2n - 2)$-step, syn-
chronization algorithm was presented by Goto [12], with
each cell having several thousands of states. Waksman [74]
presented a 16-state optimum-time synchronization algo-
rithm. Afterward, Balzer [1] and Gerken [9] developed
an eight-state algorithm and a seven-state algorithm, re-
spectively, thus decreasing the number of states required
for the synchronization. In 1987, Mazoyer [25] developed
a six-state synchronization algorithm which, at present, is
the algorithm having the fewest states.

Firing Squad Synchronization Algorithm

Section "Firing Squad Synchronization Algorithm" briefly
sketches the design scheme for the firing squad synchro-
nization algorithm according to Waksman [74] in which
the first transition rule set was presented. It is quoted from
Waksman [74].

The code book of the state transitions of machines
is so arranged to cause the array to progressively
divide itself into 2^k equal parts, where k is an in-
teger and an increasing function of time. The end
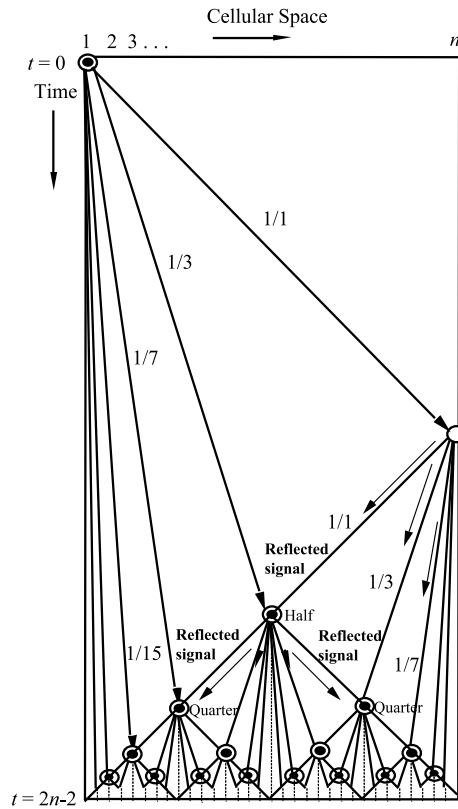
**Firing Squad Synchronization Problem in Cellular Automata, Fig-
ure 2**
**Time-space diagram for Waksman's optimum-time firing squad
synchronization algorithm**

machines in each partition assume a special state so
that when the last partition occurs, all the machines
have for both neighbors machines at this state. This
is made the only condition for any machine to as-
sume terminal state.

Figure 2 is a time-space diagram for the Waksman's
optimum-step firing squad synchronization algorithm.
The general at time $t = 0$ emits an infinite number of sig-
nals which propagate at $1/(2^{k+1} - 1)$ speed, where k is
positive integer. These signals meet with a reflected sig-
nal at half point, quarter points, ..., etc., denoted by $\odot$ in
Fig. 2. It is noted that these cells indicated by $\odot$ are syn-
chronized. By increasing the number of synchronized cells
exponentially, eventually all of the cells are synchronized.

Complexity Measures and Properties in Firing Squad Synchronization Algorithms

Time Complexity   Any solution to the firing squad syn-
chronization problem can easily be shown to require

$(2n-2)$-steps for synchronizing n cells, since signals on the array can propagate no faster than one cell per step, and the time from the general's instruction until the synchronization must be at least $2n-2$. See Balzer [1], Goto [12] and Waksman [74] for a proof. The next two theorems show the optimum-time complexity for synchronizing n cells on one-dimensional arrays.

Theorem 1 ([1,12,74]) *Synchronization of n cells in less than $(2n-2)$-steps is impossible.*

Theorem 2 ([1,12,74]) *Synchronization of n cells at exactly $(2n-2)$-steps is possible.*

Number of States   The following three distinct states: the *quiescent* state, the *general* state, and the *firing* state, are required in order to define any cellular automaton that can solve the firing squad synchronization problem. The boundary state for C_0 and C_{n+1} is not generally counted as an internal state. Balzer [1] implemented a search strategy in order to prove that there exists no four-state solution. He showed that no four-state optimum-time solution exists. Sanders [40] studied a similar problem on a parallel computer and showed that the Balzer's backtrack heuristic was not correct, rendering the proof incomplete and gave a proof based on a computer simulation for the non-existence of four-state solution. Balzer [1] also showed that there exist no five-state optimum-time solution satisfying special conditions. It is noted that the Balzer's special conditions do not hold for the Mazoyer's six-state solution with the fewest states known at present. The question that remains is: "What is the minimum number of states for an optimum-time solution of the problem?" At present, that number is *five* or *six*. Section "Partial Solutions" gives some 4- and 5-state partial solutions that can synchronize infinite cells, but not all.

Theorem 3 ([1,40]) *There is no four-state CA that can synchronize n cells.*

Berthiaume, Bittner, Perković, Settle and Simon [2] considered the state lower bound on ring-connected cellular automata. It is shown that there exists no three-state solution and no four-state symmetric solution for rings.

Theorem 4 ([2]) *There is no four-state symmetric optimum-time solution for ring-connected cellular automata.*

Number of Transition Rules   Any k-state transition table for the synchronization has at most $(k-1)k^2$ entries in $(k-1)$ matrices of size $k \times k$. The number of transition rules reflects the complexity of synchronization algorithms.

Transition Rule Sets for Optimum-Time Firing Squad Synchronization Algorithms

Section "Transition Rule Sets for Optimum-Time Firing Squad Synchronization Algorithms" implements most of the transition rule sets for the synchronization algorithms above mentioned on a computer and check whether these rule sets yield successful firing configurations at exactly $t = 2n - 2$ steps for any n such that $2 \le n \le 10,000$.

Waksman's 16-State Algorithm   Waksman [74] proposed a 16-state firing squad synchronization algorithm, which, together with an unpublished algorithm by Goto [12], is referred to as the first-in-the-world optimum-time synchronization algorithm. Waksman presented the first set of transition rules described in terms of a state transition table that is defined on the following state set $\mathcal{D}$ consisting of 16 states such that $\mathcal{D} = \{Q, T, P_0, P_1, B_0, B_1, R_0, R_1, A_{000}, A_{001}, A_{010}, A_{011}, A_{100}, A_{101}, A_{110}, A_{111}\}$, where Q is a quiescent state, T is a firing state, P_0 and P_1 are prefiring states, B_0 and B_1 are states for signals propagating at various speeds, R_0 and R_1 are trigger states which cause the B_0 and B_1 states move in the left or right direction and A_{ijk}, $i, j, k \in \{0, 1\}$ are control states which generate the state R_0 or R_1 either with a unit delay or without any delay. The state P_0 also acts as an initial general.

USN Transition Rule Set   Cellular automata researchers have reported that some errors are included in the Waksman's transition table. A computer simulation made in Umeo, Sogabe and Nomura [64] reveals this to be true. They corrected some errors included in the original Waksman's transition rule set. The correction procedures can be found in Umeo, Sogabe and Nomura [64]. This subsection gives a complete list of the transition rules which yield successful synchronizations for any n. Figure 3 is the complete list, which consists of 202 transition rules. The list is referred to as the *USN transition rule set*. In the correction, a ninety-three percent reduction in the number of transition rules is realized compared to the Waksman's original list. The computer simulation based on the table of Fig. 3 gives the following observation. Figure 4 shows snapshots of the Waksman's 16-state optimum-time synchronization algorithm on 21 cells.

Observation 3.1 ([54,64]) *The set of rules given in Fig. 3 is the smallest transition rule set for Waksman's optimum-time firing squad synchronization algorithm.*

Balzer's Eight-State Algorithm   Balzer [1] constructed an eight-state, 182-rule synchronization algorithm and

——— Q State ———				
1: Q	Q	Q	→	Q
2: Q	Q	B0	→	Q
3: Q	Q	Bl	→	Q
4: Q	Q	R0	→	R0
5: Q	Q	Rl	→	Q
6: Q	Q	P0	→	A000
7: Q	Q	P1	→	A100
8: Q	Q	A000	→	A001
9: Q	Q	A001	→	A000
10: Q	Q	A100	→	A101
11: Q	Q	A101	→	A100
12: Q	Q	A010	→	R0
13: Q	Q	A110	→	Q
14: Q	Q	*	→	Q
15: B0	Q	Q	→	Q
16: B0	Q	B0	→	Q
17: B0	Q	R0	→	R0
18: B0	Q	P1	→	A100
19: B0	Q	A000	→	A001
20: B0	Q	A101	→	A100
21: B0	Q	A010	→	R0
22: B0	Q	A110	→	Q
23: B1	Q	Q	→	Q
24: B1	Q	Bl	→	Q
25: B1	Q	R0	→	R0
26: B1	Q	Rl	→	Q
27: B1	Q	P0	→	A000
28: B1	Q	A001	→	A000
29: B1	Q	A100	→	A101
30: R0	Q	Q	→	Q
31: R0	Q	Bl	→	Q
32: R0	Q	P0	→	A000
33: R0	Q	A000	→	A001
34: R0	Q	A001	→	A000
35: R0	Q	A011	→	Q
36: R1	Q	Q	→	R1
37: R1	Q	B0	→	R1
38: R1	Q	Bl	→	R1
39: P0	Q	Q	→	A010
40: P0	Q	Bl	→	A010
41: P0	Q	Rl	→	A010
42: P0	Q	*	→	P1
43: P1	Q	Q	→	A110
44: P1	Q	B0	→	A110
45: A000	Q	Q	→	R1
46: A000	Q	B0	→	R1
47: A001	Q	Rl	→	Q
48: A100	Q	Q	→	Q
49: A100	Q	B0	→	Q
50: A010	Q	Q	→	A011
51: A010	Q	B0	→	A011
52: A010	Q	Rl	→	A011
53: A010	Q	*	→	P0
54: A011	Q	Q	→	A010
55: A011	Q	Bl	→	A010
56: A011	Q	Rl	→	A010
57: A011	Q	*	→	P1
58: A110	Q	Q	→	A111
59: A110	Q	Bl	→	A111
60: A111	Q	Q	→	A110

61: A111	Q	B0	→	A110
——— B0 State ———				
62: Q	B0	Q	→	B0
63: Q	B0	R0	→	R0
64: Q	B0	P0	→	B0
65: Q	B0	P1	→	B0
66: Q	B0	A001	→	P1
67: Q	B0	A100	→	P1
68: B1	B0	P0	→	B0
69: B1	B0	P1	→	B0
70: R1	B0	Q	→	R1
71: R1	B0	P0	→	R1
72: R1	B0	P1	→	R1
73: P0	B0	Q	→	B0
74: P0	B0	B1	→	B0
75: P0	B0	R0	→	R0
76: P0	B0	P0	→	P0
77: P0	B0	P1	→	P0
78: P0	B0	A100	→	P1
79: P0	B0	A011	→	B0
80: P1	B0	Q	→	B0
81: P1	B0	B1	→	B0
82: P1	B0	R0	→	R0
83: P1	B0	P0	→	P0
84: P1	B0	A100	→	P1
85: A001	B0	P0	→	B0
86: A011	B0	Q	→	P1
87: A110	B0	Q	→	P1
88: A110	B0	P0	→	P1
89: A110	B0	P1	→	P1
——— B1 State ———				
90: Q	Bl	Q	→	B1
91: Q	Bl	B0	→	B1
92: Q	Bl	R0	→	Q
93: Q	Bl	R1	→	B1
94: Q	Bl	A000	→	P0
95: Q	Bl	A101	→	P0
96: B0	B1	Q	→	B1
97: B0	B1	R0	→	Q
98: B0	B1	A000	→	P0
99: B0	B1	A101	→	P0
100: R0	B1	Q	→	B1
101: R0	B1	A000	→	P0
102: R1	B1	Q	→	Q
103: R1	B1	B0	→	Q
104: A010	B1	Q	→	P0
105: A010	B1	B0	→	P0
106: A010	B1	R1	→	P0
107: A111	B1	Q	→	P0
108: A111	B1	B0	→	P0
——— R0 State ———				
109: Q	R0	Q	→	Q
110: Q	R0	Bl	→	Q
111: Q	R0	A111	→	Q
112: B0	R0	Q	→	B1
113: B1	R0	Q	→	B0

114: P0	R0	B1	→	B0
115: P1	R0	B1	→	B0
116: P1	R0	A111	→	B0
——— R1 State ———				
117: Q	R1	Q	→	Q
118: Q	R1	B0	→	B1
119: Q	R1	B1	→	B0
120: B1	R1	Q	→	Q
121: B1	R1	P0	→	B0
122: B1	R1	P1	→	B0
123: A101	R1	Q	→	Q
124: A101	R1	P1	→	B0
——— P0 State ———				
125: Q	P0	Q	→	P0
126: Q	P0	P0	→	P0
127: Q	P0	*	→	P0
128: B0	P0	B0	→	P0
129: B0	P0	P0	→	P0
130: B0	P0	*	→	P0
131: R1	P0	R0	→	P0
132: R1	P0	P0	→	P0
133: R1	P0	*	→	P0
134: P0	P0	Q	→	P0
135: P0	P0	B0	→	P0
136: P0	P0	R0	→	P0
137: P0	P0	P0	→	T
138: P0	P0	P1	→	T
139: P0	P0	A010	→	P0
140: P0	P0	*	→	T
141: P1	P0	P0	→	T
142: P1	P0	P1	→	T
143: P1	P0	*	→	T
144: A000	P0	P0	→	P0
145: A000	P0	A010	→	P0
146: A000	P0	*	→	P0
147: *	P0	Q	→	P0
148: *	P0	B0	→	P0
149: *	P0	R0	→	P0
150: *	P0	P0	→	T
151: *	P0	P1	→	T
152: *	P0	A010	→	P0
——— P1 State ———				
153: Q	P1	Q	→	P1
154: Q	P1	P1	→	P1
155: Q	P1	*	→	P1
156: B0	P1	B0	→	P1
157: B0	P1	P1	→	P1
158: B0	P1	*	→	P1
159: R1	P1	R0	→	P1
160: R1	P1	P1	→	P1
161: R1	P1	*	→	P1
162: P0	P1	P0	→	T
163: P0	P1	P1	→	T
164: P0	P1	*	→	T
165: P1	P1	Q	→	P1
166: P1	P1	B0	→	P1

167: P1	P1	R0	→	P1
168: P1	P1	P0	→	T
169: P1	P1	P1	→	T
170: P1	P1	A110	→	P1
171: P1	P1	*	→	T
172: A100	P1	P1	→	P1
173: A100	P1	A110	→	P1
174: A100	P1	*	→	P1
——— A000 State ———				
175: Q	A000	Q	→	Q
176: Q	A000	P0	→	B0
177: B1	A000	Q	→	Q
178: B1	A000	P0	→	B0
——— A001 State ———				
179: Q	A001	Q	→	Q
180: Q	A001	B0	→	Q
181: B0	A001	Q	→	Q
182: B0	A001	B0	→	Q
——— A100 State ———				
183: Q	A100	Q	→	R1
184: Q	A100	P1	→	R1
185: B0	A100	Q	→	P1
186: B0	A100	P1	→	P1
——— A101 State ———				
187: Q	AI01	R1	→	Q
188: B1	A101	R1	→	P0
——— A010 State ———				
189: Q	A010	Q	→	Q
190: Q	A010	B1	→	Q
191: P0	A010	Q	→	B0
192: P0	A010	B1	→	B0
——— A011 State ———				
193: Q	A011	Q	→	Q
194: Q	A011	B0	→	Q
195: B0	A011	Q	→	Q
196: B0	A011	B0	→	Q
——— A110 State ———				
197: Q	AI10	Q	→	R0
198: Q	A110	B0	→	P1
199: P1	A110	Q	→	R0
200: P1	A110	B0	→	P1
——— A111 State ———				
201: R0	A111	Q	→	Q
202: R0	A111	B1	→	P0

Firing Squad Synchronization Problem in Cellular Automata, Figure 3
USN transition table consisting of 202 rules that realize Waksman's synchronization algorithm. The symbol ∗ represents the boundary state

the structure of which is completely identical to that of Waksman [74]. A computer examination made by Umeo, Hisaoka and Sogabe [54] revealed no errors, however, 17 rules were found to be redundant. Figure 5 gives a list of transition rules for Balzer's algorithm and snapshots for synchronization operations on 28 cells. Those redundant rules are indicated by shaded squares. In the transition table, the symbols "M", "L", "F" and "X" represent the general, quiescent, firing and boundary states, respectively.

Noguchi [34] also constructed an eight-state, 119-rule optimum-time synchronization algorithm.

Gerken's Seven-State Algorithm Gerken [9] constructed a seven-state, 118-rule synchronization algorithm. In the computer examination, no errors were found, however, 13 rules were found to be redundant. Figure 6 gives a list of the transition rules for Gerken's algorithm and snapshots for synchronization operations on 28 cells. The

	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	21
0	P0	Q	Q	Q	Q	Q	Q	Q	Q	Q	Q	Q	Q	Q	Q	Q	Q	Q	Q	Q	Q
1	P0	A01	Q	Q	Q	Q	Q	Q	Q	Q	Q	Q	Q	Q	Q	Q	Q	Q	Q	Q	Q
2	P0	B0	A01	Q	Q	Q	Q	Q	Q	Q	Q	Q	Q	Q	Q	Q	Q	Q	Q	Q	Q
3	P0	B0	Q	A01	Q	Q	Q	Q	Q	Q	Q	Q	Q	Q	Q	Q	Q	Q	Q	Q	Q
4	P0	B0	R0	Q	A01	Q	Q	Q	Q	Q	Q	Q	Q	Q	Q	Q	Q	Q	Q	Q	Q
5	P0	R0	B1	Q	Q	A01	Q	Q	Q	Q	Q	Q	Q	Q	Q	Q	Q	Q	Q	Q	Q
6	P0	B0	B1	Q	R0	Q	A01	Q	Q	Q	Q	Q	Q	Q	Q	Q	Q	Q	Q	Q	Q
7	P0	B0	B1	R0	Q	Q	Q	A01	Q	Q	Q	Q	Q	Q	Q	Q	Q	Q	Q	Q	Q
8	P0	B0	Q	B0	Q	Q	R0	Q	A01	Q	Q	Q	Q	Q	Q	Q	Q	Q	Q	Q	Q
9	P0	B0	Q	B0	Q	R0	Q	Q	Q	A01	Q	Q	Q	Q	Q	Q	Q	Q	Q	Q	Q
10	P0	B0	Q	B0	R0	Q	Q	Q	Q	Q	A01	Q	Q	Q	Q	Q	Q	Q	Q	Q	Q
11	P0	B0	Q	R0	B1	Q	Q	R0	Q	Q	Q	A01	Q	Q	Q	Q	Q	Q	Q	Q	Q
12	P0	B0	R0	Q	B1	Q	R0	Q	Q	Q	R0	Q	A01	Q	Q	Q	Q	Q	Q	Q	Q
13	P0	R0	B1	Q	B1	R0	Q	Q	Q	R0	Q	Q	Q	A01	Q	Q	Q	Q	Q	Q	Q
14	P0	B0	B1	Q	Q	B0	Q	Q	R0	Q	Q	Q	R0	Q	A01	Q	Q	Q	Q	Q	Q
15	P0	B0	B1	Q	Q	B0	Q	R0	Q	Q	Q	R0	Q	Q	Q	A01	Q	Q	Q	Q	Q
16	P0	B0	B1	Q	Q	B0	R0	Q	Q	Q	R0	Q	Q	Q	R0	Q	A01	Q	Q	Q	Q
17	P0	B0	B1	Q	R0	B1	Q	Q	R0	Q	Q	R0	Q	Q	Q	A01	Q	Q	Q	Q	Q
18	P0	B0	B1	Q	R0	B1	Q	R0	Q	Q	R0	Q	Q	Q	R0	Q	A01	Q	Q	Q	Q
19	P0	B0	B1	R0	Q	Q	B1	R0	Q	Q	R0	Q	Q	R0	Q	Q	Q	A01	Q	Q	Q
20	P0	B0	Q	B0	Q	Q	B0	Q	Q	R0	Q	Q	R0	Q	Q	Q	R0	Q	Q	R0	P0
21	P0	B0	Q	B0	Q	Q	B0	Q	R0	Q	Q	R0	Q	Q	Q	R0	Q	Q	R0	A00	P0
22	P0	B0	Q	B0	Q	Q	B0	R0	Q	Q	R0	Q	Q	Q	R0	Q	Q	R0	A00	B0	P0
23	P0	B0	Q	B0	Q	Q	R0	B1	Q	R0	Q	Q	Q	R0	Q	Q	R0	A00	Q	B0	P0
24	P0	B0	Q	B0	Q	Q	R0	Q	B1	Q	R0	Q	Q	R0	Q	Q	R0	A00	Q	R1	B0	P0
25	P0	B0	Q	B0	Q	R0	Q	Q	B1	R0	Q	Q	R0	Q	Q	R0	A00	Q	B1	R1	P0
26	P0	B0	Q	B0	R0	Q	Q	Q	B0	Q	Q	R0	Q	Q	R0	A00	Q	R1	Q	B1	B0	P0
27	P0	B0	Q	R0	B1	Q	Q	Q	B0	Q	R0	Q	Q	R0	A00	Q	Q	R1	B1	B0	P0
28	P0	B0	R0	Q	B1	Q	Q	Q	B0	R0	Q	A00	Q	R1	Q	Q	B0	Q	B0	P0
29	P0	R0	B1	Q	B1	Q	Q	Q	R0	B1	A00	Q	Q	R1	Q	B0	Q	B0	P0
30	P0	B0	B1	Q	B1	Q	Q	R0	Q	P0	Q	P1	Q	R1	Q	Q	R1	B0	Q	B0	P0
31	P0	B0	B1	Q	B1	Q	R0	Q	A00	P0	A01	Q	R1	Q	Q	B1	R1	Q	B0	P0
32	P0	B0	B1	Q	B1	Q	R0	A00	B0	P0	B0	A01	Q	R1	Q	B1	Q	R1	B0	P0
33	P0	B0	B1	Q	B1	R0	A00	Q	B0	P0	B0	Q	A01	Q	R1	B1	Q	B1	R1	P0
34	P0	B0	B1	Q	Q	B0	A00	Q	R1	B0	P0	B0	R0	Q	A01	B0	Q	B1	B0	P0
35	P0	B0	B1	Q	Q	P1	Q	Q	B1	R1	P0	R0	B1	Q	Q	P1	Q	Q	B1	B0	P0
36	P0	B0	B1	A10	P1	A11	Q	B1	B0	P0	B0	B1	Q	A10	P1	A11	Q	B1	B0	P0
37	P0	B0	B1	A10	R1	P1	R0	A11	B1	B0	P0	B0	B1	A10	R1	P1	R0	A11	B1	B0	P0
38	P0	B0	P0	P0	B0	P1	B0	P0	P0	B0	P0	B0	P0	P0	B0	P1	B0	P0	P0	B0	P0
39	P0	P0	P0	P0	P0	P1	P0	P0	P0	P0	P0	P0	P0	P0	P1	P0	P0	P0	P0	P0
40	T	T	T	T	T	T	T	T	T	T	T	T	T	T	T	T	T	T	T	T	T

Firing Squad Synchronization Problem in Cellular Automata, Figure 4

Snapshots of the Waksman's 16-state optimum-time synchronization algorithm on 21 cells

13 redundant rules are marked by shaded squares in the table. The symbols ">", "/", "…" and "#" represent the general, quiescent, firing and boundary states, respectively. The symbol "…" is replaced by "F" in the configuration (right) at time $t = 54$.

Mazoyer's Six-State Algorithm   Mazoyer [25] proposed a six-state, 120-rule synchronization algorithm, the structure of which differs greatly from the previous three algorithms discussed above. The computer examination re-

vealed no errors and only one redundant rule. Figure 7 presents a list of transition rules for Mazoyer's algorithm and snapshots of configurations on 28 cells. In the transition table, the letters "G", "L", "F" and "X" represent the general, quiescent, firing and boundary states, respectively.

Goto's Algorithm   The first synchronization algorithm presented by Goto [12] was not published as a journal paper. According to Prof. Goto, the original note Goto [12] is now unavailable, and the only existing material that treats the algorithm is Goto [13]. The Goto's study presents one figure (Fig. 3.8 in Goto [13]) demonstrating how the algorithm works on 13 cells with a very short description in Japanese. Umeo [50] reconstructed the Goto's algorithm based on this figure. Mazoyer [27] also reconstructed this algorithm again. The algorithm that Umeo [50] reconstructed is a non-recursive algorithm consisting of a marking phase and a $3n$-step synchronization phase. In the first phase, by printing a special marker in the cellular space, the entire cellular space is divided into many smaller subspaces, the lengths of which increase exponentially with a common ratio of two, that is 2^j, for any integer j such that $1 \leq j \leq \lfloor \log_2 n \rfloor - 1$. The marking is made from both the left and right ends. In the second phase, each subspace is synchronized using a well-known conventional $3n$-step simple synchronization algorithm. A time-space diagram of the reconstructed algorithm is shown in Fig. 8.

Gerken's 155-State Algorithm   Gerken [9] constructed two kinds of optimum-time synchronization algorithms. One seven-state algorithm has been discussed in the previous subsection, and the other is a 155-state algorithm having $\Theta(n \log n)$ state-change complexity. The transition table given in Gerken [9] is described in terms of two-layer construction with 32 states and 347 rules. An expansion of the transition table into a single-layer format yields a 155-state table consisting of 2371 rules. Figure 9 shows a configuration on 28 cells.

State Change Complexity

Vollmar [73] introduced a state-change complexity in order to measure the efficiency of cellular algorithms and showed that $\Omega(n \log n)$ state-changes are required for the synchronization of n cells in $(2n - 2)$ steps.

Theorem 5 ([73]) $\Omega(n \log n)$ *state-change is necessary for synchronizing n cells in $(2n - 2)$ steps.*

Theorem 6 ([9,54])   *Each optimum-time synchronization algorithm developed by Balzer [1], Gerken [9], Ma-*

Transition table A

A	A	B	C	L	M	Q	R	X
A	A			A			A	
B	C			C	R		C	
C	C			C	Q		C	
L	L						L	
M	B			B			B	
Q	A			A	Q		A	
R	L						L	
X								

Transition table B

B	A	B	C	L	M	Q	R	X
A								
B	Q	B	B	R	A	Q	R	
C	Q	B	B	R	A	Q	R	
L								
M	L		C			C		
Q								
R	Q	B	B		A	Q		
X								

Transition table C

C	A	B	C	L	M	Q	R	X
A								
B			C	R	R		M	C
C	M		C	R	R		M	C
L	L						L	
M	L			C	C	M	M	C
Q								
R			C	B	B		M	C
X								

Transition table L

L	A	B	C	L	M	Q	R	X
A				A		Q		
B			R		Q			
C	C		R	C	Q	C		M
L	L			L		L		
M	C		C	C	M	C		M
Q	L			L		L		
R			B	A	A	Q		
X								

Transition table Q

Q	A	B	C	L	M	Q	R	X
A	Q			Q		Q		
B					R	M	R	
C	M			M	M	M	R	
L	Q			Q	Q	Q		
M					M			
Q	L			A	Q	L		
R	L			A	Q	L		
X								

Transition table M

M	A	B	C	L	M	Q	R	X
A		M	M		M			M
B				M		M	M	M
C				M	M	M	M	M
L	M			M		M		
M	M	M	M	M	F	M		F
Q		M	M		M			M
R				M	M		M	M
X	M	M	M	M	F	M		F

Transition table R

R	A	B	C	L	M	Q	R	X
A								
B			R		Q	Q	R	
C	C	R	C	Q	Q	R		
L								
M	C			B	M	C		
Q	L			A	Q	L		
R		B	R	B	Q	Q	R	
X						M	M	R

Snapshots for synchronization operations on 28 cells:

	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	21	22	23	24	25	26	27	28
0	M	L	L	L	L	L	L	L	L	L	L	L	L	L	L	L	L	L	L	L	L	L	L	L	L	L	L	L
1	M	C	L	L	L	L	L	L	L	L	L	L	L	L	L	L	L	L	L	L	L	L	L	L	L	L	L	L
2	M	C	C	L	L	L	L	L	L	L	L	L	L	L	L	L	L	L	L	L	L	L	L	L	L	L	L	L
3	M	C	R	C	L	L	L	L	L	L	L	L	L	L	L	L	L	L	L	L	L	L	L	L	L	L	L	L
4	M	C	R	B	C	L	L	L	L	L	L	L	L	L	L	L	L	L	L	L	L	L	L	L	L	L	L	L
5	M	C	C	B	R	C	L	L	L	L	L	L	L	L	L	L	L	L	L	L	L	L	L	L	L	L	L	L
6	M	C	C	R	R	B	C	L	L	L	L	L	L	L	L	L	L	L	L	L	L	L	L	L	L	L	L	L
7	M	C	C	R	B	B	R	C	L	L	L	L	L	L	L	L	L	L	L	L	L	L	L	L	L	L	L	L
8	M	C	C	C	B	R	R	B	C	L	L	L	L	L	L	L	L	L	L	L	L	L	L	L	L	L	L	L
9	M	C	R	C	R	R	B	B	R	C	L	L	L	L	L	L	L	L	L	L	L	L	L	L	L	L	L	L
10	M	C	R	C	R	B	B	R	R	B	C	L	L	L	L	L	L	L	L	L	L	L	L	L	L	L	L	L
11	M	C	R	C	C	B	R	R	B	B	R	C	L	L	L	L	L	L	L	L	L	L	L	L	L	L	L	L
12	M	C	R	B	C	R	R	B	B	R	R	B	C	L	L	L	L	L	L	L	L	L	L	L	L	L	L	L
13	M	C	C	B	C	R	B	B	R	R	B	B	R	C	L	L	L	L	L	L	L	L	L	L	L	L	L	L
14	M	C	C	B	C	C	B	R	R	B	B	R	R	B	C	L	L	L	L	L	L	L	L	L	L	L	L	L
15	M	C	C	B	R	C	B	R	R	B	B	R	R	B	B	R	C	L	L	L	L	L	L	L	L	L	L	L
16	M	C	C	R	R	C	R	B	B	R	R	B	B	R	R	B	R	B	C	L	L	L	L	L	L	L	L	L
17	M	C	C	R	R	C	C	B	R	R	B	B	R	R	B	B	R	B	R	C	L	L	L	L	L	L	L	L
18	M	C	C	R	R	B	C	R	R	B	B	R	R	B	B	R	R	B	B	R	B	C	L	L	L	L	L	L
19	M	C	C	R	B	B	C	R	B	B	R	R	B	B	R	R	B	B	R	R	B	B	R	C	L	L	L	L
20	M	C	C	C	B	B	C	C	B	R	R	B	B	R	R	B	B	R	R	B	R	R	B	R	C	L	L	L
21	M	C	R	C	B	B	B	C	R	R	B	B	R	R	B	B	R	R	B	B	R	B	B	R	B	C	L	L
22	M	C	R	C	B	R	R	C	R	B	B	R	R	B	B	R	R	B	B	R	R	B	R	R	B	C	L	L
23	M	C	R	C	R	R	C	C	B	R	R	B	B	R	R	B	B	R	R	B	B	R	C	L	L	L	L	L
24	M	C	R	C	R	R	B	C	R	R	B	B	R	R	B	B	R	R	B	B	R	R	B	C	L	L	L	L
25	M	C	R	C	R	R	B	B	C	R	B	B	R	R	B	B	R	R	B	B	R	R	B	R	C	L	L	L
26	M	C	R	C	R	B	B	B	C	C	B	R	R	B	B	R	R	B	B	R	R	B	R	R	B	C	L	L
27	M	C	R	C	C	B	B	B	R	C	R	R	B	B	R	R	B	B	R	R	B	B	R	R	B	B	R	M
28	M	C	R	B	C	B	B	R	R	C	R	B	B	R	R	B	B	R	R	B	B	R	R	B	B	R	Q	M
29	M	C	C	B	C	B	R	R	R	C	C	B	R	R	B	B	R	R	B	B	R	R	B	B	R	Q	Q	M
30	M	C	C	B	C	R	R	R	R	R	C	B	R	R	B	B	R	R	B	B	R	R	B	Q	L	Q	M
31	M	C	C	B	C	R	R	R	B	B	C	R	B	B	R	R	B	B	R	R	B	B	R	Q	A	L	Q	M
32	M	C	C	B	C	R	R	B	B	B	C	C	B	R	R	B	B	R	R	B	B	R	Q	L	A	Q	Q	M
33	M	C	C	B	C	R	B	B	B	B	R	C	R	R	B	B	R	R	B	B	R	Q	A	L	L	Q	Q	M
34	M	C	C	B	C	C	B	B	B	R	R	C	R	B	B	R	R	B	B	R	Q	L	A	A	L	Q	Q	M
35	M	C	C	B	R	C	B	B	R	R	C	C	B	R	R	B	B	R	Q	A	L	L	A	Q	Q	Q	M
36	M	C	C	R	R	C	B	R	R	R	R	C	R	R	B	B	R	Q	L	A	A	L	L	Q	L	Q	M
37	M	C	C	R	R	C	R	R	R	R	B	B	C	R	B	B	R	Q	A	L	L	A	A	L	Q	L	Q	M
38	M	C	C	R	R	C	R	R	R	B	B	B	C	C	B	R	Q	L	A	A	L	L	A	Q	Q	L	Q	M
39	M	C	C	R	R	C	R	R	B	B	B	B	R	C	R	Q	A	L	L	A	A	L	L	Q	A	L	Q	M
40	M	C	C	R	R	C	R	B	B	B	B	R	R	C	Q	L	A	A	L	L	A	A	L	Q	A	L	Q	M
41	M	C	C	R	R	C	C	B	B	B	R	R	R	M	M	L	L	A	A	L	L	A	Q	Q	A	Q	Q	M
42	M	C	C	R	R	B	C	B	B	B	R	R	Q	M	M	M	C	L	L	A	A	L	L	Q	L	A	Q	M
43	M	C	C	R	B	B	C	B	R	R	R	Q	Q	M	M	M	C	C	L	L	A	A	L	Q	L	L	Q	M
44	M	C	C	C	B	B	C	R	R	R	Q	L	Q	M	M	M	C	R	C	L	L	A	Q	Q	L	L	Q	M
45	M	C	R	C	B	B	C	R	R	Q	A	L	Q	M	M	M	C	R	B	C	L	L	Q	A	L	Q	Q	M
46	M	C	R	C	B	B	C	R	Q	L	A	Q	Q	M	M	M	C	C	B	R	C	L	Q	A	A	L	Q	M
47	M	C	R	C	B	B	C	Q	A	L	L	Q	Q	M	M	M	C	C	R	R	B	C	Q	A	A	Q	Q	M
48	M	C	R	C	B	B	M	M	A	A	L	Q	Q	M	M	M	C	C	R	B	B	M	M	A	A	Q	L	Q	M
49	M	C	R	C	B	A	M	M	B	A	Q	Q	Q	M	M	M	C	C	C	B	A	M	M	B	A	Q	L	Q	M
50	M	C	R	C	Q	M	M	L	C	Q	L	Q	M	M	M	C	R	C	Q	R	M	M	L	C	Q	L	Q	M
51	M	C	R	M	R	Q	M	M	C	L	M	L	Q	M	M	C	R	M	R	Q	M	M	C	L	M	L	Q	M
52	M	C	Q	M	C	Q	M	M	C	Q	M	C	Q	M	M	C	Q	M	C	Q	M	M	C	Q	M	C	Q	M
53	M	M	M	M	M	M	M	M	M	M	M	M	M	M	M	M	M	M	M	M	M	M	M	M	M	M	M	M
54	F	F	F	F	F	F	F	F	F	F	F	F	F	F	F	F	F	F	F	F	F	F	F	F	F	F	F	F

Firing Squad Synchronization Problem in Cellular Automata, Figure 5
Transition table for the Balzer's eight-state protocol (*left*) and its snapshots for synchronization operations on 28 cells (*right*)

zoyer [25] *and Waksman* [74] *has an O(n²) state-change complexity, respectively.*

Theorem 7 ([9]) *Gerken's 155-state synchronization algorithm has a* $\Theta(n \log n)$ *state-change complexity.*

It has been shown that any 3n-step *thread-like* synchronization algorithm has a $\Theta(n \log n)$ state-change complexity and such a 3n-step thread-like synchronization algorithm can be used for subspace synchronization in the Goto's time-optimum synchronization algorithm. Umeo, Hisaoka and Sogabe [54] has shown that:

Theorem 8 ([54]) *Goto's time-optimum synchronization algorithm as reconstructed by Umeo* [50] *has* $\Theta(n \log n)$ *state-change complexity.*

Figure 10 shows a comparison of the state-change complexities in several optimum-time synchronization algorithms.
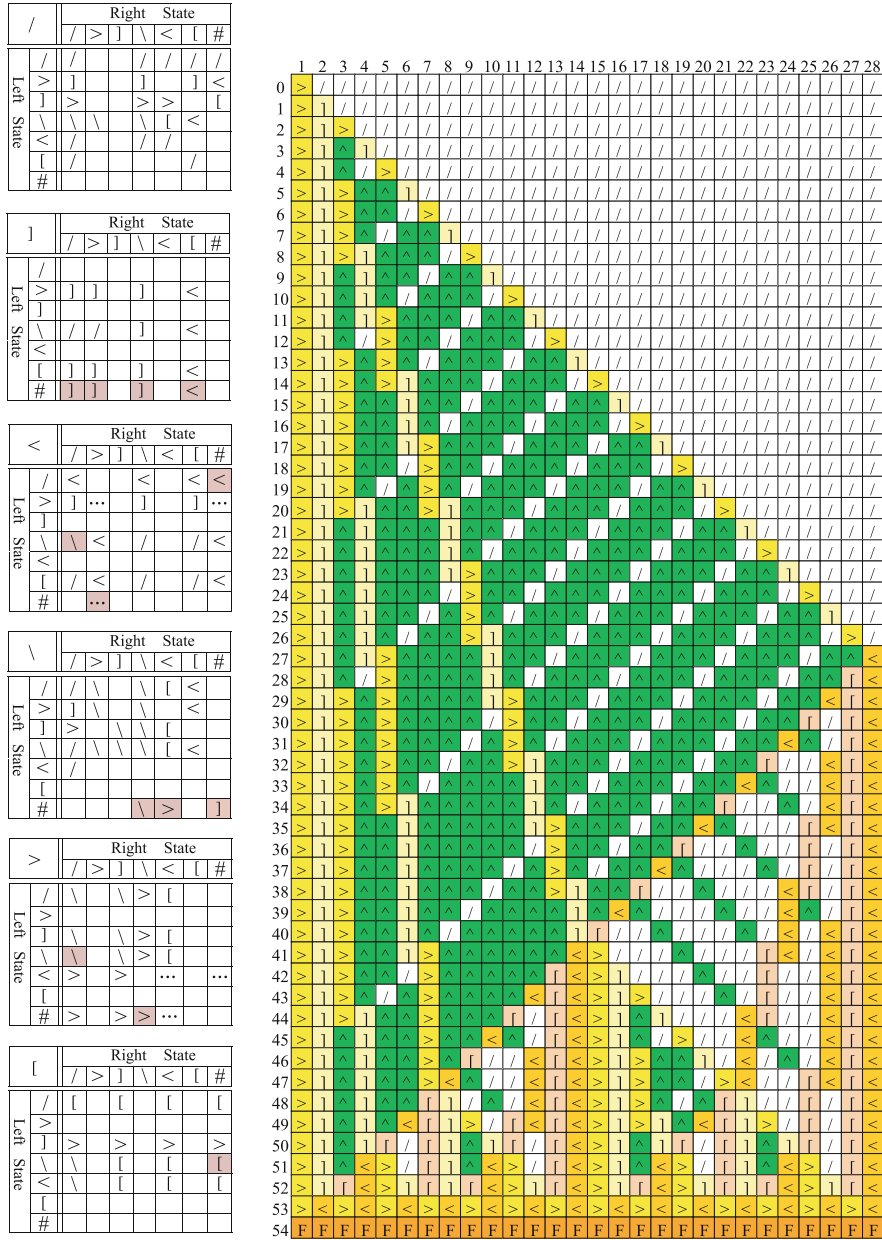
Firing Squad Synchronization Problem in Cellular Automata, Figure 6
Transition table for the Gerken's seven-state protocol (*left*) and snapshots for synchronization operations on 28 cells (*right*)

A Comparison of Quantitative Aspects of Optimum-Time Synchronization Algorithms

Section "A Comparison of Quantitative Aspects of Optimum-Time Synchronization Algorithms" presents a table based on a quantitative comparison of optimum-time synchronization algorithms and their transition tables discussed above with respect to the number of internal states of each finite state automaton, the number of transition rules realizing the synchronization, and the number of state-changes on the array.

One-Sided vs. Two-Sided Recursive Algorithms

Firing squad synchronization algorithms have been designed on the basis of parallel divide-and-conquer strategy that calls itself recursively in parallel. Those recursive calls are implemented by generating many *Generals* that

Transition table for state A

A	Right State					
Left State	A	B	C	G	L	X
A	A	B	C	B	A	F
B		G	C	C	G	C
C	A				A	
G			C	C		C
L	A	L	G			
X	F		G			

Transition table for state B

B	Right State					
Left State	A	B	C	G	L	X
A	B	B	L		G	
B	A	B	C	B	G	
C	A			L	L	L
G	C		B	G	C	G
L	G	B	L	B		
X						

Transition table for state C

C	Right State					
Left State	A	B	C	G	L	X
A		B		B	B	B
B			C	G	C	G
C	A	B	C	B	C	
G		B		B	B	B
L	A	G	C	G	C	
X						

Transition table for state G

G	Right State					
Left State	A	B	C	G	L	X
A		G	G		B	
B	G	G	G	B	G	
C	G	G	A	A	A	
G	G	G	F	B	F	
L	G	G	G			
X	G	G	F	A		

Transition table for state L

L	Right State					
Left State	A	B	C	G	L	X
A	L	L	L	C	G	C
B	L	L	L	L	L	L
C	L	L	L	G	A	G
G	L	L	L	A	C	A
L		L	L	L	L	L
X					L	

Snapshots of configurations on 28 cells

```
       1  2  3  4  5  6  7  8  9 10 11 12 13 14 15 16 17 18 19 20 21 22 23 24 25 26 27 28
 0  G  L  L  L  L  L  L  L  L  L  L  L  L  L  L  L  L  L  L  L  L  L  L  L  L  L  L  L
 1  A  C  L  L  L  L  L  L  L  L  L  L  L  L  L  L  L  L  L  L  L  L  L  L  L  L  L  L
 2  G  B  A  L  L  L  L  L  L  L  L  L  L  L  L  L  L  L  L  L  L  L  L  L  L  L  L  L
 3  G  C  G  G  L  L  L  L  L  L  L  L  L  L  L  L  L  L  L  L  L  L  L  L  L  L  L  L
 4  G  B  A  B  C  L  L  L  L  L  L  L  L  L  L  L  L  L  L  L  L  L  L  L  L  L  L  L
 5  G  C  G  L  C  A  L  L  L  L  L  L  L  L  L  L  L  L  L  L  L  L  L  L  L  L  L  L
 6  G  B  A  L  A  A  G  L  L  L  L  L  L  L  L  L  L  L  L  L  L  L  L  L  L  L  L  L
 7  G  C  G  L  A  B  B  C  L  L  L  L  L  L  L  L  L  L  L  L  L  L  L  L  L  L  L  L
 8  G  B  A  L  L  B  C  C  A  L  L  L  L  L  L  L  L  L  L  L  L  L  L  L  L  L  L  L
 9  G  C  G  G  L  L  C  A  A  G  L  L  L  L  L  L  L  L  L  L  L  L  L  L  L  L  L  L
10  G  B  A  B  C  L  A  A  B  B  C  L  L  L  L  L  L  L  L  L  L  L  L  L  L  L  L  L
11  G  C  G  L  C  L  A  B  B  C  C  A  L  L  L  L  L  L  L  L  L  L  L  L  L  L  L  L
12  G  B  A  L  C  L  L  B  C  C  A  A  G  L  L  L  L  L  L  L  L  L  L  L  L  L  L  L
13  G  C  G  L  C  A  L  L  C  A  A  B  B  C  L  L  L  L  L  L  L  L  L  L  L  L  L  L
14  G  B  A  L  A  A  G  L  A  A  B  B  C  C  A  L  L  L  L  L  L  L  L  L  L  L  L  L
15  G  C  G  L  A  B  B  L  A  B  B  C  C  A  A  G  L  L  L  L  L  L  L  L  L  L  L  L
16  G  B  A  L  L  B  G  L  L  B  C  C  A  A  B  B  C  L  L  L  L  L  L  L  L  L  L  L
17  G  C  G  G  L  B  B  C  L  L  C  A  A  B  B  C  C  A  L  L  L  L  L  L  L  L  L  L
18  G  B  A  B  L  B  C  C  A  L  A  A  B  B  C  C  A  A  G  L  L  L  L  L  L  L  L  L
19  G  C  G  G  L  L  C  A  A  L  A  B  B  C  C  A  A  B  B  C  L  L  L  L  L  L  L  L
20  G  B  A  B  C  L  A  A  A  L  L  B  C  C  A  A  B  B  C  C  A  L  L  L  L  L  L  L
21  G  C  G  L  C  L  A  A  A  G  L  L  C  A  A  B  B  C  C  A  A  G  L  L  L  L  L  L
22  G  B  A  L  C  L  A  A  B  B  C  L  A  A  B  B  C  C  A  A  B  B  C  L  L  L  L  L
23  G  C  G  L  C  L  A  B  B  C  C  L  A  B  B  C  C  A  A  B  B  C  C  A  L  L  L  L
24  G  B  A  L  C  L  L  B  C  C  C  L  L  B  C  C  A  A  B  B  C  C  A  A  G  L  L  L
25  G  C  G  L  C  A  L  L  C  C  C  A  L  L  C  A  A  B  B  C  C  A  A  B  B  C  L  L
26  G  B  A  L  A  A  G  L  C  C  A  A  G  L  A  A  B  B  C  C  A  A  B  B  C  C  A  L
27  G  C  G  L  A  B  B  L  C  A  A  B  B  L  A  B  B  C  C  A  A  B  B  C  C  A  A  C
28  G  B  A  L  L  B  G  L  A  A  B  B  G  L  L  B  C  C  A  A  B  B  C  C  A  A  C  B
29  G  C  G  G  L  B  B  L  A  B  B  B  C  L  L  C  A  A  B  B  C  C  A  A  C  B  L
30  G  B  A  B  L  B  G  L  L  B  B  B  C  C  A  L  A  A  B  B  C  C  A  A  C  B  L  L
31  G  C  G  G  L  B  B  C  L  B  B  C  C  A  A  L  A  B  B  C  C  A  A  C  B  L  L  L
32  G  B  A  B  L  B  C  C  L  B  C  C  A  A  A  L  L  B  C  C  A  A  C  B  L  L  L  L
33  G  C  G  G  L  L  C  C  L  L  C  A  A  A  A  G  L  L  C  A  A  C  B  L  L  L  L  L
34  G  B  A  B  C  L  C  C  A  L  A  A  A  A  B  B  C  L  A  A  C  B  L  L  L  L  L  L
35  G  C  G  L  C  L  C  A  A  L  A  A  A  B  B  C  C  L  A  C  B  L  L  L  L  L  L  L
36  G  B  A  L  C  L  A  A  A  L  A  A  B  B  C  C  C  L  G  B  L  L  L  L  L  L  L  L
37  G  C  G  L  C  L  A  A  A  L  A  B  B  C  C  C  C  G  G  C  L  L  L  L  L  L  L  L
38  G  B  A  L  C  L  A  A  A  L  L  B  C  C  C  C  B  A  G  B  A  L  L  L  L  L  L  L
39  G  C  G  L  C  L  A  A  A  G  L  L  C  C  C  B  A  C  G  C  G  G  L  L  L  L  L  L
40  G  B  A  L  C  L  A  A  B  B  C  L  C  C  B  A  C  B  G  B  A  B  C  L  L  L  L  L
41  G  C  G  L  C  L  A  B  B  C  C  L  C  B  A  C  B  G  B  G  C  G  L  C  A  L  L  L
42  G  B  A  L  C  L  L  B  C  C  C  L  G  A  C  B  L  L  G  B  A  L  A  A  G  L  L  L
43  G  C  G  L  C  A  L  L  C  C  C  G  G  C  B  L  L  L  G  C  G  L  A  B  B  C  L  L
44  G  B  A  L  A  A  G  L  C  C  B  A  G  B  L  L  L  L  G  B  A  L  L  B  C  C  A  L
45  G  C  G  L  A  B  B  L  C  B  A  C  G  C  L  L  L  L  G  C  G  G  L  L  C  A  A  C
46  G  B  A  L  L  B  G  L  G  A  C  B  G  B  A  L  L  L  G  B  A  B  C  L  A  A  C  B
47  G  C  G  G  L  B  B  A  G  C  B  L  G  C  G  G  L  L  G  C  G  L  C  L  A  C  B  L
48  G  B  A  B  L  B  A  C  G  B  L  L  G  B  A  L  C  L  G  B  A  L  C  L  G  B  L  L
49  G  C  G  G  L  G  C  B  G  C  L  L  G  C  G  L  C  G  G  G  C  L  C  G  G  C  L  L
50  G  B  A  B  A  G  B  L  G  B  A  L  G  B  A  L  G  A  G  B  A  L  G  A  G  B  A  L
51  G  C  G  B  C  G  C  L  G  C  G  C  G  C  G  C  G  C  G  C  G  C  G  C  G  C  G  C
52  G  B  G  B  G  B  G  B  G  B  G  B  G  B  G  B  G  B  G  B  G  B  G  B  G  B  G  B
53  G  G  G  G  G  G  G  G  G  G  G  G  G  G  G  G  G  G  G  G  G  G  G  G  G  G  G  G
54  F  F  F  F  F  F  F  F  F  F  F  F  F  F  F  F  F  F  F  F  F  F  F  F  F  F  F  F
```

Firing Squad Synchronization Problem in Cellular Automata, Figure 7
Transition table for the Mazoyer's six-state protocol (*left*) and its snapshots of configurations on 28 cells (*right*)

work for synchronizing divided small areas in the cellular space. Initially a *General* G_0 located at the left end works for synchronizing the whole cellular space consisting of n cells. In Fig. 11 (left), G_1 synchronizes the subspace between G_1 and the right end of the array. The ith *General* G_i, $i = 2, 3, \ldots$, works for synchronizing the cellular space between G_{i-1} and G_i, respectively. Thus, all of the *Generals* generated by G_0 are located at the left end of the divided cellular spaces to be synchronized. On the other hand, in Fig. 11 (right), the *General* G_0 generates *General* G_i, $i = 1, 2, 3, \ldots$. Each G_i, $i = 1, 2, 3, \ldots$, synchronizes

the divided space between G_i and G_{i+1}, respectively. In addition, G_i, $i = 2, 3, \ldots$, does the same operations as G_0. Thus, in Fig. 11 (right) one can find *Generals* located at either end of the subspace for which they are responsible. If all of the recursive calls for the synchronization are issued by *Generals* located at one (both two) end(s) of partitioned cellular spaces for which the *General* works, the synchronization algorithm is said to have *one-sided* (*two-sided*) recursive property, respectively. A synchronization algorithm with the one-sided (two-sided) recursive property is referred to as one-sided (two-sided) recursive synchro-

Firing Squad Synchronization Problem in Cellular Automata, Table 1
Quantitative comparison of transition rule sets for optimum-time firing squad synchronization algorithms. The ∗ symbol shows the correction and reduction of transition rules made in Umeo, Hisaoka and Sogabe [54]. The ∗∗ symbol indicates the number of states and rules obtained after the expansion of the original two-layer construction

Algorithm	# of states	# of transition rules	State change complexity
Goto [12]	many thousands	–	$\Theta(n \log n)$
Waksman [74]	16	202*(3216)	$O(n^2)$
Balzer [1]	8	165* (182)	$O(n^2)$
Noguchi [34]	8	119	$O(n^2)$
Gerken [9]	7	105* (118)	$O(n^2)$
Mazoyer [25]	6	119* (120)	$O(n^2)$
Gerken [9]	155**(32)	2371**(347)	$\Theta(n \log n)$

Firing Squad Synchronization Problem in Cellular Automata, Table 2
A qualitative comparison of optimum-time firing squad synchronization algorithms

Algorithm	One-/two-sided	Recursive/non-recursive	# of signals
Goto [12]	–	non-recursive	finite
Waksman [74]	two-sided	recursive	infinite
Balzer [1]	two-sided	recursive	infinite
Noguchi [34]	two-sided	recursive	infinite
Gerken [9]	two-sided	recursive	infinite
Mazoyer [25]	one-sided	recursive	infinite
Gerken [9]	two-sided	recursive	finite

nization algorithm. Figure 11 illustrates a time-space diagram for one-sided (Fig. 11 (left)) and two-sided (Fig. 11 (right)) recursive synchronization algorithms both operating in optimum $2n - 2$ steps. It is noted that optimum-time synchronization algorithms developed by Balzer [1], Gerken [9], Noguchi [34] and Waksman [74] are two-sided ones and an algorithm proposed by Mazoyer [25] is an only synchronization algorithm with the one-sided recursive property.

Observation 3.2 ([54]) *Optimum-time synchronization algorithms developed by Balzer [1], Gerken [9], Noguchi [34] and Waksman [74] are two-sided ones. The algorithm proposed by Mazoyer [25] is a one-sided one.*

A more general design scheme for one-sided recursive optimum-time synchronization algorithms can be found in Mazoyer [24].

Recursive vs. Non-recursive algorithms

As is shown in the previous section, the optimum-time synchronization algorithms developed by Balzer [1], Gerken [9], Mazoyer [25], Noguchi [34] and Waksman [74] are recursive ones. On the other hand, it is noted that overall structure of the reconstructed Goto's algorithm is a non-recursive one, however divided sub-

spaces are synchronized by using a recursive $3n + O(1)$-step synchronization algorithm.

Number of Signals

Waksman [74] devised an efficient way to cause a general cell to generate infinite signals at propagating speeds of $1/1, 1/3, 1/7, .., 1/(2^k - 1)$, where k is any natural number. These signals play an important role in dividing the array into two, four, eight, …, equal parts synchronously. The same set of signals is used in Balzer [1]. Gerken [9] had a similar idea in the construction of his seven-state algorithm. Thus *infinite* set of signals with different propagation speed is used in the first three algorithms. On the other hand, *finite* sets of signals with propagating speed $\{1/5, 1/2, 1/1\}$ and $\{1/3, 1/2, 3/5, 1/1\}$ are made use of in Gerken's 155-state algorithm and the reconstructed Goto's algorithm, respectively.

A Comparison of Qualitative Aspects of Optimum-Time Synchronization Algorithms

Section "A Comparison of Qualitative Aspects of Optimum-Time Synchronization Algorithms" presents a table based on a qualitative comparison of optimum-time synchronization algorithms with respect to one/two-sided re-
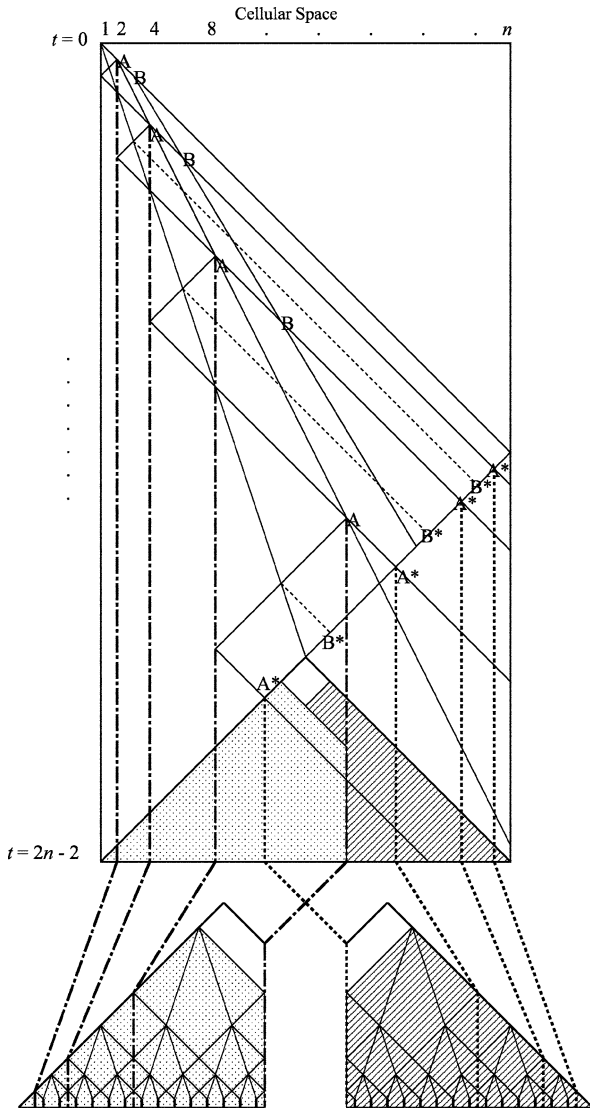
Firing Squad Synchronization Problem in Cellular Automata, Figure 8
Time-space diagram for Goto's algorithm as reconstructed by Umeo [50]
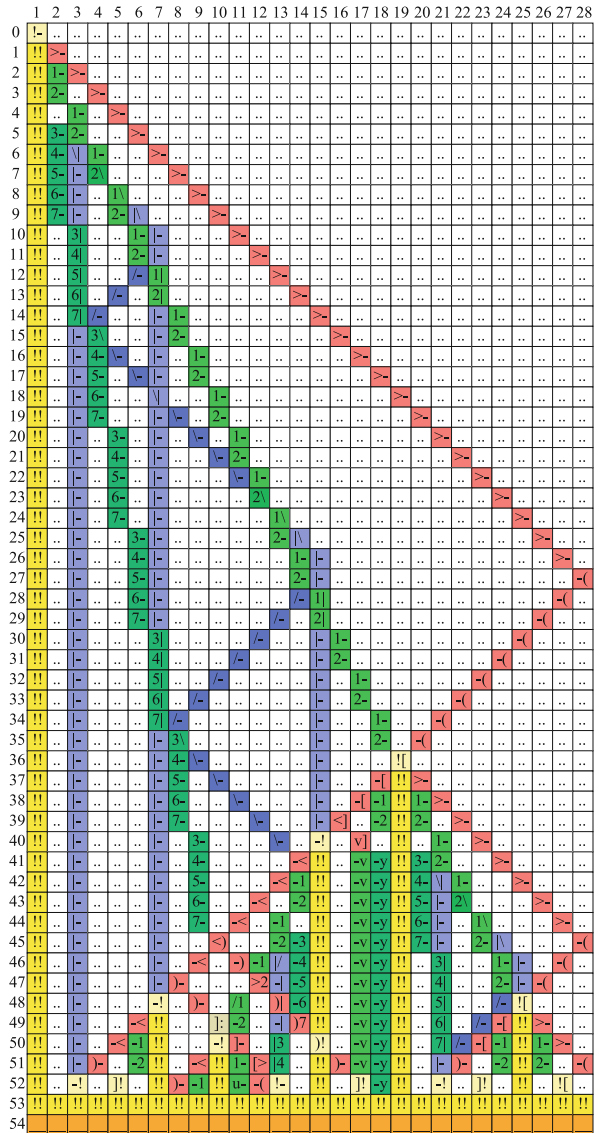
Firing Squad Synchronization Problem in Cellular Automata, Figure 9
Snapshots of the Gerken's 155-state algorithm on 28 cells

cursive properties and the number of signals being used for simultaneous space divisions.

Variants of the Firing Squad Synchronization Problem

Generalized Firing Squad Synchronization Problem

Section "Generalized Firing Squad Synchronization Problem" considers a *generalized* firing squad synchronization problem which allows the general to be located any-

where on the array. It has been shown to be impossible to synchronize any array of length n less than $n - 2 + \max(k, n - k + 1)$ steps, where the general is located on C_k. Moore and Langdon [30], Szwerinski [45] and Varshavsky, Marakhovsky and Peschansky [69] developed a generalized optimum-time synchronization algorithm with 17, 10 and 10 internal states, respectively, that can synchronize any array of length n at exactly $n - 2 + \max(k, n - k + 1)$ steps. Recently, Settle and Simon [43] and Umeo, Hisaoka, Michisaka, Nishioka and Maeda [56]
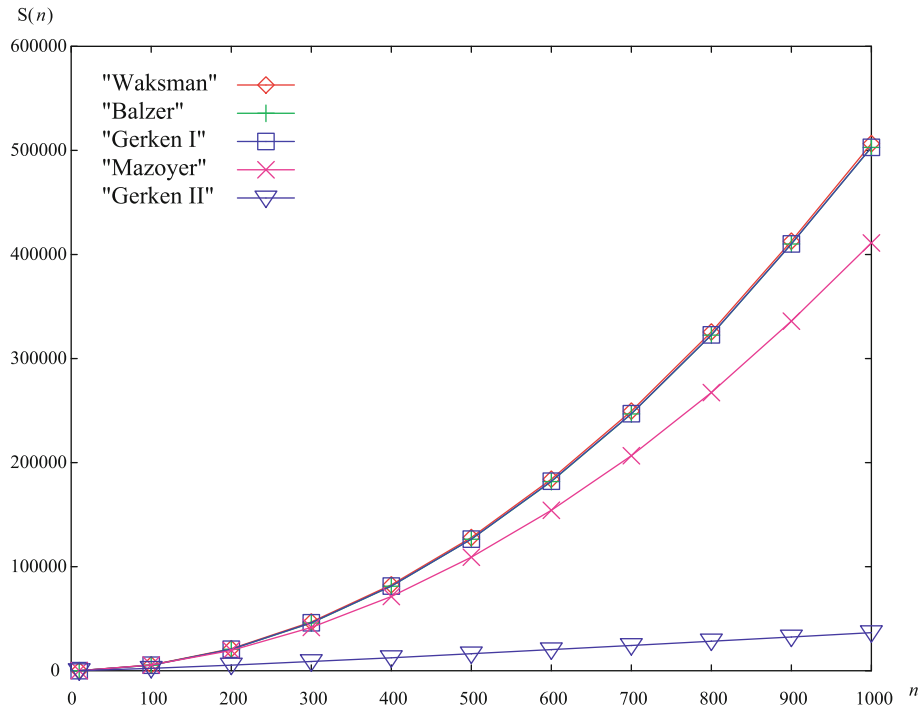
Firing Squad Synchronization Problem in Cellular Automata, Figure 10
A comparison of state-change complexities in optimum-time synchronization algorithms
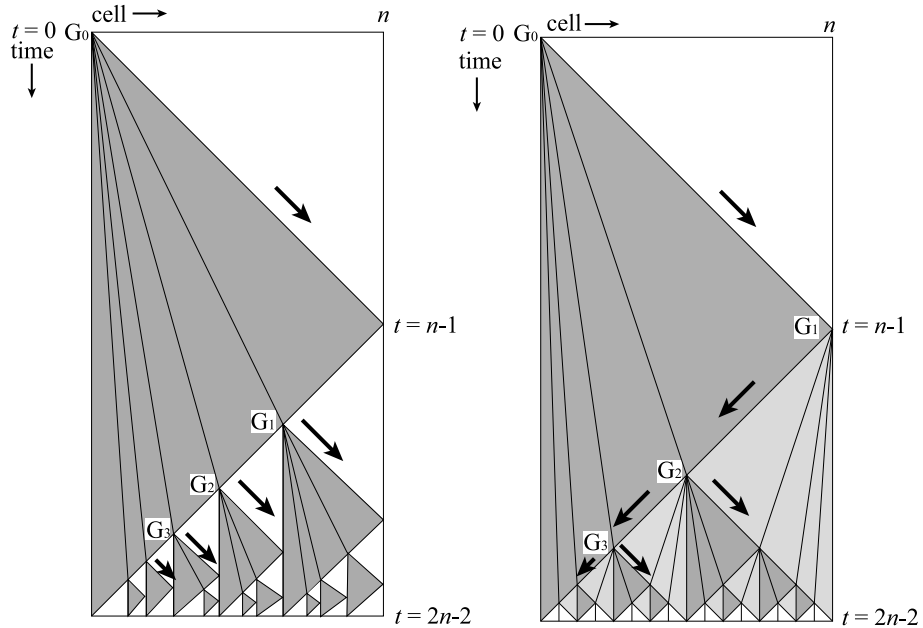
Firing Squad Synchronization Problem in Cellular Automata, Figure 11
One-sided recursive synchronization scheme (*left*) and two-sided recursive synchronization scheme (*right*)
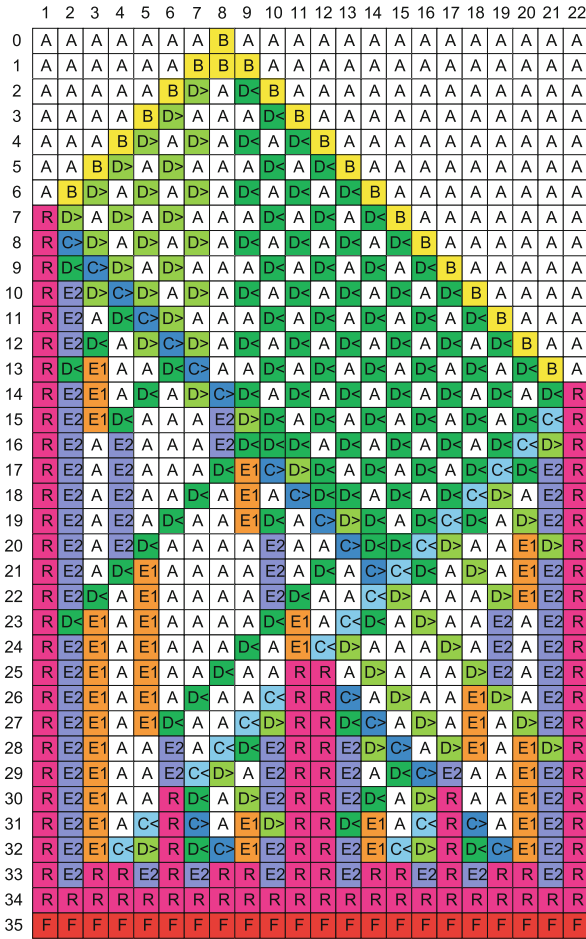
Firing Squad Synchronization Problem in Cellular Automata, Figure 12

Snapshots of the three Russian's 10-state generalized optimum-time synchronization algorithm on 22 cells

State Q

L \ R	Q	P	R	Z	M	*
Q	Q	P	Q	Q		Q
P	P	P				P
R	Q		Q			
Z	Q			Q		Q
M						
*	Q	P	Q	Q		

State Z

L \ R	Q	P	R	Z	M	*
Q			Q	P	Q	
P		Z		Z		
R	Q		Q	Q		Q
Z	P	Z	Q	F	Q	F
M	Q			Q	Q	Q
*		Z	Q	F	Q	

State P

L \ R	Q	P	R	Z	M	*
Q	Z	Z	R	R		
P	Z		Z	Z		
R	R	Z	Z			Z
Z	R	Z		Z		Z
M						
*	Z		Z	Z		

State M

L \ R	Q	P	R	Z	M	*
Q						
P						
R			R	Z		
Z			Z			
M						
*						

State R

L \ R	Q	P	R	Z	M	*
Q			R	P	Z	
P			M	R	M	
R	R	M				M
Z	P	R			R	
M	Z	M	M	R	M	
*						

Firing Squad Synchronization Problem in Cellular Automata, Figure 13

Transition table for symmetric six-state protocol (*left*) and snapshots for synchronization algorithm on 14 cells

have proposed a 9-state generalized synchronization algorithm operating in optimum-step. Figure 12 shows snapshots for synchronization configurations based on the rule set of Varshavsky, Marakhovsky and Peschansky [69].

Theorem 9 ([30,43,45,56,69]) *There exists a cellular automaton that can synchronize any one-dimensional array of length n in optimum $n - 2 + \max(k, n - k + 1)$ steps, where the general is located on the kth cell from left end.*

Non-Optimum-Time 3n-Step Synchronization Algorithms

Non-optimum-time 3n-step algorithm is a simple and straightforward one that exploits a parallel divide-and-conquer strategy based on an efficient use of 1/1- and 1/3-speed of signals. Minsky and MacCarthy [28] gave an idea for designing the 3n-step synchronization algorithm, and

Fischer [8] implemented the 3n-step algorithm, yielding a 15-state implementation, respectively. Yunès [76] developed two seven-state synchronization algorithms, thus decreasing the number of internal states of each cellular automaton. This section presents a new symmetric six-state 3n-step firing squad synchronization algorithm developed in Umeo, Maeda and Hongyo [61]. The number *six* is the smallest one known at present in the class of 3n-step synchronization algorithms. Figure 13 shows the 6-state transition table and snapshots for synchronization on 14 cells. In the transition table, the symbols "P", "Q", "F" and "*" represent the general, quiescent, firing and boundary states, respectively. Yunès [79] also developed a symmetric 6-state 3n-step solution.

Theorem 10 ([61,79]) *There exists a symmetric 6-state cellular automaton that can synchronize any n cells in 3n + O(log n) steps.*

State Q

L \ R	Q	P	R	M	Z	*
Q	Q	P	Q	M	Q	Q
P	P	P				P
R	Q		Q			Q
M	M					M
Z	Q				Q	Q
*	Q	P	Q	M	Q	

State Z

L \ R	Q	P	R	M	Z	*
Q	P	Q	Q	Q	P	
P	Q	Z			Z	Q
R	Q		Q		Q	Q
M	Q			Q	Q	Q
Z	P	Z	Q	Q	F	F
*		Q	Q	Q	F	

State M

L \ R	Q	P	R	M	Z	*
Q	P	M	R	M		
P	M			M	Z	Z
R	R		R	R	Z	Z
M	M	M	R	M	Z	Z
Z			Z	Z	Z	
*	R	Z	Z	Z		

State R

L \ R	Q	P	R	M	Z	*
Q			R	Z	P	
P			M	M	R	
R	R	M		M		
M	Z	M	M	M	R	Z
Z	P	R		R		
*			Z			

State P

L \ R	Q	P	R	M	Z	*
Q	Z	Z	R		R	
P	Z		Z		Z	
R	R	Z	Z			Z
M				P	R	
Z	R	Z		R	Z	Z
*		Z		Z		

Snapshots for synchronization algorithm on 15 cells (rows 0–39, columns 1–15):

	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15
0	Q	Q	Q	Q	M	Q	Q	Q	Q	Q	Q	Q	Q	Q	Q
1	Q	Q	Q	M	P	M	Q	Q	Q	Q	Q	Q	Q	Q	Q
2	Q	Q	M	M	P	M	M	Q	Q	Q	Q	Q	Q	Q	Q
3	Q	M	M	M	P	M	M	M	Q	Q	Q	Q	Q	Q	Q
4	M	M	M	M	P	M	M	M	M	Q	Q	Q	Q	Q	Q
5	Z	M	M	M	P	M	M	M	M	M	Q	Q	Q	Q	Q
6	Q	Z	M	M	P	M	M	M	M	M	M	Q	Q	Q	Q
7	Q	Q	Z	M	P	M	M	M	M	M	M	M	Q	Q	Q
8	Q	Q	Q	Z	P	M	M	M	M	M	M	M	M	Q	Q
9	Q	Q	Q	Q	R	M	M	M	M	M	M	M	M	M	Q
10	Q	Q	Q	Q	Z	R	M	M	M	M	M	M	M	M	M
11	Q	Q	Q	Q	Q	R	R	M	M	M	M	M	M	M	Z
12	Q	Q	Q	Q	Q	R	M	R	M	M	M	M	M	Z	Q
13	Q	Q	Q	Q	Q	Z	R	M	R	M	M	M	Z	Q	Q
14	Q	Q	Q	Q	Q	Q	R	R	M	R	M	M	Z	Q	Q
15	Q	Q	Q	Q	Q	Q	R	M	R	M	R	M	Z	Q	Q
16	Q	Q	Q	Q	Q	Q	Z	R	M	Z	Q	Q	Q	Q	Q
17	Q	Q	Q	Q	Q	Q	Q	R	Z	Q	Q	Q	Q	Q	Q
18	Q	Q	Q	Q	Q	Q	Q	P	Q	Q	Q	Q	Q	Q	Q
19	Q	Q	Q	Q	Q	Q	P	Z	P	Q	Q	Q	Q	Q	Q
20	Q	Q	Q	Q	Q	P	R	Z	R	P	Q	Q	Q	Q	Q
21	Q	Q	Q	Q	P	R	R	Q	R	R	P	Q	Q	Q	Q
22	Q	Q	Q	P	R	M	R	Q	R	M	R	P	Q	Q	Q
23	Q	Q	P	R	M	R	Z	Q	Z	R	M	R	P	Q	Q
24	Q	P	R	M	R	R	Q	Q	Q	R	R	M	R	P	Q
25	P	R	M	R	R	Q	Q	Q	Q	Q	R	R	M	R	P
26	Z	M	R	M	R	Z	Q	Q	Q	Z	R	M	R	M	Z
27	Q	Z	M	R	R	Q	Q	Q	Q	Q	R	R	M	Z	Q
28	Q	Q	Z	M	R	Q	Q	Q	Q	Q	R	M	Z	Q	Q
29	Q	Q	Q	Z	Z	Q	Q	Q	Q	Q	Z	Z	Q	Q	Q
30	Q	Q	Q	P	P	Q	Q	Q	Q	Q	P	P	Q	Q	Q
31	Q	Q	P	Z	Z	P	Q	Q	Q	P	Z	Z	P	Q	Q
32	Q	P	R	Z	Z	R	P	Q	P	R	Z	Z	R	P	Q
33	P	R	R	Q	Q	R	R	P	R	R	Q	Q	R	R	P
34	Z	M	R	Q	Q	Q	R	M	Z	M	R	Q	Q	R	M	Z
35	Q	Z	Z	Q	Q	Z	Z	Q	Z	Z	Q	Q	Z	Z	Q
36	Q	P	P	Q	Q	P	P	Q	P	P	Q	Q	P	P	Q
37	P	Z	Z	P	P	Z	Z	P	Z	Z	P	P	Z	Z	P
38	Z	Z	Z	Z	Z	Z	Z	Z	Z	Z	Z	Z	Z	Z	Z
39	F	F	F	F	F	F	F	F	F	F	F	F	F	F	F

Firing Squad Synchronization Problem in Cellular Automata, Figure 14

Transition table for generalized symmetric six-state protocol (*left*) and snapshots for synchronization algorithm on 15 cells with a *General* on C_5

A non-trivial, new symmetric six-state $3n$-step generalized firing squad synchronization algorithm is also presented in Umeo, Maeda and Hongyo [61]. Figure 14 gives a list of transition rules for the 6-state generalized synchronization algorithm and snapshots of configurations on 15 cells. The symbol "M" is the general state.

Theorem 11 ([61]) *There exists a symmetric 6-state cellular automaton that can solve the generalized firing squad synchronization problem in* $\max(k, n - k + 1) + 2n + O(\log n)$ *steps.*

In addition, a state-change complexity is studied in $3n$-step firing squad synchronization algorithms. It has been shown that the six-state algorithms presented above have $O(n^2)$ state-change complexity, on the other hand, the *thread-like* $3n$-step algorithms developed so far have $O(n \log n)$ state-change complexity. Here, the following

table presents a quantitative comparison of the $3n$-step synchronization algorithms developed so far.

Delayed Firing Squad Synchronization Algorithm

This section introduces a *freezing-thawing* technique that yields a delayed synchronization algorithm for one-dimensional arrays. The technique is very useful in the design of time-efficient synchronization algorithms for one- and two-dimensional arrays in Umeo [52], Yunès [77] and Umeo and Uchino [65]. A similar technique was used by Romani [37] in the tree synchronization. The technique is stated as in the following theorem.

Theorem 12 ([52]) *Let t_0, t_1, t_2 and Δt be any integer such that $t_0 \geq 0$, $t_0 \leq t_1 \leq t_0 + n - 1$, $t_1 \leq t_2$ and $\Delta t = t_2 - t_1$. It is assumed that a usual optimum-time synchronization operation is started at time $t = t_0$ by generating a special signal at the left end of one-dimensional array and the right end cell of the array receives another special signals from outside at time $t = t_1$ and t_2, respectively. Then, there exists a one-dimensional cellular automaton that can synchronize the array of length n at time $t = t_0 + 2n - 2 + \Delta t$.*

The array operates as follows:

1. Start an optimum-time firing squad synchronization algorithm at time $t = t_0$ at the left end of the array. A 1/1-speed signal is propagated towards the right direction to wake-up cells in quiescent state. The signal is referred to as *wake-up signal*. A *freezing* signal is given from outside at time $t = t_1$ at the right end of the array. The signal is propagated in the left direction at its maximum speed, that is, 1 cell per 1 step, and freezes the configuration progressively. Any cell that receives the freezing signal from its right neighbor has to stop its state-change and transmits the freezing signal to its left neighbor. The frozen cell keeps its state as long as no thawing signal will arrive.

2. A special signal supplied with outside at time $t = t_2$ is used as a *thawing* signal that thaws the frozen configuration. The thawing signal forces the frozen cell to resume its state-change procedures immediately. See Fig. 15 (left). The signal is also transmitted toward the left end at speed 1/1.

The readers can see how those three signals work. The entire configuration can be frozen during Δt steps and the synchronization on the array is delayed for Δt steps. It is easily seen that the freezing signal can be replaced by the reflected signal of the wake-up signal, that is generated at

Firing Squad Synchronization Problem in Cellular Automata, Table 3
A comparison of 3n-step firing squad synchronization algorithms

Algorithm	# States	# Rules	Time complexity	State-change complexity	Generals's position	Type	Notes	Ref.
Minsky and MacCarthy	13	–	$3n + \theta_n \log n + c$	$O(n \log n)$	left	thread	$0 \le \theta_n < 1$	[28]
Fischer	15	–	$3n - 4$	$O(n \log n)$	left	thread	–	[8]
Yunès	7	105	$3n \pm 2\theta_n \log n + c$	$O(n \log n)$	left	thread	$0 \le \theta_n < 1$	[76]
Yunès	7	107	$3n \pm 2\theta_n \log n + c$	$O(n \log n)$	left	thread	$0 \le \theta_n < 1$	[76]
Settle and Simon	6	134	$3n + 1$	$O(n^2)$	right	plane	–	[43]
Settle and Simon	7	127	$2n - 2 + k$	$O(n^2)$	arbitrary	plane	–	[43]
Umeo et al.	6	78	$3n + O(\log n)$	$O(n^2)$	left	plane	–	[61]
Umeo et al.	6	115	$\max(k, n - k + 1) + 2n + O(\log n)$	$O(n^2)$	arbitrary	plane	–	[61]
Umeo and Yanagihara	5	67	$3n - 3$	$O(n^2)$	left/right	plane	$n = 2^k$, $k = 1, 2, \ldots$	[67]
Yunès	6	105	$3n + \lceil \log n \rceil - 3$	$O(n \log n)$	left	thread		[79]
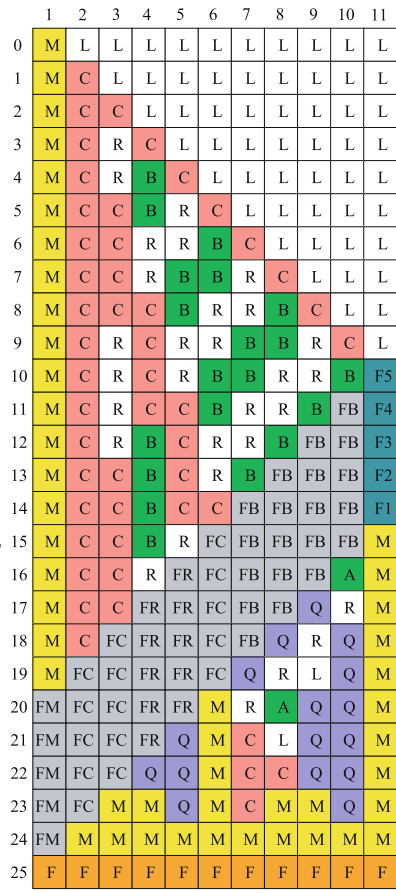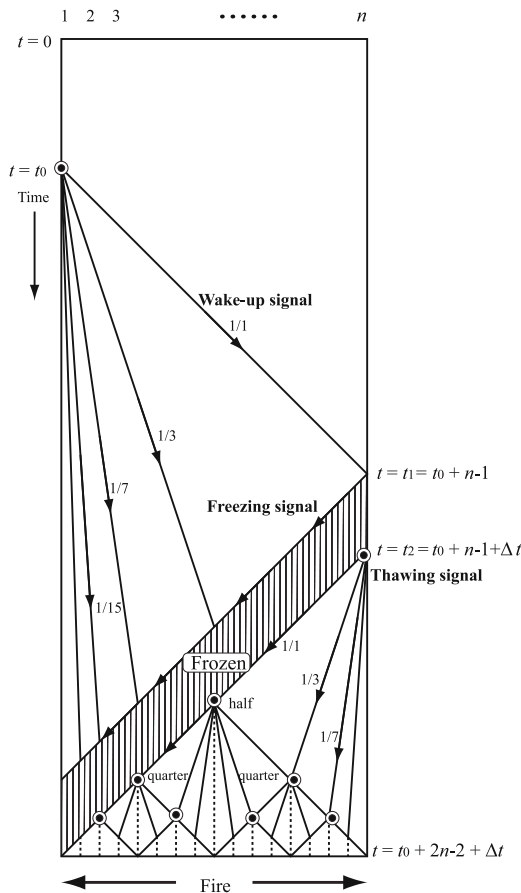
Firing Squad Synchronization Problem in Cellular Automata, Figure 15
Time-space diagram for delayed firing squad synchronization scheme based on the *freezing-thawing* technique (*left*) and a delayed
(for $\Delta t = 5$ steps) configuration in Balzer's optimum-time firing squad synchronization algorithm on $n = 11$ cells (*right*)

the right end cell at time $t = t_0 + n - 1$. See Fig. 15. The scheme is referred to as *freezing-thawing* technique.

Fault-Tolerant Firing Squad Synchronization Problem

Consider a one-dimensional array of cells, some of which are defective. At time $t = 0$, the left end cell C_1 is in the *fire-when-ready* state, which is the initialization signal for the array. The *fault-tolerant* firing squad synchronization problem for cellular automata with *defective* cells is to determine a description of cells that ensures all *intact* cells enter the *fire* state at exactly the same time and for the first time. The fault-tolerant firing squad synchronization problem has been studied in Kutrib and Vollmar [22,23], Umeo [52] and Yunès [77].

Cellular Automata with Defective Cells

- **Intact and Defective Cells:** Each cell has its own self-diagnosis circuit that diagnoses itself before its operation. A consecutive defective (intact) cells are referred to as a *defective* (*intact*) segment, respectively. Any defective and intact cell can detect whether its neighbor cells are defective or not. Cellular arrays are assumed to have an intact segment at its left and right ends. New defections do not occur during the operational lifetime on any cell.

- **Signal Propagation in a Defective Segment:** It is assumed that any cell in defective segment can only transmit the signal to its right or left neighbor depending on the direction in which it comes to the defective segment. The speed of the signal in any defective segment is fixed to 1/1, that is, one cell per one step. In defective segments, both the information carried by the signal and the direction in which the signal is propagated are preserved without any modifications. Thus, one can see that any defective segment has two one-way pipelines that can transfer one state at 1/1 speed in either direction. Note that from a standard viewpoint of state transition of usual CA each cell in a defective segment can change its internal states in a specific manner.

The array consists of p defective segments and $(p + 1)$ intact segments, where they are denoted by I_i and D_j, respectively and p is any positive integer. Let n_i and m_j be number of cells on the ith intact and jth defective segments, where i and j be any integer such that $1 \leq i \leq p + 1$ and $1 \leq j \leq p$. Let n be the number of cells of the array such that $n = (n_1+m_1)+(n_2+m_2)+,\ldots,+(n_p+m_p)+n_{p+1}$.

Fault-Tolerant Firing Squad Synchronization Algorithms   Umeo [52] studied the synchronization algo-
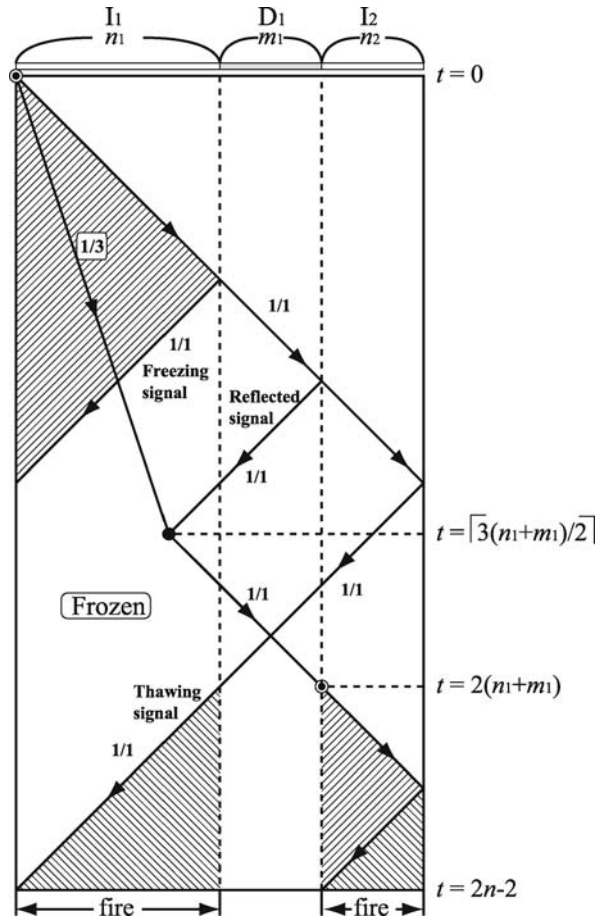
Firing Squad Synchronization Problem in Cellular Automata, Figure 16
Time-space diagram for optimum-time firing squad synchronization algorithm with one defective segment

rithms for such arrays that there are locally more intact cells than defective ones, i.e., $n_i \geq m_i$ for any i such that $1 \leq i \leq p$. First, consider the case $p = 1$ where the array has one defective segment and $n_1 \geq m_1$. Figure 16 illustrates a simple synchronization scheme. The fault-tolerant synchronization algorithm for one defective segment is stated as follows:

Theorem 13 ([52]) *Let M be any cellular array of length n with one defective and two intact segments such that $n_1 \geq m_1$, where n_1 and m_1 denote the number of cells on the first intact and defective segments, respectively. Then, M is synchronizable in $2n - 2$ optimum-time.*

The synchronization scheme above can be generalized to arrays with multiple defective segments more than two. Figure 17 shows the synchronization scheme for a cellu-
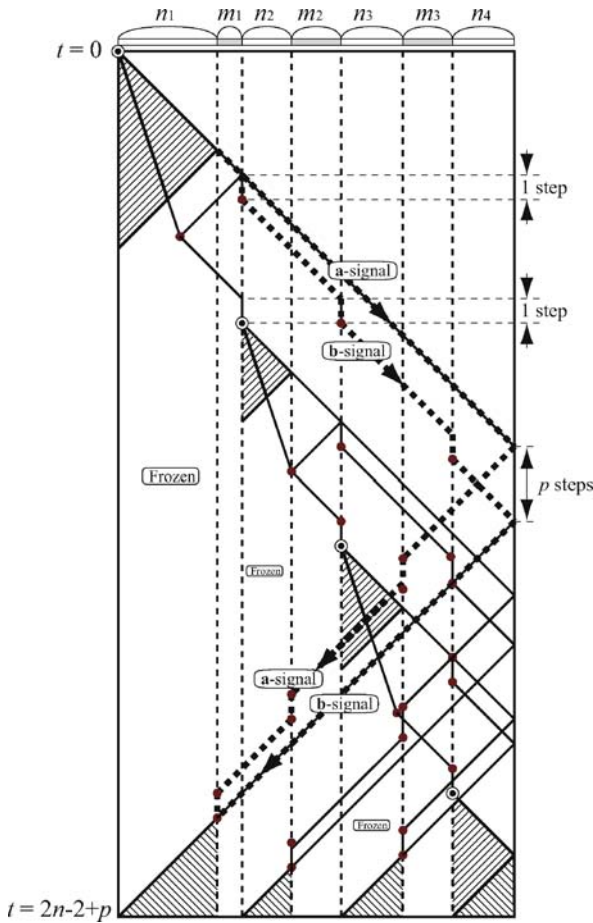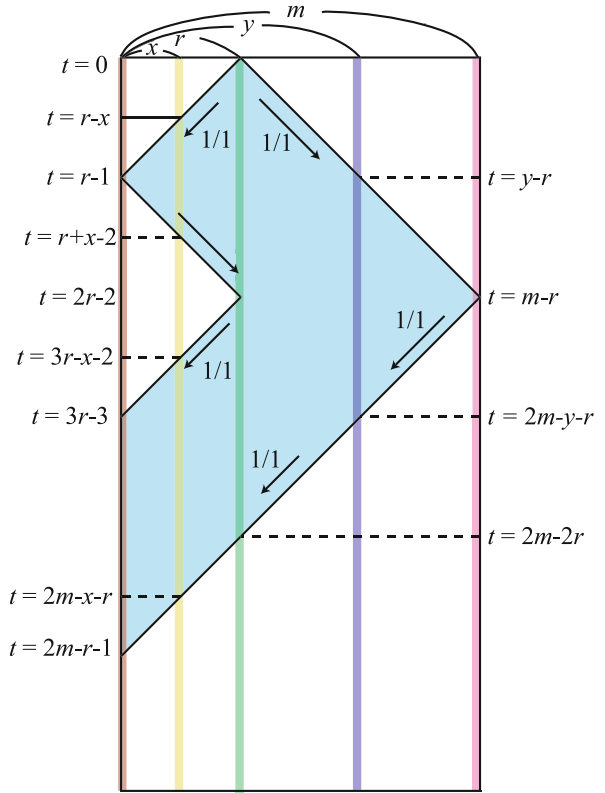
Firing Squad Synchronization Problem in Cellular Automata, Figure 17

Time-space diagram for optimum-time firing squad synchronization algorithm with three defective segments

lar array with three defective segments. Details of the algorithm can be found in Umeo [52].

Theorem 14 ([52]) *Let p be any positive integer and M be any cellular array of length n with p defective segments, where $n_i \geq m_i$ and $n_i + m_i \geq p - i$, for any i such that $1 \leq i \leq p$. Then, M is synchronizable in $2n - 2 + p$ steps.*

Partial Solutions

The original firing squad synchronization problem is defined to synchronize all cells of one-dimensional array. In this section, consider a *partial FSSP solution* that can synchronize an infinite number of cells, but not all. The first partial solution was given by Umeo and Yanagihara [67]. They proposed a five-state solution that can synchronize any one-dimensional cellular array of length $n = 2^k$ in $3n - 3$ steps for any positive integer k. Figure 18 shows

the five-state transition table consisting of 67 rules and its snapshots for $n = 8$ and 16. In the transition table, the symbols "R", "Q", "F" and "∗" represent the general, quiescent, firing and boundary states, respectively.

Theorem 15 ([67]) *There exists a 5-state cellular automaton that can synchronize any array of length $n = 2^k$ in $3n - 3$ steps, where k is any positive integer.*

Surprisingly, Yunès [80] and Umeo, Kamikawa and Yunès [58] proposed four-state synchronization protocols which are based on an algebraic property of Wolfram's two-state cellular automata.

Theorem 16 ([58,80]) *There exists a 4-state cellular automaton that can synchronize any array of length $n = 2^k$ in non-optimum $2n - 1$ steps, where k is any positive integer.*

Figure 19 shows the four-state transition table given by Yunès [80] which is based on Wolfram's Rule 60. It consists of 32 rules. Snapshots for $n = 16$ cells are also illustrated in Fig. 19. In the transition table, the symbols "G", "Q", "F" and "∗" represent the general, quiescent, firing and boundary states, respectively. Figure 20 shows the four-state transition table given by Umeo, Kamikawa and Yunès [58] which is based on Wolfram's Rule 150. It has 32 transition rules. Snapshots for $n = 16$ are given in the figure. In the transition table, the symbols "G", "Q", "F" and "∗" represent the general, quiescent, firing and boundary states, respectively.

Umeo, Kamikawa and Yunès [58] proposed a different, but looking-similar, 4-state protocol based on Wolfram's Rule 150. Figure 21 shows the four-state transition table consisting of 37 rules and its snapshots for $n = 9$ and 17. The four-state protocol has some desirable properties. Note that the algorithm operates in optimum-step and its transition rule is symmetric. The state of a general can be either "G" or "A". Its initial position can be at the left or right end.

Theorem 17 ([58]) *There exists a symmetric 4-state cellular automaton that can synchronize any array of length $n = 2^k + 1$ in $2n - 2$ optimum-steps, where k is any positive integer.*

Yunès [80] has given a state lower bound for the partial solution:

Theorem 18 ([80]) *There is no 3-state partial solution.*

Thus, the 4-state partial solutions given above are optimum in state-number complexity in partial solutions.

State Q

Left\Right	Q	R	L	S	*
Q	Q	Q	L	Q	Q
R	R	R			R
L	Q		L	S	Q
S	Q	S			
*	Q	Q	L	Q	

State L

Left\Right	Q	R	L	S	*
Q	L	S	Q	Q	
R	Q	Q	R	R	Q
L	L		L	L	
S	R	F			F
*			R	R	

State R

Left\Right	Q	R	L	S	*
Q	R	R	Q	L	
R	Q	R	L		L
L	S		Q	F	
S	Q	R	L		L
*	S		Q	F	

State S

Left\Right	Q	R	L	S	*
Q	Q	S	L	Q	
R	R			F	
L	S			S	
S	Q	S	F		
*	Q	S	F		
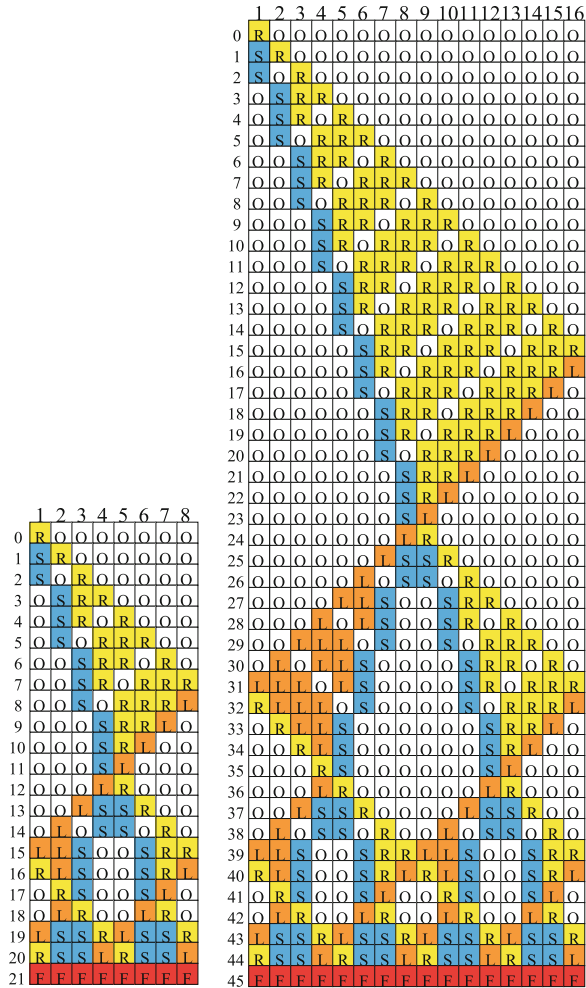
Firing Squad Synchronization Problem in Cellular Automata, Figure 18
Transition table for the five-state protocol (*left*) and its snapshots of configurations on 8 and 16 cells (*right*)

Synchronization Algorithm for One-Bit Communication Cellular Automata

In the study of cellular automata, the amount of bit-information exchanged at one step between neighboring cells has been assumed to be O(1)-bit. An O(1)-bit communication CA is a *conventional* cellular automaton in which the number of communication bits exchanged at one step between neighboring cells is assumed to be O(1)-bit, however, such an inter-cell bit-information exchange has been hidden behind the definition of conventional automata-theoretic finite state description. On the other hand, the 1-bit inter-cell communication model is a new cellular automaton in which inter-cell communication is restricted to 1-bit data, referred to as the 1-bit CA model (CA_{1-bit}). The number of internal states of the CA_{1-bit} is assumed to be finite in the usual sense. The next state of each cell is determined by the present state of that cell and two binary 1-bit inputs from its left and right neighbor cells. Thus, the CA_{1-bit} can be thought of as one of the most powerless and the simplest models in a variety of CA's. A precise definition of the CA_{1-bit} can be found in Umeo [51] and Umeo and Kamikawa [57]. Mazoyer [26] and Nishimura and Umeo [32] each designed an optimum-time synchronization algorithm on the CA_{1-bit} based on Balzer's algorithm and Waksman's algorithm, respectively. Figure 22 shows a configuration of the 1-bit synchronization algorithm on 15 cells that is based on the design of Nishimura and Umeo [32]. Each cell has 78 internal states and 208 transition rules. The small black triangles ► and ◄ indicate a 1-signal transfer in the right or left direction, respectively, between neigh-

Q		Right State		
		Q	R	G	*
Left State	Q	Q	R	Q	Q
	R	Q	R		
	G	G		G	G
	*				

G		Right State		
		Q	R	G	*
Left State	Q	G		G	
	R	G	F	G	
	G	Q	R	Q	R
	*	G	F	G	

R		Right State		
		Q	R	G	*
Left State	Q	R	Q	R	R
	R	R	Q	R	R
	G		G	F	F
	*				

Firing Squad Synchronization Problem in Cellular Automata, Figure 19

Transition table for the four-state protocol based on Wolfram's Rule 60 (*left*) and its snapshots of configurations on 16 cells (*right*)
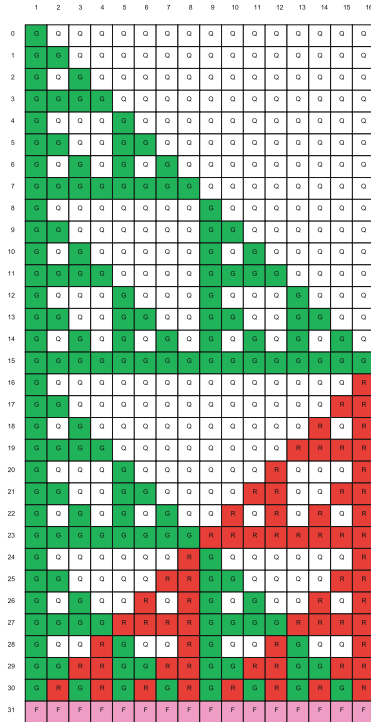
Q		Right State		
		Q	G	R	*
Left State	Q	Q	G	R	Q
	G	G	Q		G
	R	R		Q	
	*				

G		Right State		
		Q	G	R	*
Left State	Q	G	Q		
	G	Q	G	R	R
	R	G	G	F	
	*	G	G	F	

R		Right State		
		Q	G	R	*
Left State	Q	R	R	Q	R
	G		F	G	F
	R	Q	R	R	R
	*				

Firing Squad Synchronization Problem in Cellular Automata, Figure 20

Transition table for the four-state protocol based on Wolfram's Rule 150 (*left*) and its snapshots of configurations on 16 cells (*right*)

boring cells. A symbol in a cell shows an internal state of the cell.

Theorem 19 ([26,32]) *There exists a CA_{1-bit} that can synchronize n cells in optimum $2n - 2$ steps.*

Synchronization Algorithms for Multi-Bit Communication Cellular Automata

Section "Synchronization Algorithms for Multi-Bit Communication Cellular Automata" studies a trade-off between internal states and communication bits in firing squad synchronization protocols for k-bit communication-restricted cellular automata (CA_{k-bit}) and propose several time-optimum state-efficient bit-transfer-based synchronization protocols. It is shown that there exists a 1-state CA_{5-bit} that can synchronize any n cells in $2n - 2$ optimum-step. The result is interesting, since one knows that there exists no 4-state synchronization algorithm on *conventional* $O(1)$-*bit communication* cellular automata. A bit-transfer complexity is also introduced to measure the efficiency of synchronization protocols. It is shown that $\Omega(n \log n)$ bit-transfer is a lower-bound

for synchronizing n cells in $(2n - 2)$ steps. In addition, each optimum-time/non-optimum-time synchronization protocols presented has an $O(n^2)$ bit-transfer complexity, respectively. Most of the results presented here are from Umeo, Yanagihara and Kanazawa [68]. A computational relation between the conventional CA and CA_{k-bit} is stated as follows:

Lemma 20 ([57]) *Let N be any s-state conventional cellular automaton with time complexity $T(n)$. Then, there exists a CA_{1-bit} which can simulate N in $kT(n)$ steps, where k is a positive constant integer such that $k = \lceil \log_2 s \rceil$.*

Lemma 21 ([68]) *Let N be any s-state conventional cellular automaton. Then, there exists an s-state CA_{k-bit} which can simulate N in real time, where k is a positive integer such that $k = \lceil \log_2 s \rceil$.*

The following theorems show a trade-off between internal states and communication bits, and present state-efficient synchronization protocols $\mathcal{P}_i$, $1 \leq i \leq 5$. In some sense, no internal state is necessary to synchronize the whole array, as is shown in Theorem 27. The protocol design is based on the 6-state Mazoyer's algorithm given in Mazoyer [25].
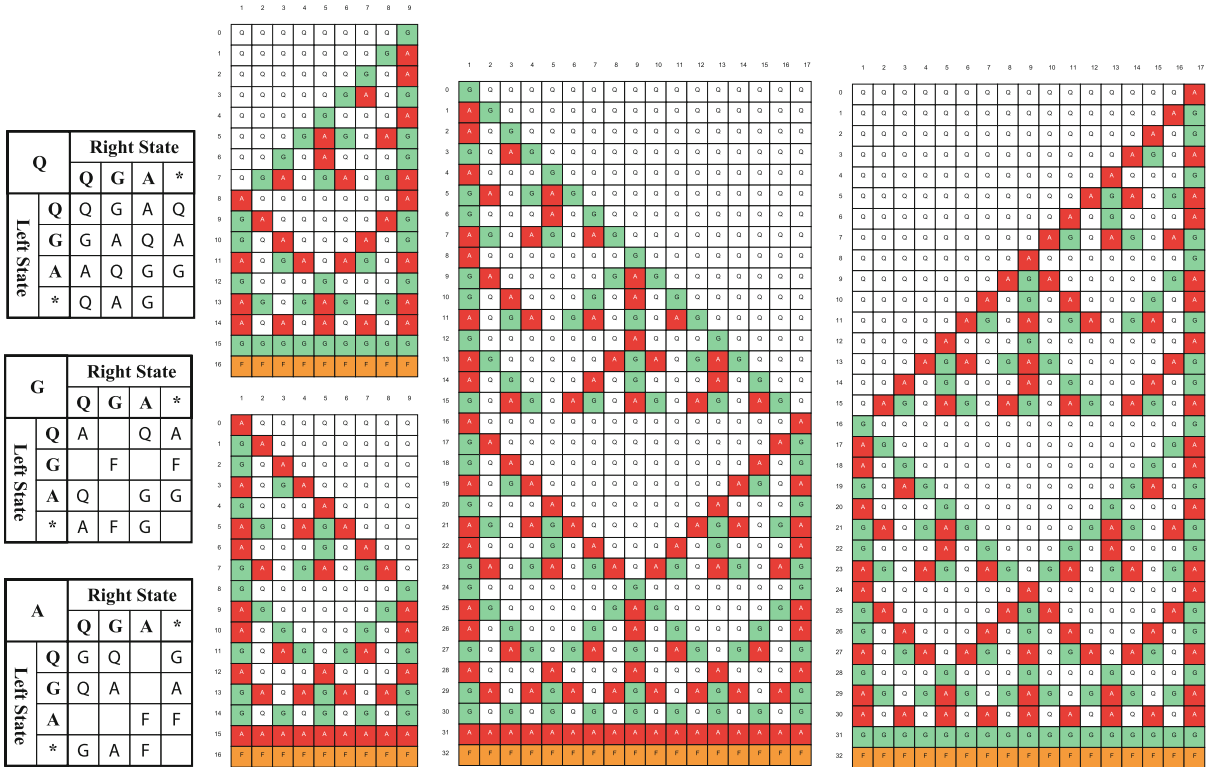
Firing Squad Synchronization Problem in Cellular Automata, Figure 21
Transition table for the four-state protocol (*left*) and its snapshots of configurations on 9 and 17 cells (*right*)

Firing Squad Synchronization Problem in Cellular Automata, Table 4
A comparison of optimum-time/non-optimum-time firing squad synchronization protocols for multi-bit communication cellular automata

Synchronization protocol	Communication bits transferred	# of states	# of Transition rules	Time complexity	One-/Two-sided recursiveness
$\mathcal{P}_1$	1	54	207	$2n-1$	One-sided
$\mathcal{P}_2$	2	6	60	$2n-2$	One-sided
$\mathcal{P}_3$	3	4	76	$2n-2$	One-sided
$\mathcal{P}_4$	4	3	87	$2n-2$	One-sided
$\mathcal{P}_4^*$	4	2	88	$2n-2$	One-sided
$\mathcal{P}_5$	5	1	114	$2n-2$	One-sided
Mazoyer [26]	1	58	–	$2n-2$	Two-sided
Mazoyer [26]	12	3	–	$2n-2$	–
Nishimura and Umeo [32]	1	78	208	$2n-2$	Two-sided

Theorem 22 ([68]) *There exists a 54-state CA_{1-bit} with protocol $\mathcal{P}_1$ that can synchronize any n cells in $2n-1$ optimum-step.*

Theorem 23 ([68]) *There exists a 6-state CA_{2-bit} with protocol $\mathcal{P}_2$ that can synchronize any n cells in $2n-2$ optimum-step.*

Theorem 24 ([68]) *There exists a 4-state CA_{3-bit} with protocol $\mathcal{P}_3$ that can synchronize any n cells in $2n-2$ optimum-step.*

Figure 23 illustrates snapshots of the 6-state $(2n-2)$-step synchronization protocol $\mathcal{P}_2$ operating on CA_{2-bit} (left) and for the 4-state protocol $\mathcal{P}_3$ operating on CA_{3-bit} with 24 cells (right).
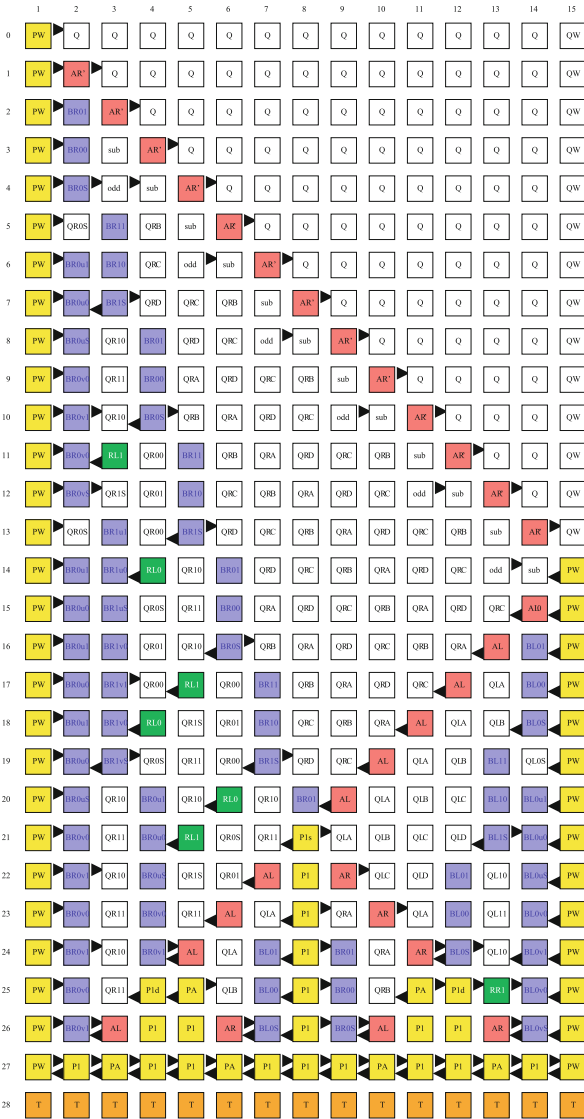
Firing Squad Synchronization Problem in Cellular Automata, Figure 22

A configuration of optimum-time synchronization algorithm with 1-bit inter-cell communication on 15 cells

Theorem 25 ([68]) *There exists a 3-state $CA_{4-\text{bit}}$ with protocol $\mathcal{P}_4$ that can synchronize any n cells in $2n-2$ optimum-step.*

Theorem 26 ([68]) *There exists a 2-state $CA_{4-\text{bit}}$ with protocol $\mathcal{P}_4^*$ that can synchronize any n cells in $2n-2$ optimum-step.*

Theorem 27 ([68]) *There exists a 1-state $CA_{5-\text{bit}}$ with protocol $\mathcal{P}_5$ that can synchronize any n cells in $2n-2$ optimum-step.*

Figures 24 and 25 illustrate snapshots of the 3-state $(2n-2)$-step synchronization protocol $\mathcal{P}_4$ operating on $CA_{4-\text{bit}}$ (left) and for the 1-state protocol $\mathcal{P}_5$ operating on $CA_{5-\text{bit}}$ (right).

Let $BT(n)$ be total number of bits transferred which is needed for synchronizing n cells on $CA_{k-\text{bit}}$. By using the similar technique developed by Vollmar [73], a lower-bound on bit-transfer complexity can be established for synchronizing n cells on $CA_{k-\text{bit}}$ in a way such that $BT(n) = \Omega(n \log n)$. In addition, it is shown that each synchronization protocol $\mathcal{P}_i$, $1 \le i \le 5$ and $\mathcal{P}_4^*$ presented above has an $O(n^2)$ bit-transfer complexity, respectively.

Theorem 28 ([68]) *$\Omega(n \log n)$ bit-transfer is a lower bound for synchronizing n cells on $CA_{k-\text{bit}}$ in $(2n-2)$ steps.*

Theorem 29 ([68]) *Each optimum-time/non-optimum-time synchronization protocols $\mathcal{P}_i$, $1 \le i \le 5$ and $\mathcal{P}_4^*$ has an $O(n^2)$ bit-transfer complexity, respectively.*

Firing Squad Synchronization Problem on Two-Dimensional Arrays

Figure 26 shows a finite two-dimensional (2-D) cellular array consisting of $m \times n$ cells. The array operates in lock-step mode in such a way that the next state of each cell (except border cells) is determined by both its own present state and the present states of its north, south, east and west neighbors. All cells (*soldiers*), except the north-west corner cell (*general*), are initially in the quiescent state at time $t = 0$ with the property that the next state of a quiescent cell with quiescent neighbors is the quiescent state again. At time $t = 0$, the north-west corner cell C_{11} is in the *fire-when-ready* state, which is the initiation signal for the synchronization. The firing squad synchronization problem is to determine a description (state set and next-state function) of cells that ensures all cells enter the *fire* state at exactly the same time and for the first time. The set of states and transition function must be independent of m and n.

Several synchronization algorithms on 2-D arrays have been proposed by Beyer [3], Grasselli [14], Shinahr [44], Szwerinski [45], Umeo, Maeda and Fujiwara [59], and Umeo, Hisaoka and Akiguchi [53]. Most of the synchronization algorithms for 2-D arrays are based on mapping schemes which map synchronized configurations for 1-D arrays onto 2-D arrays. Section "Firing Squad Synchronization Problem on Two-Dimensional Arrays" presents
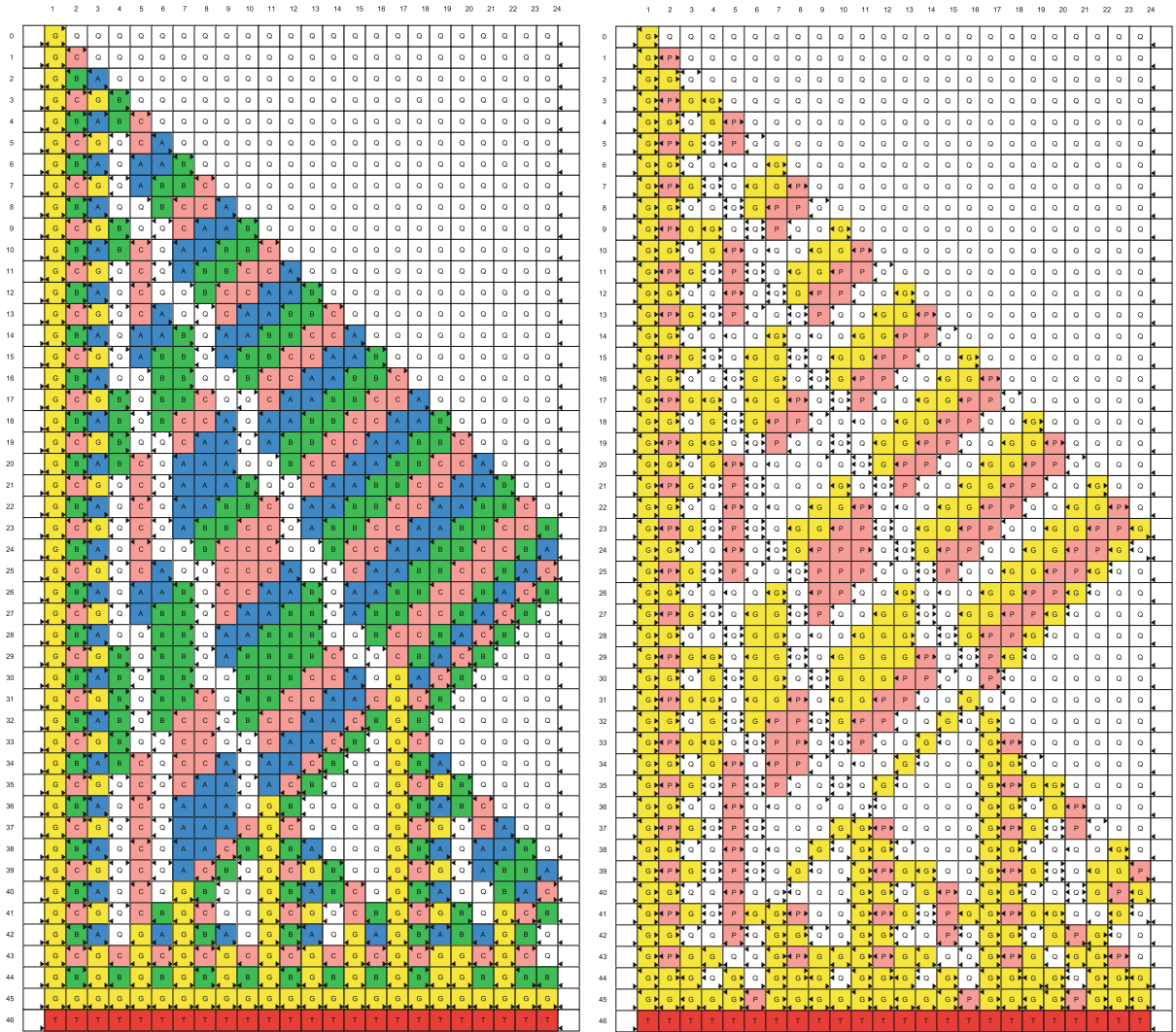
Firing Squad Synchronization Problem in Cellular Automata, Figure 23

Snapshots for the 6-state protocol $\mathcal{P}_\in$ operating on CA_{2-bit} with 24 cells (*left*) and for the 4-state protocol $\mathcal{P}_3$ operating on CA_{3-bit} with 24 cells (*right*)

such several mapping schemes that yield time-efficient 2-D synchronization algorithms.

Orthogonal Mapping:
A Simple Linear-Time Algorithm

In this section, a very simple synchronization algorithm is provided for 2-D arrays. The overall of the algorithm is as follows:

1. First, *synchronize* the first column cells using a usual optimum-step 1-D algorithm with a general at one end, thus requiring $2m - 2$ steps.

2. Then, *start the row synchronization operation* on each row simultaneously. Additional $2n - 2$ steps are required for the row synchronization. Totally, its time complexity is $2(m + n) - 4$ steps.

The implementation is referred to as *orthogonal mapping*. It is shown that $s + 2$ states are enough for the implementation of the algorithm above, where s is the number of internal states of the 1-D base algorithm. Figure 27 shows snapshots of the 8-state synchronization algorithm running on a rectangular array of size 4×6.

Firing Squad Synchronization Problem in Cellular Automata, Figure 24
Snapshots for 3-state $(2n-2)$-step synchronization protocol $\mathcal{P}_4$ operating on CA_{4-bit} with 24 cells
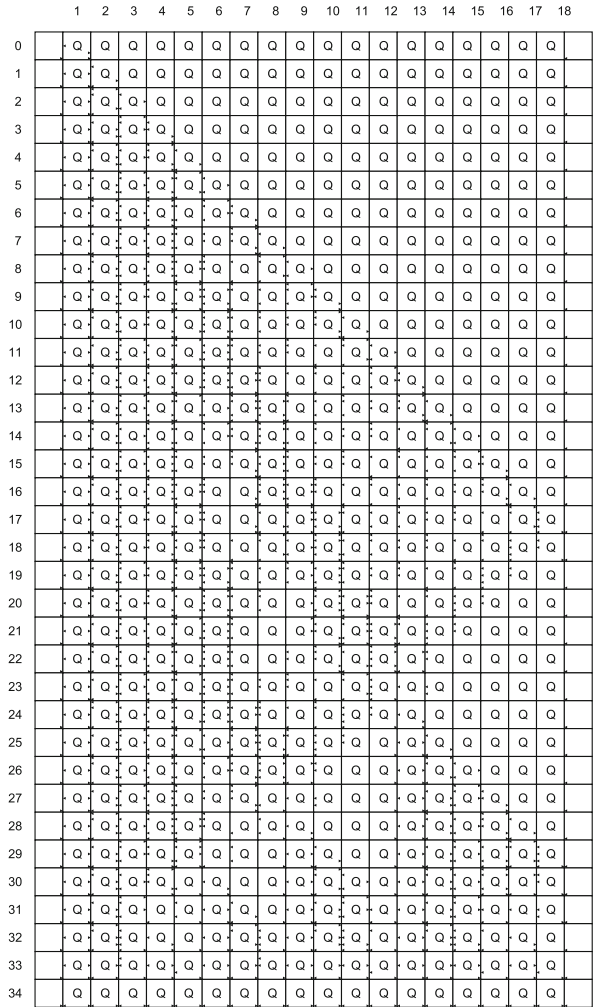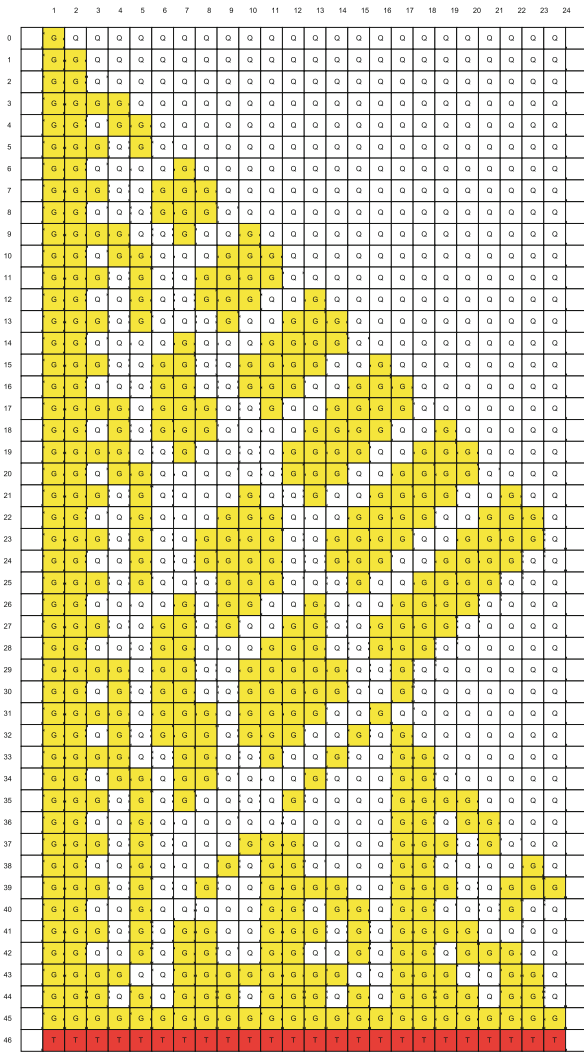
Firing Squad Synchronization Problem in Cellular Automata, Figure 25
Snapshots for 1-state $(2n-2)$-step synchronization protocol $\mathcal{P}_5$ operating on CA_{5-bit} with 18 cells

Theorem 30 *There exists an $(s+2)$-state protocol for synchronizing any $m \times n$ rectangular arrays in $2(m+n)-4$ steps, where s is number of states of any optimum-time 1-D synchronization protocol.*

L-shaped Mapping:
Shinar's Optimum-Time Algorithm

The first optimum-time 2-D synchronization algorithm was developed by Shinar [44] and Beyer [3]. The rectangular array of size $m \times n$ is regarded as $\min(m, n)$ rotated (90° in clockwise direction) L-shaped 1-D arrays, where

they are synchronized independently using the generalized firing squad synchronization algorithm. The configuration of the generalized synchronization on 1-D array can be mapped on 2-D array. See Fig. 28. Thus, an $m \times n$ array synchronization problem is reduced to independent $\min(m, n)$ 1-D generalized synchronization problems such that: $\mathcal{P}(m, m+n-1)$, $\mathcal{P}(m-1, m+n-3)$, ..., $\mathcal{P}(1, n-m+1)$ in the case $m \leq n$ and $\mathcal{P}(m, m+n-1)$, $\mathcal{P}(m-1, m+n-3)$, ..., $\mathcal{P}(m-n+1, m-n+1)$ in the case $m > n$, where $\mathcal{P}(k, \ell)$ means the 1-D generalized synchronization problem for ℓ cells with a general on the kth cell from left end. Beyer [3] and Shinahr [44] have shown that an optimum-time complexity for syn-
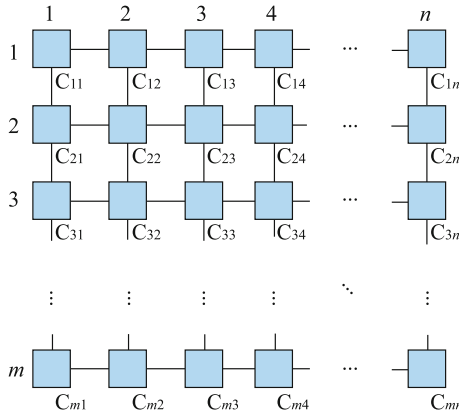
Firing Squad Synchronization Problem in Cellular Automata, Figure 26
A two-dimensional cellular automaton

chronizing any $m \times n$ arrays is $m + n + \max(m, n) - 3$ steps. Shinahr [44] has also given a 28-state implementation.

Theorem 31 ([3,44]) *There exists an optimum-time 28-state protocol for synchronizing any $m \times n$ rectangular arrays in $m + n + \max(m, n) - 3$ steps.*

Diagonal Mapping I: Six-State Linear-Time Algorithm

The proposal is a simple and state-efficient mapping scheme that enables us to embed any 1-D firing squad synchronization algorithm with a general at one end onto two-dimensional arrays without introducing additional states. Consider a 2-D array of size $m \times n$, where $m, n \geq 2$. Firstly, divide mn cells on the array into $m + n - 1$ groups

$g_k, 1 \leq k \leq m + n - 1$, defined as follows.

$$g_k = \{C_{i,j} | (i - 1) + (j - 1) = k - 1\}, \quad \text{i. e.,}$$
$$g_1 = \{C_{1,1}\}, \quad g_2 = \{C_{1,2}, C_{2,1}\},$$
$$g_3 = \{C_{1,3}, C_{2,2}, C_{3,1}\}, \quad \dots, \quad g_{m+n-1} = \{C_{m,n}\}.$$

Figure 29 shows the division of the two-dimensional array of size $m \times n$ into $m + n - 1$ groups. Let M be any 1-D CA that synchronizes ℓ cells in $T(\ell)$ steps. Assume that M has $m + n - 1$ cells, denoted by C_i, where $1 \leq i \leq m + n - 1$. Then, consider a one-to-one correspondence between the ith group g_i and the ith cell C_i on M such that $g_i \leftrightarrow C_i$, where $1 \leq i \leq m + n - 1$. One can construct a 2-D CA N such that all cells in g_i simulate the ith cell C_i in real-time and N can synchronize any $m \times n$ arrays at time $t = T(m + n - 1)$ if and only if M synchronizes 1-D arrays of length $m + n - 1$ at time $t = T(m + n - 1)$. It is noted that the set of internal states of N constructed is the same as M. Thus an $m \times n$ 2-D array synchronization problem is reduced to one 1-D synchronization problem with the general at the left end. The algorithm obtained is slightly slower than the optimum ones, but the number of internal states is considerably smaller. Figure 30 shows snapshots of the proposed 6-state linear-time firing squad synchronization algorithm on rectangular arrays. For the details of the construction of the transition rule set, see Umeo, Maeda, Hisaoka and Teraoka [60].

Theorem 32 ([60]) *Let A be any s-state firing synchronization algorithm operating in $T(\ell)$ steps on 1-D ℓ cells. Then, there exists a 2-D s-state cellular automaton that can synchronize any $m \times n$ rectangular array in $T(m + n - 1)$ steps.*
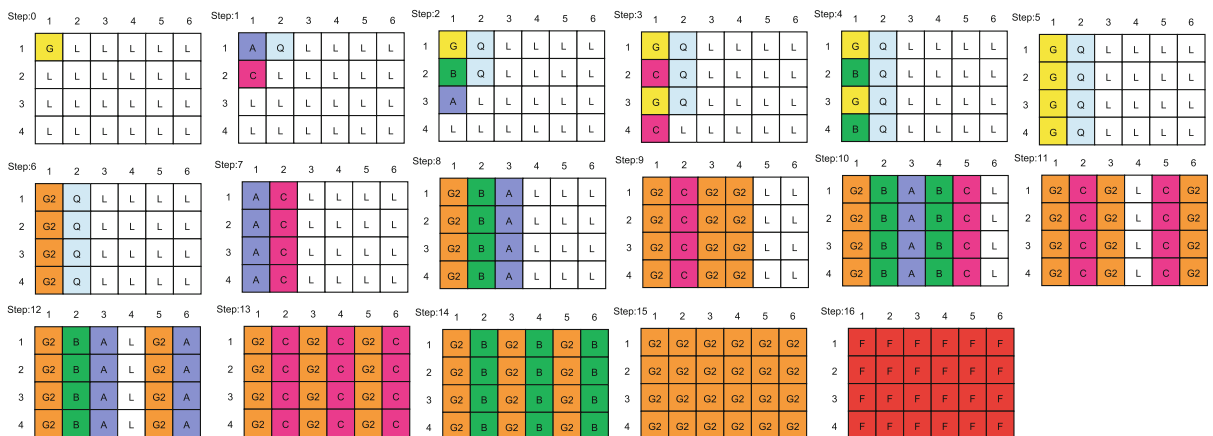
Firing Squad Synchronization Problem in Cellular Automata, Figure 27
Snapshots of the synchronization process on 4 × 6 array
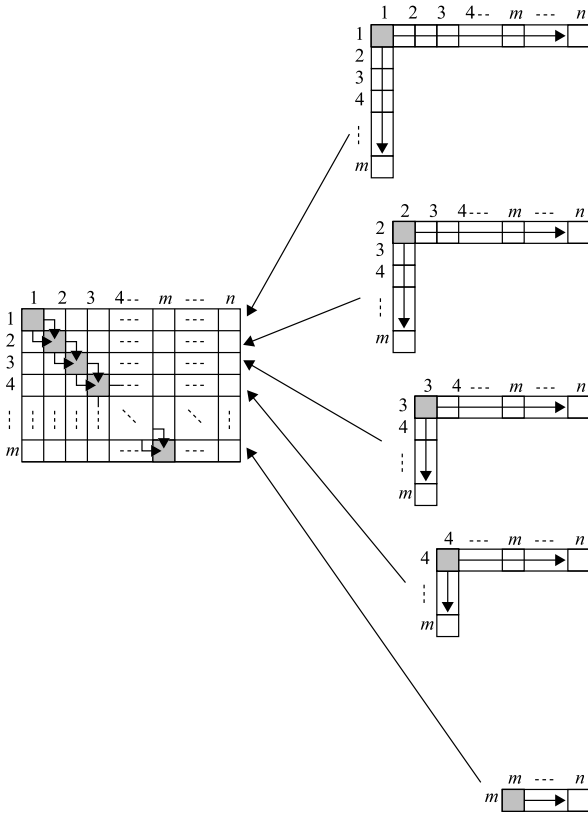
Firing Squad Synchronization Problem in Cellular Automata, Figure 28

An optimum-time synchronization scheme for rectangular arrays

Firing Squad Synchronization Problem in Cellular Automata, Figure 29

A correspondence between 1-D and 2-D arrays

Theorem 33 ([60]) *There exists a 6-state 2-D CA that can synchronize any $m \times n$ rectangular array in $2(m + n) - 4$ steps.*

Theorem 34 ([60]) *There exists a 6-state 2-D CA that can synchronize any $m \times n$ rectangular array containing isolated rectangular holes in $2(m + n) - 4$ steps.*

Theorem 35 ([60]) *There exists a 6-state firing squad synchronization algorithm that can synchronize any 3-D $m \times n \times \ell$ solid arrays in $2(m + n + \ell) - 6$ steps.*

Diagonal Mapping II:
Twelve-State Time-Optimum Algorithm

The second diagonal mapping scheme in this section enables us to embed a special class of 1-D generalized synchronization algorithm onto two-dimensional arrays without introducing additional states. An $m \times n$ 2-D array synchronization problem is reduced to one 1-D generalized synchronization problem: $\mathcal{P}(m, m + n - 1)$. Di-

vide mn cells into $m + n - 1$ groups g_k defined as follows, where k is any integer such that $-(m - 1) \le k \le n - 1$:

$$g_k = \{C_{i,j} | j - i = k\}, \quad -(m - 1) \le k \le n - 1 .$$

Figure 31 shows the correspondence between 1-D and 2-D arrays.

Property $\mathcal{A}$: Let S_i^t denote the state of C_i at step t. It is said that a generalized firing algorithm has a *property $\mathcal{A}$*, where any state S_i^t appearing *in the area $\mathcal{A}$* can be computed from its left and right neighbor states S_{i-1}^{t-1} and S_{i+1}^{t-1} but it never depends on its own previous state S_i^{t-1}. Figure 32 shows the area $\mathcal{A}$ in the time-space diagram for the generalized optimum-step firing squad synchronization algorithm.

Any one-dimensional generalized firing squad synchronization algorithm with the property $\mathcal{A}$ can be easily embedded onto two-dimensional arrays without introducing additional states.
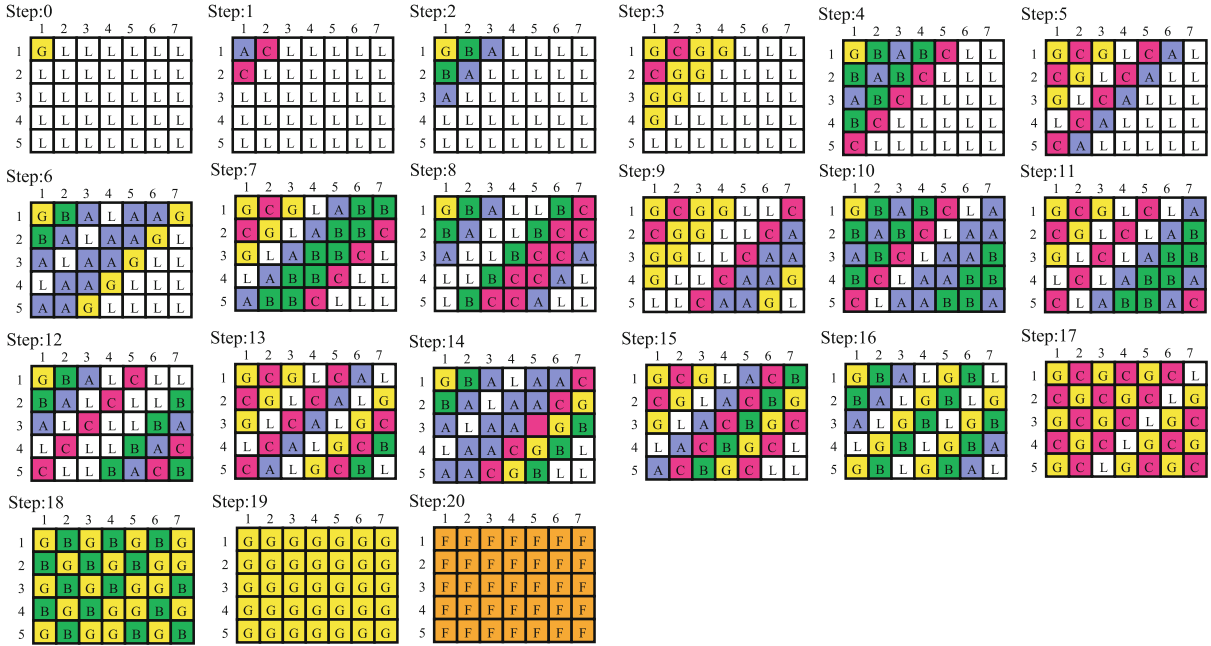
Firing Squad Synchronization Problem in Cellular Automata, Figure 30
Snapshots of the proposed 6-state linear-time firing squad synchronization algorithm on rectangular arrays

Theorem 36 ([53]) *Let M be any s-state generalized synchronization algorithm with the property $\mathcal{A}$ operating in $T(k, \ell)$ steps on 1-D ℓ cells with a general on the kth cell from the left end. Then, based on M, one can construct a 2-D s-state cellular automaton that can synchronize any $m \times n$ rectangular array in $T(m, m + n - 1)$ steps.*

It has been shown in Umeo, Hisaoka and Akiguchi [53] that there exists a 12-state implementation of the generalized optimum-time synchronization algorithms having the property $\mathcal{A}$. Then, one can get a 12-state optimum-time synchronization algorithm for rectangular arrays. Figure 33 shows snapshots of the proposed 12-state optimum-time firing squad synchronization algorithm operating on a 7×9 array.

Theorem 37 ([53]) *There exists a 12-state firing squad synchronization algorithm that can synchronize any $m \times n$ rectangular array in optimum $m + n + \max(m, n) - 3$ steps.*

Rotated L-Shaped Mapping: Time-Optimum Algorithm

In this section we present an optimum-time synchronization algorithm based on a rotated L-shaped mapping. The synchronization scheme is quite different from previous designs. The scheme uses the freezing-thawing technique. Without loss of generality, it is assumed that $m \leq n$.

A rectangular array of size $m \times n$ is regarded as m rotated (90° in counterclockwise direction) *L-shaped* 1-D arrays. Each L-shaped array is denoted by $L_i, 1 \leq i \leq m$. See Fig 34. Each L_i consists of three segments of length i, $n - m$, and i, respectively. Each segment can be synchronized by the freezing-thawing technique. Synchronization operations for $L_i, 1 \leq i \leq m$ are as follows: Figure 35 shows a time-space diagram for synchronizing L_i. The wake-up signals for the three segments of L_i are generated at time $t = m + 2(m - i) - 1, 3m - i - 2$, and $n + 2(m - i) - 1$, respectively. Synchronization operations on each segments are delayed for $\Delta t_{i_j}, 1 \leq j \leq 3$ such that:

$$\Delta t_{i_j} = \begin{cases} 2(n - m) & j = 1 \\ i & j = 2 \\ n - m & j = 3 \end{cases} . \qquad (1)$$

The synchronization for the first segment of L_i is started at time $t = m + 2(m - i) - 1$ and its operations are delayed for $\Delta t = \Delta t_{i_1} = 2(n - m)$ steps. Now letting $t_0 = m + 2(m - i) - 1, \Delta t = \Delta t_{i_1} = 2(n - m)$ in freezing-thawing technique, the first segment of L_i can be synchronized at time $t = t_0 + 2i - 2 + \Delta t = m + 2n - 3$. In a similar way, the second and the third segments can be synchronized at time $t = m + 2n - 3$. Thus, L_i can be synchronized at time $t = m + 2n - 3$. Figure 36 shows some snapshots of the synchronization process operating

Firing Squad Synchronization Problem in Cellular Automata, Figure 31
Correspondence between 1-D and 2-D cellular arrays
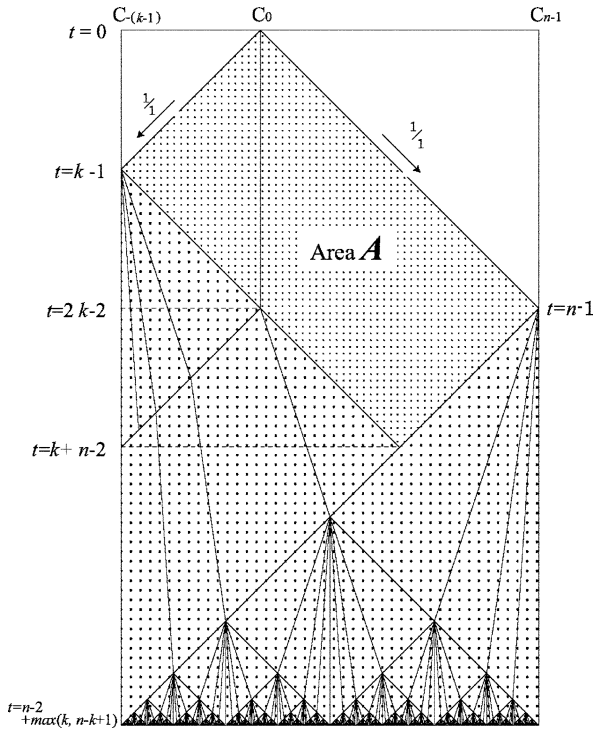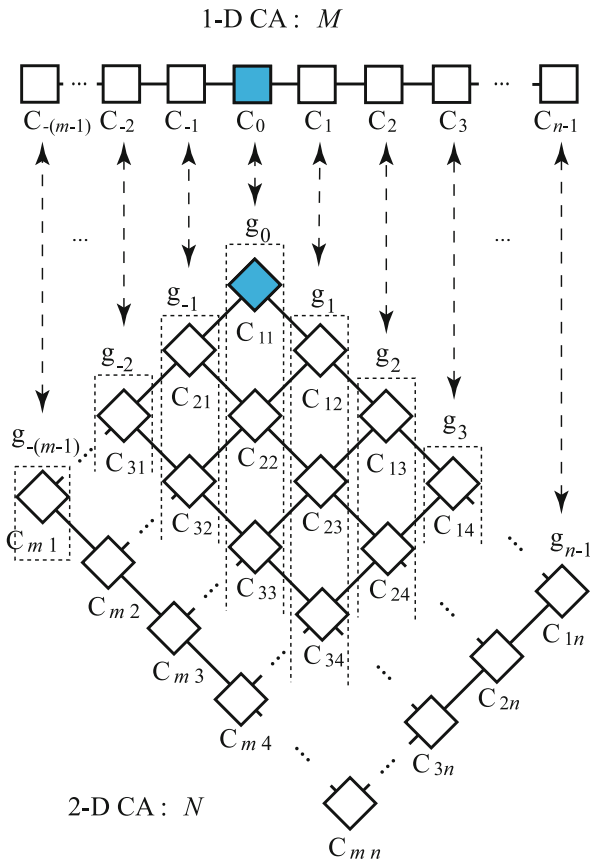
Firing Squad Synchronization Problem in Cellular Automata, Figure 32
Time-space diagram for generalized optimum-step firing squad synchronization algorithm

in optimum-steps on 5×8 array. Now the next theorem can be established.

Theorem 38 ([65]) *The algorithm above can synchronize any $m \times n$ rectangular array in optimum $m + n + \max(m, n) - 3$ steps.*

Frame Mapping: Time-Optimum Algorithm

Section "Frame Mapping: Time-Optimum Algorithm" presents an optimum-time 2-D synchronization algorithm based on frame mapping. Without loss of generality, it is assumed that $m \le n$. A rectangular array of size $m \times n$ is regarded as consisting of rectangle-shaped *frames* of width 1. See Fig. 37. Each frame L_i, $0 \le i \le \lceil m/2 \rceil - 1$, is divided into six segments and these six segments are synchronized using the freezing-thawing technique. The length of each segment of L_i is $m - 2i$, $m - 2i$, $n - m$, $m - 2i$, $m - 2i$, and $n - m$, respectively. Figure 38 shows a time-space diagram of the synchronization operations

for the outermost frame L_0. Synchronization operations on jth segment of L_0 are delayed for Δt_{0_j} steps, $1 \le j \le 6$, such that:

$$\Delta t_{0_j} = \begin{cases} 2(n - m) & j = 1 \\ 2(n - m) & j = 2 \\ m & j = 3 \\ n - m & j = 4 \\ n - m & j = 5 \\ m & j = 6 \end{cases}. \qquad (2)$$

Using the freezing-thawing technique, L_0 can be synchronized at time $t = m + 2n - 3$. The synchronization operation for L_i, $i \ge 1$ can be done similarly. Note that the ith frame is of size $(m - 2i) \times (n - 2i)$. Let T_i be steps required for synchronizing the ith frame with the synchronization operations given above starting at time $t = 0$. Then, $T_i = m + 2n - 3 - 6i = T_0 - 6i$, for any i such that $0 \le i \le \lceil m/2 \rceil - 1$. Thus, $T_i - T_{i-1} = 6$. Therefore, the starting time for synchronizing each frame is delayed for 6 steps so that synchronization operations for each frame can be finished simultaneously. In order to start the
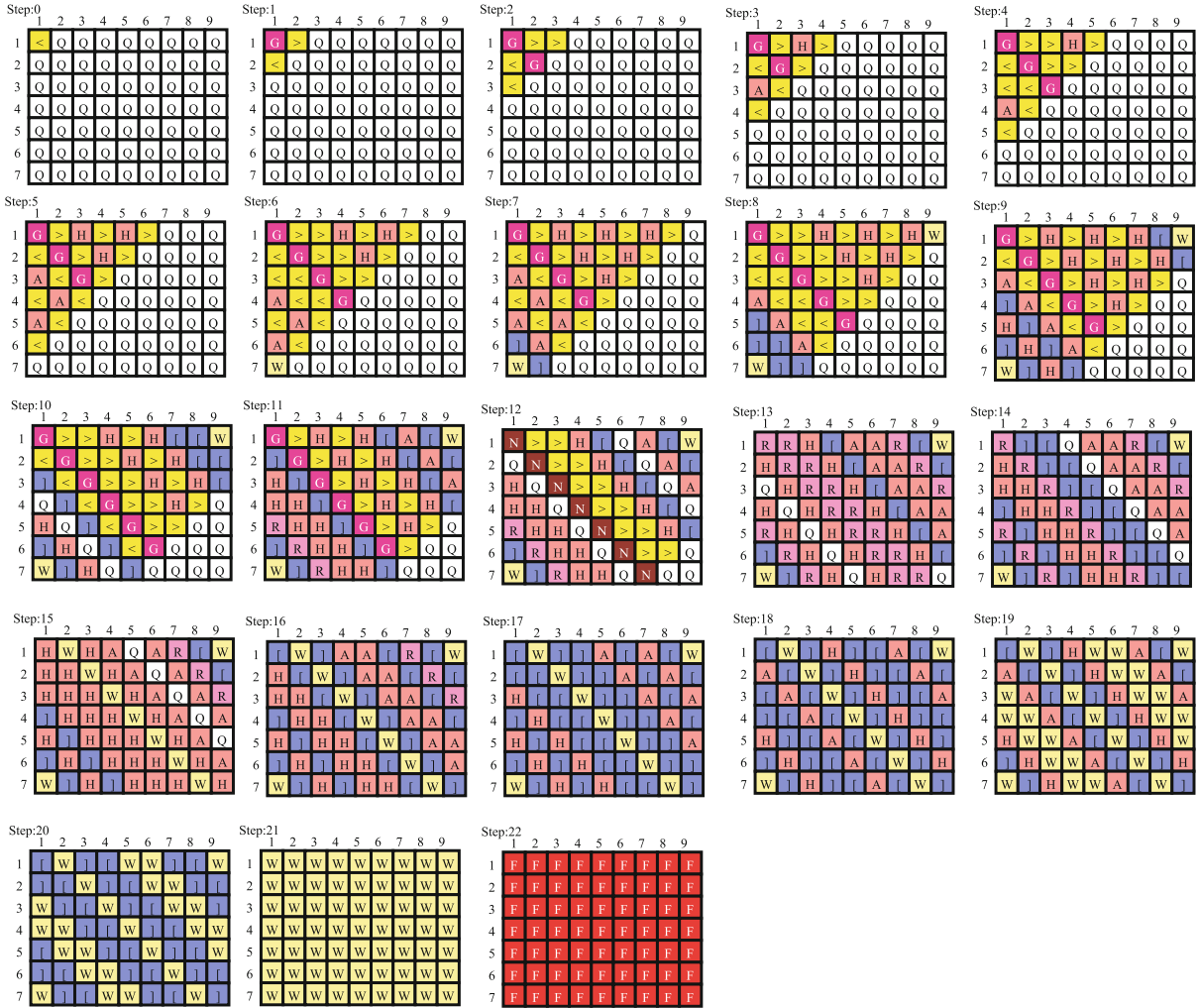
Firing Squad Synchronization Problem in Cellular Automata, Figure 33
Snapshots of the proposed 12-state optimum-time firing squad synchronization algorithm on rectangular arrays
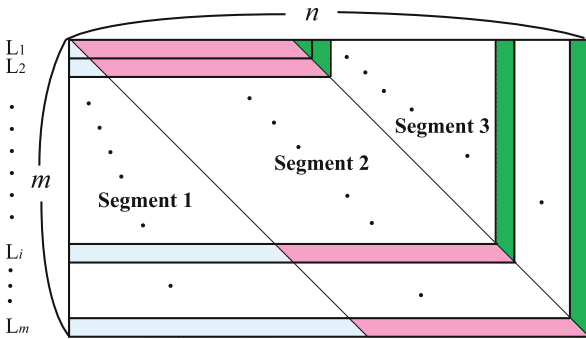
Firing Squad Synchronization Problem in Cellular Automata, Figure 34
A two-dimensional array of size $m \times n$ is regarded as consisting of m rotated (90° in counterclockwise direction) L-shaped 1-D array

operation progressively, an activating signal that travels in the diagonal direction is given to each north-west corner of each L_i at time $t = 6i$. In this way all of the frames can be synchronized. Figure 39 illustrates some snapshots of the synchronization process operating in optimum-steps on 5×8 array.

Theorem 39 ([66]) *The algorithm based on frame mapping can synchronize any $m \times n$ rectangular array in $m + n + \max(m, n) - 3$ optimum steps.*

Generalized Firing Squad Synchronization Algorithms for Two-Dimensional Rectangular Cellular Automata

Szwerinski [45] presented a first optimum-time generalized firing squad synchronization algorithm for rectangular arrays. Umeo, Hisaoka, Teraoka and Maeda [55]
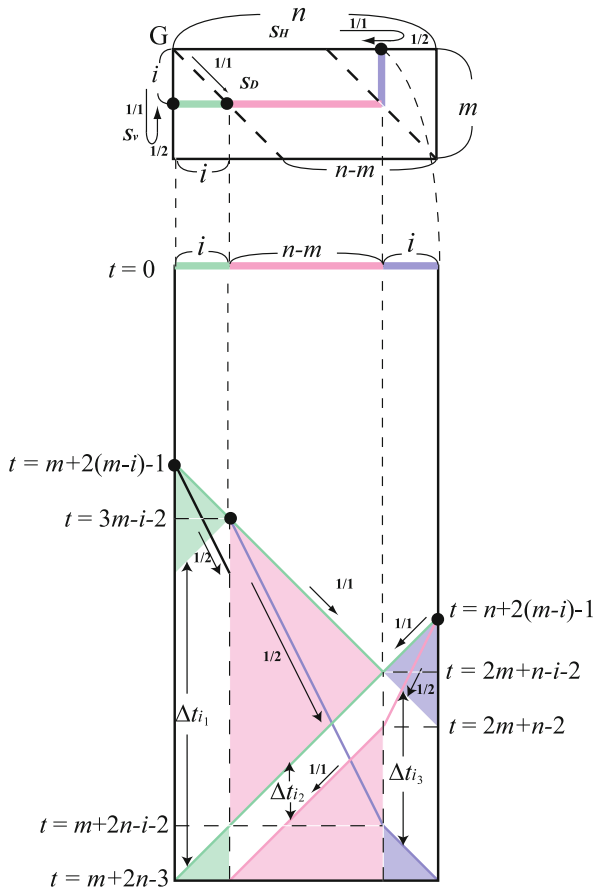
Firing Squad Synchronization Problem in Cellular Automata, Figure 35

Time-space diagram for synchronizing L$_i$

proposed a new optimum-time generalized synchronization algorithm. Section "Generalized Firing Squad Synchronization Algorithms for Two-Dimensional Rectangular Cellular Automata" presents the generalized optimum-time algorithm based on Umeo et al. [55]. The algorithm can synchronize any rectangular arrays of size $m \times n$ with the general being located at any position (r, s) on the array in $m + n + \max(m, n) - \min(r, m - r + 1) - \min(s, n - s + 1) - 1$ optimum steps, where $1 \le r \le m, 1 \le s \le n$. The following theorem is a useful technique for delaying the generalized synchronization on 1-D arrays.

Theorem 40 ([55]) *Let A be any 1-D cellular automaton that runs a generalized $T(s, n)$-step synchronization algorithm on n cells with a general on $C_s (1 \le s \le n)$ and t_0, t_1, t_2, ℓ be any integer satisfying the following conditions such that $\Delta t = t_2 - t_1 = 2\ell$, $\ell \ge 1$, $t_2 > t_1 \ge t_0 \ge 0$ and $t_1 + t_2 - 2t_0 \le 2T(s, n) - 2\max(s, n - s + 1) + 2$. It is*

assumed that three special signals are given to cell C_s at step $t = t_0$, t_1, and t_2. These signals play an important role of initializing the generalized synchronization process, starting the delayed operation, and stopping the delayed operation, respectively. Then, one can construct a cellular automaton B that synchronizes the array at time $t = t_0 + \ell + T(s, n)$. In the case where $T(s, n)$ is an optimum time complexity such that $T(s, n) = n - 2 + \max(s, n - s + 1)$, the constructed B synchronizes at time $t = t_0 + \ell + n - 2 + \max(s, n - s + 1)$. Thus the synchronization operation is delayed for ℓ steps. Figure 40 is a time-space diagram for the delayed optimum-time generalized synchronization algorithm operating on n cells. In the darker area of the diagram, each cell simulates the operation of A at speed 1/2 by repeating a simulate-one-step of A then keep-the-state operations alternatively at each step.

Consider the case where $m \le n$, $1 \le r < \lceil m/2 \rceil$ and $1 \le s < \lceil n/2 \rceil$. An array of size $m \times n$ is regarded as consisting of m rows of length n. An optimum-time generalized synchronization algorithm with a general at $C_{i,s}$ $(1 \le i \le m)$ is used for the synchronization of the ith row. The operation is referred to as *row-synchronization*. To synchronize all rows simultaneously, an efficient timing control scheme shown in Fig. 41 is developed. Figure 41 is a time-space diagram for giving special signals to each cell on the s-column. These special signals act as a timing $t = t_0$, t_1 and t_2 stated in Theorem 40. For example, the row-synchronization on the yth ($r \le y \le m$) row is started at time $t = t_0 = t_1 = y - r$ and the process is delayed from time $t = t_1 = y - r$ to $t = t_2 = 2m - y - r$, shown in the darker area in Fig. 41. Thus $\ell = m - y$. Based on Theorem 40 the yth row is synchronized at time $t = m + 2n - r - s - 1$. Figure 42 is a time-space diagram for the row-synchronization on the xth and yth row, where $1 \le x < r$ and $r \le y \le m$. The readers can see how all rows are synchronized at time $t = m + 2n - r - s - 1$.

Thus, the following theorem can be developed.

Theorem 41 ([55]) *In the row-synchronization process, all of the rows can be synchronized simultaneously at time $t = m + 2n - r - s - 1$ in the case $m \le n, 1 \le r < \lceil m/2 \rceil$ and $1 \le s < \lceil n/2 \rceil$.*

In the column-synchronization process, all of the cells take a firing state prior to the row-synchronization, but the column-synchronization fails to synchronize. Symmetrically, the following theorem holds in the case where the array is longer than is wide.

Theorem 42 ([55]) *In the column-synchronization process, all of the columns can be synchronized simultane-*
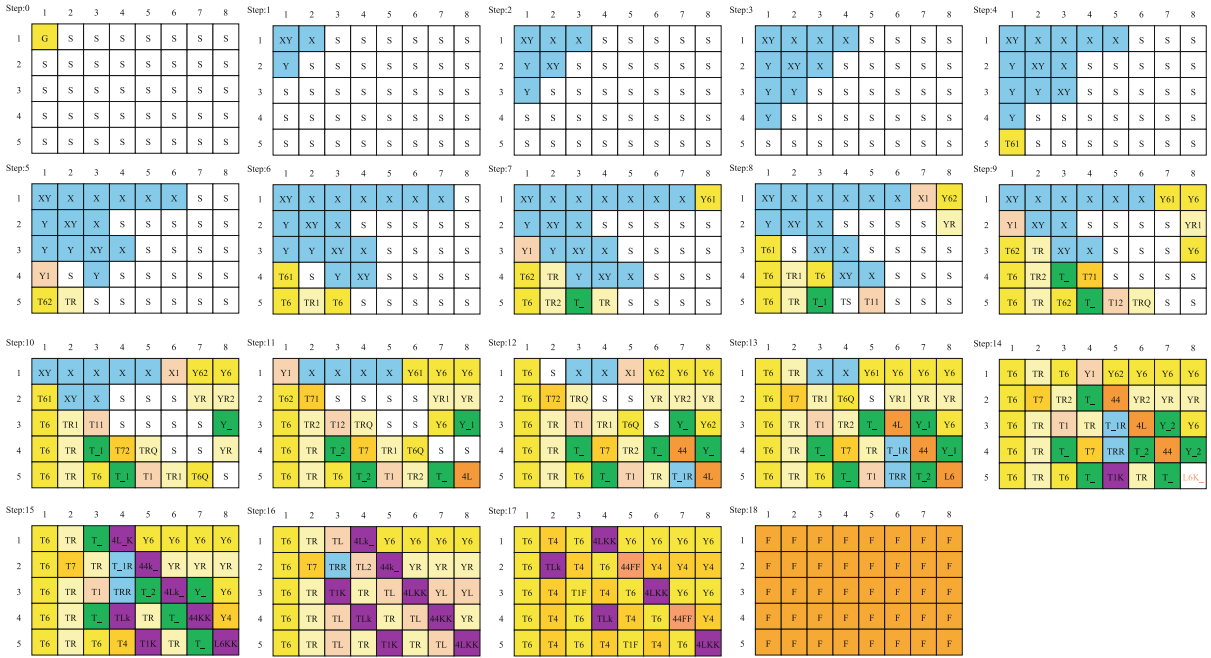
Firing Squad Synchronization Problem in Cellular Automata, Figure 36
Snapshots of the synchronization process on 5 × 8 array
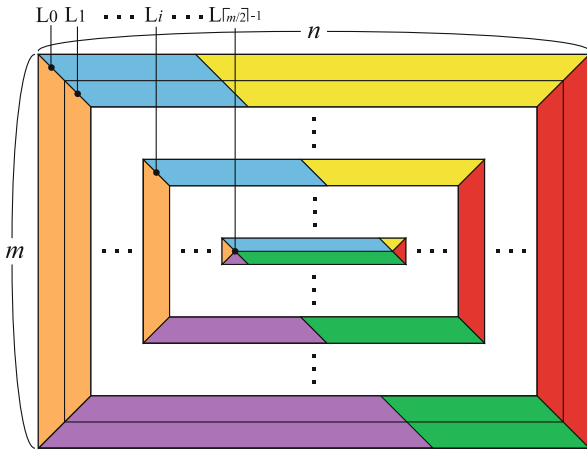
Firing Squad Synchronization Problem in Cellular Automata, Figure 37
A two-dimensional array of size m × n is regarded as consisting of ⌈m/2⌉ frames

Firing Squad Synchronization Problem in Cellular Automata, Figure 38
Time-space diagram for synchronizing L_0

ously at time $t = n + 2m - r - s - 1$ in the case $m \geq n$, $1 \leq r < \lceil m/2 \rceil$ and $1 \leq s < \lceil n/2 \rceil$.

To synchronize the array in optimum-time, the array performs both row- and column-synchronization operations. Each cell should take a firing state on each layer at two different times. The first one is false and it should be ignored. The second one is the right firing state. By combining the Theorems 41, 42, the following theorem can be established.
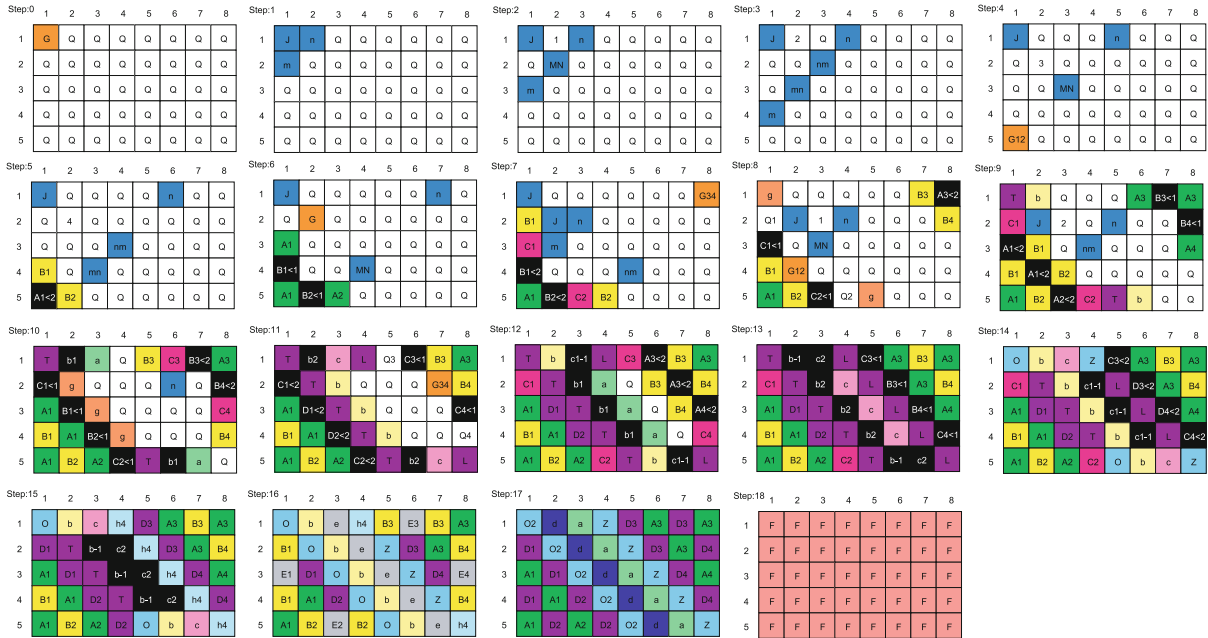
Firing Squad Synchronization Problem in Cellular Automata, Figure 39
Snapshots of the synchronization process on 5 × 8 array

Theorem 43 ([55]) *The scheme given above can synchronize any $m \times n$ array in $m + n + \max(m, n) - \min(r, m - r + 1) - \min(s, n - s + 1) - 1$ optimum steps, where (r, s) is the general's initial position such that $1 \le r \le m, 1 \le s \le n$.*

The number of internal states of an automaton realizing the Szwerinski's algorithm is 25,600. Umeo, Maeda, Hisaoka and Teraoka [60] presented a 14-state implementation for a non-optimum-time generalized algorithm. The 2-D generalized synchronization algorithm is $\max(r + s, m + n - r - s + 2) - \max(m, n) + \min(r, m - r + 1) + \min(s, n - s + 1) - 3$ steps larger than the optimum algorithm proposed by Szwerinski [45]. However, the number of internal states required to yield the synchronizing condition is the smallest known at present. Snapshots of the 14-state generalized synchronization algorithm running on a rectangular array of size 6×8 with the general at $C_{2,3}$ are shown in Fig. 43.

Theorem 44 ([60]) *There exists a 14-state 2-D CA that can synchronize any $m \times n$ rectangular array in $m + n + \max(r + s, m + n - r - s + 2) - 4$ steps with the general at an arbitrary initial position (r, s).*

Firing Squad Synchronization Algorithms for Two-Dimensional Square Cellular Automata

An optimum-time square synchronization algorithm has been proposed by Shinahr [44]. The algorithm operates as follows: By dividing the entire square array into n rotated L-shaped 1-D arrays such that the length of the ith L is $2n - 2i + 1$ $(1 \le i \le n)$, one treats the square synchronization as n independent 1-D synchronizations with the general located at the center cell. On the ith L, a general is generated at $C_{i,i}$ at time $t = 2i - 2$, and the general initiates the horizontal and vertical synchronizations on the row and column arrays via an optimum-time synchronization algorithm. The array can be synchronized in optimum time $t = 2i - 2 + 2(n - i + 1) - 2 = 2n - 2$. It has been shown in Umeo, Maeda and Fujiwara [59] that 9 states are sufficient for the optimum-time square synchronization. The implementation is based on Mazoyer's 6-state algorithm. Figure 44 shows snapshots of configurations of the 9-state synchronization algorithm running on a square of size 8×8.

Theorem 45 ([44,59]) *There exists a 9-state 2-D CA that can synchronize any $n \times n$ square array in $2n - 2$ steps.*

Firing Squad Synchronization Algorithms for Two-Dimensional One-Bit Communication Cellular Automata

The firing squad synchronization problem for 2-D one-bit communication cellular automata has been studied by Torre, Napoli and Parente [49], Umeo, Michisaka and Kamikawa [62], Gruska, Torre, and Parente [17],
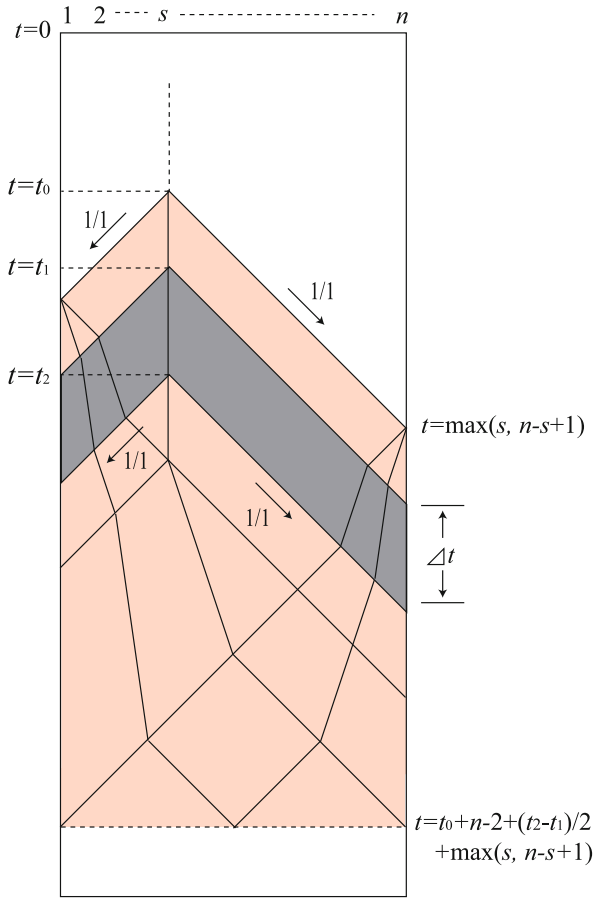
Firing Squad Synchronization Problem in Cellular Automata, Figure 40
Time-space diagram for delayed optimum-time generalized synchronization algorithm

Firing Squad Synchronization Problem in Cellular Automata, Figure 41
Time-space diagram for delaying row synchronization

and Umeo, Michisaka, Kamikawa and Kanazawa [63]. Section "Firing Squad Synchronization Algorithms for Two-Dimensional One-Bit Communication Cellular Automata" presents two implementations of the square and rectangular synchronization algorithms on CA_{1-bit}. The first one is for square arrays given in Umeo, Michisaka, Kamikawa and Kanazawa [63]. It runs in $(2n - 1)$ steps on $n \times n$ square arrays. The proposed implementation is one step slower than optimum-time for the O(1)-bit communication model. The total numbers of internal states and transition rules of the CA_{1-bit} are 127 and 405, respectively. Figure 45 shows snapshots of configurations of the 127-state implementation running on a square of size 8×8. Gruska, Torre, and Parente [17] presented an optimum-time algorithm.

Theorem 46 ([17]) *There exists a 2-D CA_{1-bit} that can synchronize any $n \times n$ square arrays in $2n - 2$ steps.*

Umeo, Michisaka, Kamikawa and Kanazawa [63] has also implemented the rectangular synchronization algorithm for 2-D CA_{1-bit}. The total numbers of internal states and transition rules of the CA_{1-bit} are 862 and 2217, respectively. Figure 46 shows snapshots of the synchronization process on a 5×8 rectangular array.

Theorem 47 ([63]) *There exists a 2-D CA_{1-bit} that can synchronize any $m \times n$ rectangular arrays in $m + n + \max(m, n)$ steps.*

Summary and Future Directions

The present article has examined via computer the state transition rule sets for which optimum-time synchronization algorithms were designed over the past forty years. The smallest transition rule sets for the well-known firing squad synchronization algorithms are useful and important for researchers who might have interests in those transition rule sets that realize the classical optimum-time firing algorithms quoted frequently in the literatures. It has also presented a survey and a comparison of the quantitative and qualitative aspects of the optimum-time
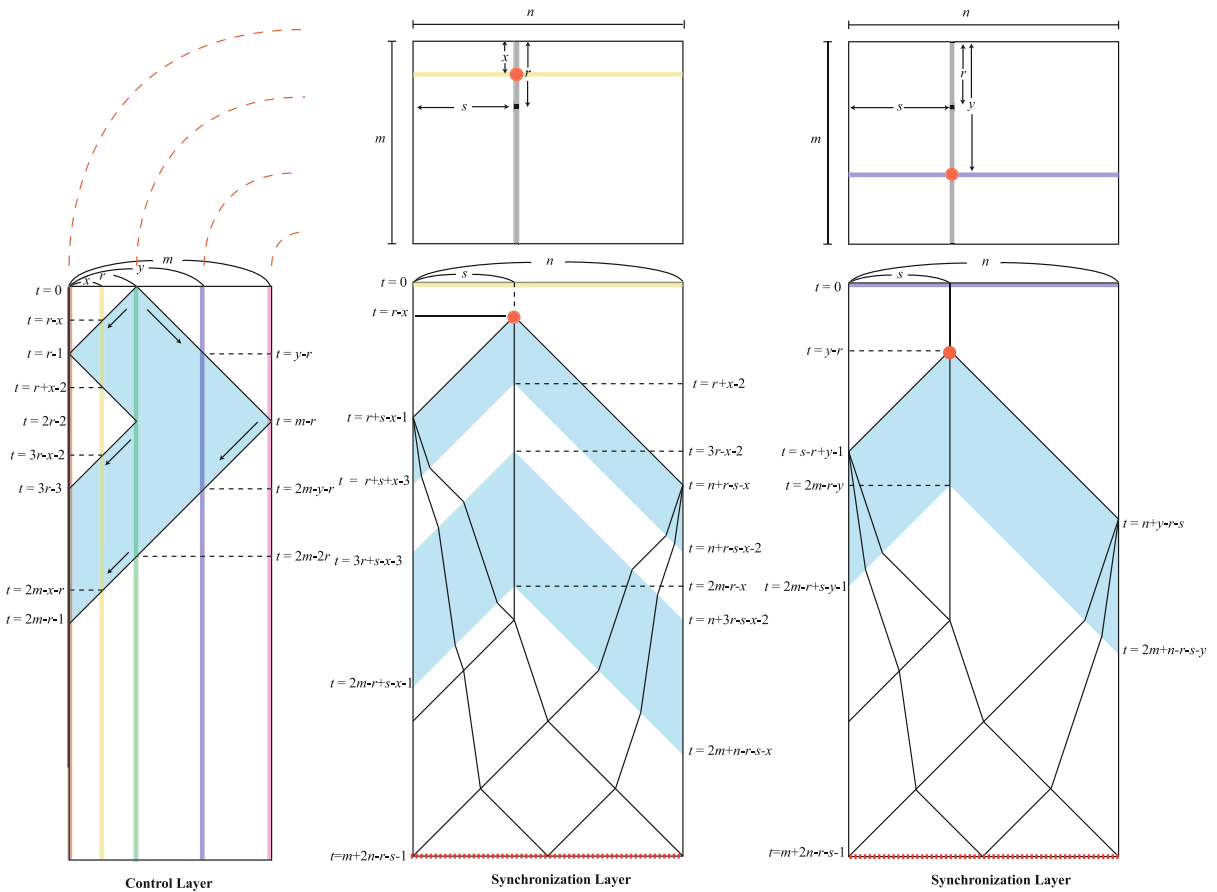
Firing Squad Synchronization Problem in Cellular Automata, Figure 42
Time-space diagram for the row-synchronization on the xth and yth row of a rectangular array of size $m \times n$, where $m \leq n$, $1 \leq r < \lceil m/2 \rceil$, $1 \leq s < \lceil n/2 \rceil$, $1 \leq x < r$ and $r \leq y \leq m$

synchronization algorithms developed thus far for one-dimensional cellular arrays. It has studied several variants of the firing squad synchronization problems including fault-tolerant synchronization protocols, 4- and 5-state partial solutions, one-bit communication protocols and non-optimum-time algorithms etc. Finally a survey on two-dimensional firing squad synchronization algorithms has been given. Several new results and new viewpoints have been provided. The question: "What is the minimum number of states for an optimum-time solution of the problem?" still remains open. A new approach should be required.

Bibliography

Primary Literature

1. Balzer R (1967) An 8-state minimal time solution to the firing squad synchronization problem. Inf Control 10:22–42

2. Berthiaume A, Bittner T, Perković L, Settle A, Simon J (2004) Bounding the firing synchronization problem on a ring. Theor Comput Sci 320:213–228

3. Beyer WT (1969) Recognition of topological invariants by iterative arrays. Ph.D. Thesis, MIT, pp 144

4. Burns JE, Lynch NA (1987) The Byzantine firing squad problem. In: Advances in Computing Research: Parallel and Distributed Computing. JAI press, 4:147–161

5. Coan BA, Dolev D, Dwork C, Stockmeyer L (1989) The distributed firing squad problem. SIAM J Comput 18(5):990–1012

6. Culik K II (1989) Variations of the firing squad synchronization problem and applications. Inf Process Lett 30:153–157

7. Culik K II, Dube S (1991) An efficient solution of the firing mob problem. Theor Comput Sci 91:57–69

8. Fischer PC (1965) Generation of primes by a one-dimensional real-time iterative array. J ACM 12(3):388–394

9. Gerken HD (1987) über Synchronisations-Probleme bei Zellularautomaten. Diplomarbeit, Institut für Theoretische Informatik, Technische Universität Braunschweig

10. Goldstein D, Kobayashi K (2004) On the complexity of network synchronization. In: Proc. of ISSAC 2004. LNCS, vol 3341. Springer, Berlin, pp 496–507
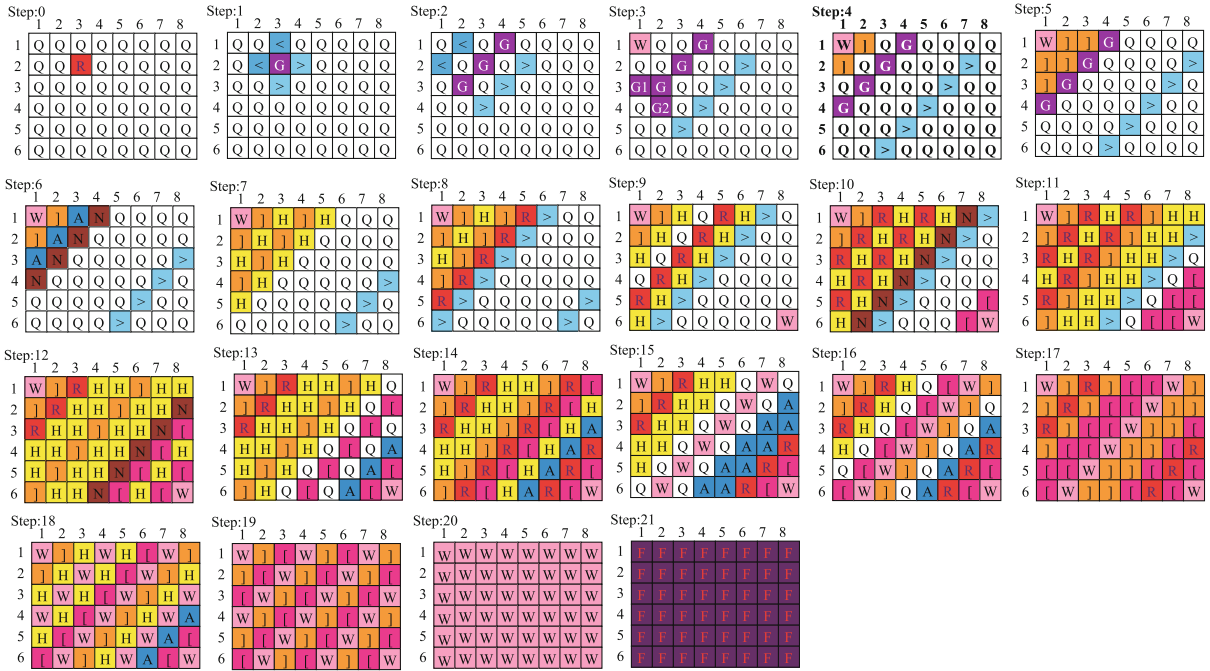
Firing Squad Synchronization Problem in Cellular Automata, Figure 43
Snapshots of the 14-state linear-time generalized firing squad synchronization algorithm on rectangular arrays
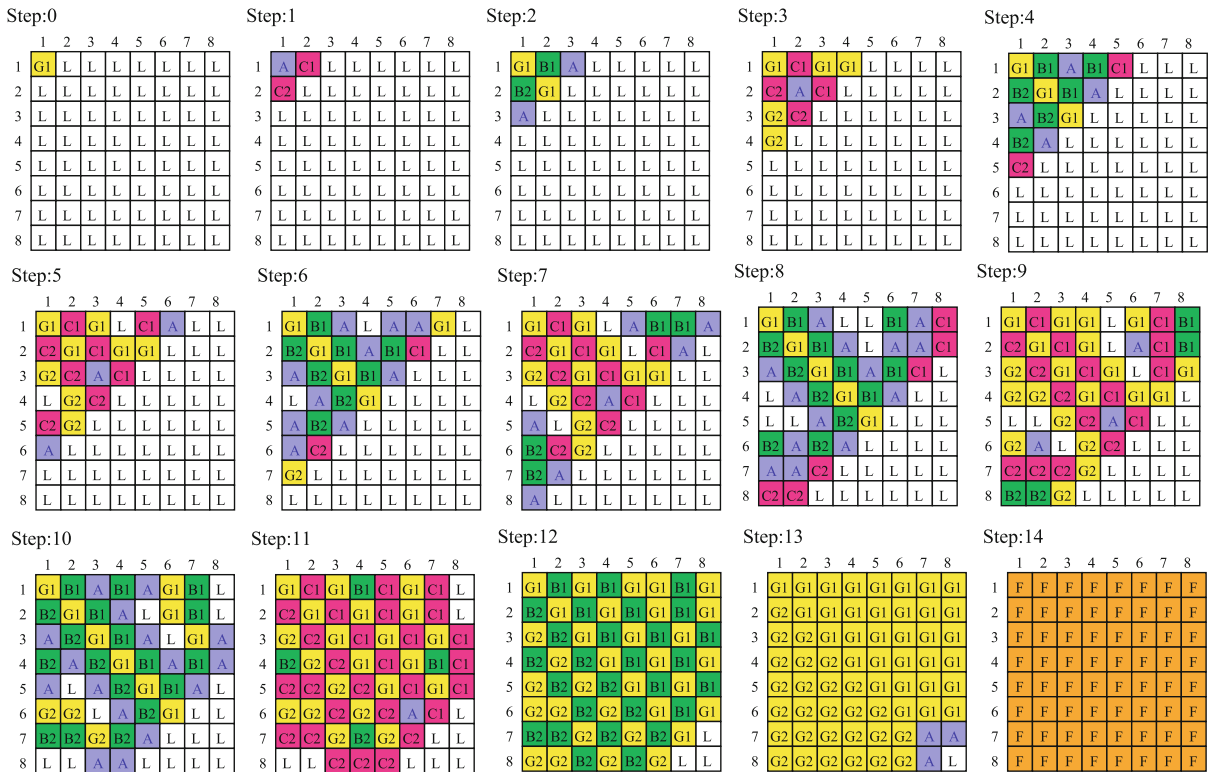
Firing Squad Synchronization Problem in Cellular Automata, Figure 44
A configuration of a 9-state implementation of optimum-time firing on square arrays
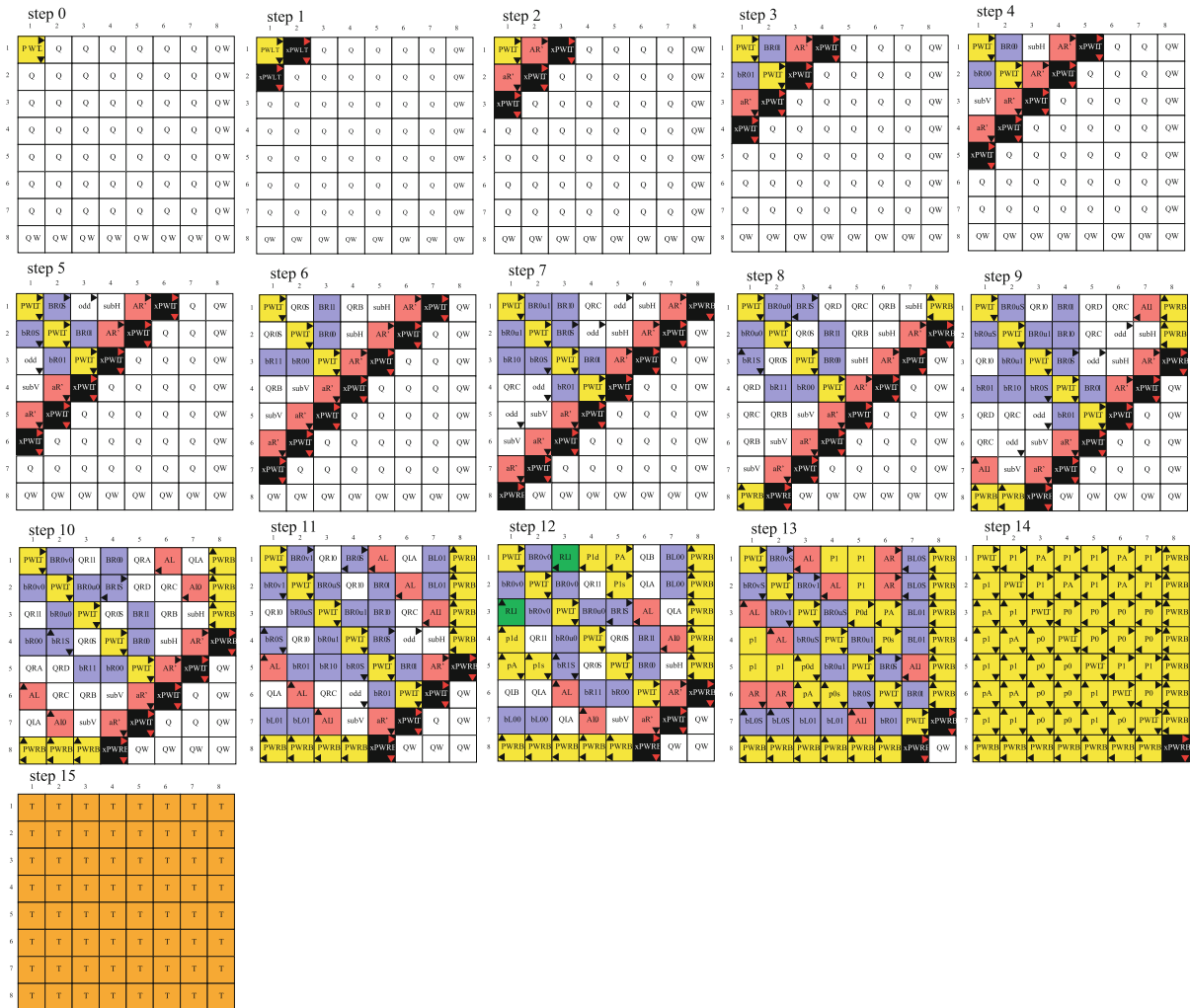
Firing Squad Synchronization Problem in Cellular Automata, Figure 45
Snapshots of the ($2n − 1$)-step square synchronization algorithm with the general on the northwest corner

11. Goldstein D, Meyer N (2002) The wake up and report problem is asymptotically time-equivalent to the firing squad synchronization problem. In: Proc. of 13th Annual ACM-SIAM Symp. on Discrete Algorithms, pp 578–587

12. Goto E (1962) A minimal time solution of the firing squad problem. Dittoed course notes for Applied Mathematics 298, Harvard University, pp 52–59, with an illustration in color

13. Goto E (1966) Some puzzles on automata. In: Kitagawa T (ed) Toward computer sciences. Kyouritsu, pp 67–91 (in Japanease)

14. Grasselli A (1975) Synchronization of cellular arrays: The firing squad problem in two dimensions. Inf Control 28:113–124

15. Grefenstette JJ (1983) Network structure abd the firing squad synchronization problem. J Comput Syst Sci 26:139–152

16. Grigorieff S (2006) Synchronization of a bounded degree graph of cellular automata with non uniform delays in time $\Delta \lfloor \log_m(\delta) \rfloor$. Theor Comput Sci 356(1):170–185

17. Gruska J, Torre SL, Parente M (2007) The firing squad synchronization problem on squares, toruses and rings. Int J Found Comput Sci, 18(3):637–654

18. Herman GT, Liu WH, Rowland S, Walker A (1974) Synchronization of growing cellular systems. Inf Control 25:103–122

19. Imai K, Morita K (1996) Firing squad synchronization problem in reversible cellular automata. Theor Comput Sci 165: 475–482

20. Jiang T (1992) The synchronization of nonuniform networks of finite automata. Inf Comput 97:234–261

21. Kobayashi K (1977) The firing squad synchronization problem for two-dimensional arrays. Inf Control 34:177–197

22. Kutrib M, Vollmar R (1991) Minimal time synchronization in restricted defective cellular automata. J Inf Process Cybern ELK 27:179–196

23. Kutrib M, Vollmar R (1995) The firing squad synchronization problem in defective cellular automata. IEICE Trans Inf Syst E78-D(7):895–900
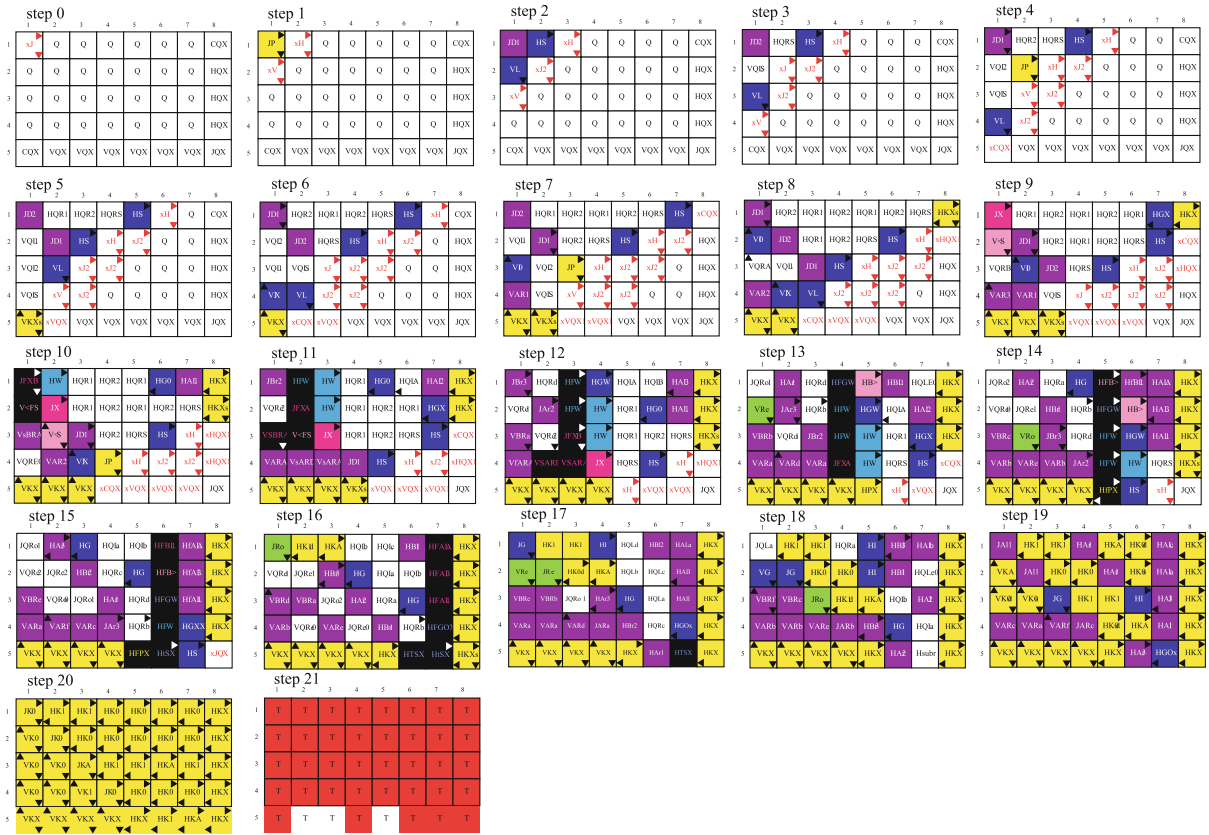
Firing Squad Synchronization Problem in Cellular Automata, Figure 46
Snapshots of the proposed rectangular firing squad synchronization algorithm with the general at the northwest corner

24. Mazoyer J (1986) An overview of the firing squad synchronization problem. Lecture Notes on Computer Science, vol 316. Springer, Berlin

25. Mazoyer J (1987) A six-state minimal time solution to the firing squad synchronization problem. Theor Comput Sci 50:183–238

26. Mazoyer J (1996) On optimal solutions to the firing squad synchronization problem. Theor Comput Sci 168:367–404

27. Mazoyer J (1997) A minimal-time solution to the FSSP without recursive call to itself and with bounded slope of signals. Draft, p 8

28. Minsky M (1967) Computation: Finite and infinite machines. Prentice Hall, Englewood Cliffs

29. Moore EF (1964) The firing squad synchronization problem. In: Moore EF (ed) Sequential Machines, Selected Papers, Addison-Wesley, Reading MA

30. Moore FR, Langdon GG (1968) A generalized firing squad problem. Inf Control 12:212–220

31. Nguyen HB, Hamacher VC (1974) Pattern synchronization in two-dimensional cellular space. Inf Control 26:12–23

32. Nishimura J, Umeo H (2005) An optimum-time synchronization protocol for 1-bit-communication cellular automaton. In: Proc. of the 9th World Multi-Conference on Systemics, Cybernetics and Informatics, vol X, pp 232–237

33. Nishitani Y, Honda N (1981) The firing squad synchronization problem. Theor Comput Sci 14:39–61

34. Noguchi K (2004) Simple 8-state minimal time solution to the firing squad synchronization problem. Theor Comput Sci 314:303–334

35. Roka Z (1995) The firing squad synchronization problem on Cayley graphs. In: Proc. of MFCS'95. LNCS, vol 969. Springer, Berlin, pp 402–411

36. Romani F (1976) Cellular automata synchronization. Inform Sci 10:299–318

37. Romani F (1977) On the fast synchronization of tree connected networks. Inform Sci, 12:229–244

38. Romani F (1978) The parallelism principle: speeding up the cellular automata synchronization. Inf Control 36:245–255

39. Rosenstiehl P, Fiksel JR, Holliger A (1972) Intelligent graphs: networks of finite automata capable of solving graph problems. In: Graph Theory and Computing, Academic Press, New York, pp 219–265

40. Sanders P (1994) Massively parallel search for transition-tables of polyautomata. In: Jesshope C, Jossifov V, Wilhelmi W (eds) Proc. of the VI International Workshop on Parallel Processing by Cellular Automata and Arrays. Akademie, Berlin, pp 99–108

41. Schmid H, Worsch T (2004) The firing squad synchronization

problem with many generals for one-dimensional. In: CA Proc. of IFIP World Congress, pp 111–124

42. Settle A, Simon J (1998) Improved bounds for the firing synchronization problem. In: SIROCCO 5: Proc. of the 5th International Colloquium on Structural Information and Communication Complexity. Carleton Scientific, Ontario, pp 66–81

43. Settle A, Simon J (2002) Smaller solutions for the firing squad. Theor Comput Sci 276:83–109

44. Shinahr I (1974) Two- and three-dimensional firing squad synchronization problems. Inf Control 24:163–180

45. Szwerinski H (1982) Time-optimum solution of the firing-squad-synchronization-problem for n-dimensional rectangles with the general at an arbitrary position. Theor Comput Sci 19:305–320

46. Torre SL, Napoli M, Parente D (1996) Synchronization of one-way connected processors. Complex Syst 10:239–255

47. Torre SL, Napoli M, Parente D (1998) Synchronization of a line of identical processors at a given time. Fundamenta Informaticae 34:103–128

48. Torre SL, Napoli M, Parente M (2000) A compositional approach to synchronize two dimensional networks of processors. Theor Inf Appl 34:549–564

49. Torre SL, Napoli M, Parente M (2001) Firing squad synchronization problem on bidimensional cellular automata with communication constraints. In: Proc. of MCU 2001. LNCS, vol 2055. Springer, Berlin, pp 264–275

50. Umeo H (1996) A note on firing squad synchronization algorithms-A reconstruction of Goto's first-in-the-world optimum-time firing squad synchronization algorithm. In: Kutrib M, Worsch T (eds) Proc. of Cellular Automata Workshop, pp 65

51. Umeo H (2001) Linear-time recognition of connectivity of binary images on 1-bit inter-cell communication cellular automaton. Parallel Computing, 27:587–599

52. Umeo H (2004) A simple design of time-efficient firing squad synchronization algorithms with fault-tolerance. IEICE Trans Inf Syst E87-D(3):733–739

53. Umeo H, Hisaoka M, Akiguchi S (2005) A twelve-state optimum-time synchronization algorithm for two-dimensional rectangular arrays. In: Proc. of the 4th International Conference on Unconventional Computation: UC 2005. LNCS, vol 3699. Springer, Berlin, pp 214–223

54. Umeo H, Hisaoka M, Sogabe T (2005) A survey on optimum-time firing squad synchronization algorithms for one-dimensional cellular automata. Int J Unconv Comput 1: 403–426

55. Umeo H, Hisaoka M, Teraoka M, Maeda M (2005) Several new generalized linear- and optimum-time synchronization algorithms for two-dimensional rectangular arrays. In: Margenstern M (ed) Proc. of the 4th International Conference on Machines, Computations and Universality: MCU 2004. LNCS, vol 3354. Springer, Berlin, pp 223–232

56. Umeo H, Hisaoka M, Michisaka K, Nishioka K, Maeda M (2002) Some new generalized synchronization algorithms and their implementations for large scale cellular automata. In: Proc. of the 3rd International Conference on Unconventional Models of Computation: UMC 2002. LNCS, vol 2509. Springer, Berlin, pp 276–286

57. Umeo H, Kamikawa N (2003) Real-time generation of primes by a 1-bit-communication cellular cutomaton. Fundamenta Informaticae 58:421–435

58. Umeo H, Kamikawa N, Yunès JB (2008) A family of smallest symmetrical four-state firing squad synchronization protocols for one-dimensional ring celluar automata. Luniver Press, Beckington, pp 174–186

59. Umeo H, Maeda M, Fujiwara N (2002) An efficient mapping scheme for embedding any one-dimensional firing squad synchronization algorithm onto two-dimensional arrays. In: Proc. of the 5th International Conference on Cellular Automata for Research and Industry. LNCS, vol 2493. Springer, Berlin, pp 69–81

60. Umeo H, Maeda M, Hisaoka M, Teraoka M (2006) A state-efficient mapping scheme for designing two-dimensional firing squad synchronization algorithms. Fundamenta Informaticae 74(4):603–623

61. Umeo H, Maeda M, Hongyo K (2006) A design of symmetrical six-state $3n$-step firing squad synchronization algorithms and their implementations. In: Proc. of 7th International Conference on Cellular Automata for Research and Industry, ACRI2006. LNCS, vol 4173. Springer, Berlin, pp 157–168

62. Umeo H, Michisaka K, Kamikawa N (2003) A synchronization problem on 1-bit-communication cellular automata. In: Proc. of International Conference on Computational Science-ICCS2003. LNCS, vol 2657. Springer, Berlin, pp 492–500

63. Umeo H, Michisaka K, Kamikawa N, Kanazawa M (2007) State-efficient one-bit-communication solutions for some classical cellular automata problems. Fundamenta Informaticae 78:449–465

64. Umeo H, Sogabe T, Nomura Y (2000) Correction, optimization and verification of transition rule set for Waksman's firing squad synchronization algorithm. In: Proc. of the Fourth Intern. Conference on Cellular Automata for Research and Industry. Springer, London, pp 152–160

65. Umeo H, Uchino H (2007) A new time-optimum synchronization algorithm for two-dimensional cellular arrays. In: Quesada-Arenciba A, Rodriguez JC, Moreno-Diaz R Jr, R Moreno-Diaz (eds) Proc. of Intern. Conf. on Computer Aided Systems Theory, EUROCAST 2007. LNCS, vol 4739. Springer, Berlin, pp 604–611

66. Umeo H, Yamawaki T, Shimizu N, Uchino H (2007) Modeling and simulation of global synchronization processes for large-scale-of two-dimensional cellular arrays. In: Proc. of Intern. Conf. on Modeling and Simulation, AMS 2007, pp 139–144

67. Umeo H, Yanagihara T (2007) A smallest five-state solution to the firing squad synchronization problem. In: Proc. of 5th Intern. Conf. on Machines, Computations, and Universality, MCU 2007. LNCS, vol 4664. Springer, Berlin, pp 292–302

68. Umeo H, Yanagihara T, Kanazawa M (2006) State-efficient firing squad synchronization protocols for communication-restricted cellular automata. In: Proc. of 7th International Conference on Cellular Automata for Research and Industry, ACRI2006. LNCS, vol 4173. Springer, Berlin, pp 169–181

69. Varshavsky VI, Marakhovsky VB, Peschansky VA (1970) Synchronization of Interacting Automata. Math Syst Theor 4(3):212–230

70. Vivien H (1996) A quasi-optimal time for synchronization two interacting finite automata. J Algebra Comput 6(2):261–267

71. Vivien H (2005) Cellular automata: a geometrical approach. Draft, pp 412

72. Vollmar R (1979) Algorithmen in Zellularautomaten, Teubner, Stuttgart, pp 192

73. Vollmar R (1982) Some remarks about the "Efficiency" of polyautomata. Int J Theor Phys 21(12):1007–1015
74. Waksman A (1966) An optimum solution to the firing squad synchronization problem. Inf Control 9:66–78
75. Worsch T (2000) Linear time language recognition on cellular automata with restricted communication. In: Gonnet GH, Panario D, Viola A (eds) Proc. of LATIN 2000:Theoretical Informatics. LNCS, vol 1776. Springer, Berlin, pp 417–426
76. Yunès JB (1994) Seven-state solution to the firing squad synchronization problem. Theor Comput Sci 127:313–332
77. Yunès JB (2006) Fault tolerant solutions to the firing squad synchronization problem in linear cellular automata. J Cell Autom 1:253–268
78. Yunès JB (2007) Simple new algorithms which solve the firing squad synchronization problem: A 7-states 4n-steps solution. In: Proc. of 5th Intern. Conf. on Machines, Computations, and Universality, MCU 2007. LNCS, vol 4664. Springer, Berlin, pp 316–324
79. Yunès JB (2008) An intrinsically non minimal-time Minsky-like 6-states solution to the firing squad synchronization problem. Theor Inf Appl 42(1):55–68
80. Yunès JB (2007) A 4-states algebraic solution to linear cellular automata synchronization. submitted to Information Processing Letters

Books and Reviews

Delorme M, Mozoyer J (eds) (1999) Cellular Automata. Kluwer Academic Publishers, Dordrecht
Ilachinski A (2001) Cellular Automata – A Discrete Universe –. World Scientific, Singapore
Wolfram S (1994) Cellular Automata and Complexity. Addison-Wesley Publishing Company, Reading
Wolfram S (2002) A New Kind of Science. Wolfram Media, Champaign

Fluctuations, Importance of: Complexity in the View of Stochastic Processes

RUDOLF FRIEDRICH[1], JOACHIM PEINKE[2], M. REZA RAHIMI TABAR[3]
[1] Institute for Theoretical Physics, University of Münster, Münster, Germany
[2] Institute of Physics, Carl-von-Ossietzky University Oldenburg, Oldenburg, Germany
[3] Dept. of Physics, Sharif University of Technology, Theran, Iran

Article Outline

Glossary

Complexity in time Complex structures may be characterized by non regular time behavior of a describing variable $q \in \mathbf{R}^d$. Thus the challenge is to understand or to model the time dependence of $q(t)$, which may be achieved by a differential equation $\frac{d}{dt} q(t) = \dots$ or by the discrete dynamics $q(t + \tau) = f(q(t), \dots)$ fixing the evolution in the future. Of special interest are not only nonlinear equations leading to chaotic dynamics but also those which include general noise terms, too.

Complexity in space Complex structures may be characterized by their spatial disorder. The disorder on a selected scale l may be measured at the location x by some scale dependent quantities, $q(l, x)$, like wavelets, increments and so on. The challenge is to understand or to model the features of the disorder variable $q(l, x)$ on different scales l. If the moments of q show power behavior $\langle q(l, x)^n \rangle \propto l^{\zeta(n)}$ the complex structures are called fractals. Well known examples of spatial complex structures are turbulence or financial market data. In the first case the complexity of velocity fluctuations over different distances l are investigated, in the second case the complexity of price changes over different time steps (time scale) are of interest.

Fokker–Planck equation The evolution of a variable $\mathbf{x}(t)$ from $\mathbf{x}'$ at t' to $\mathbf{x}$ at t, with $t' > t$, is described in a statistical manner by the conditional probability distribution $p(\mathbf{x}, t | \mathbf{x}', t')$. The conditional probability is subject to a Fokker–Planck equation, also known as second Kolmogorov equation, if

$$
\frac{\partial}{\partial t} p(\mathbf{x}, t | \mathbf{x}', t') =
$$
$$
- \sum_{i=1}^{d} \frac{\partial}{\partial x_i} D_i^{(1)}(\mathbf{x}, t) p(\mathbf{x}, t | \mathbf{x}', t')
$$
$$
+ \frac{1}{2} \sum_{i,j=1}^{d} \frac{\partial^2}{\partial x_i \partial x_j} D_{ij}^{(2)}(\mathbf{x}, t) p(\mathbf{x}, t | \mathbf{x}', t') ,
$$

holds. Here $\mathbf{D}^{(1)}$ and $\mathbf{D}^{(2)}$ are the drift vector and the diffusion matrix, respectively.

Kramers–Moyal coefficients Knowing for a stochastic process the conditional probability distribution $p(\mathbf{x(t)}, t|\mathbf{x}', t')$, for all t and t' the Kamers–Moyal coefficients can be estimated as nth order moments of the conditional probability distribution. In this way also the drift and diffusion coefficient of the Fokker–Planck equation can be obtained form the empirically measured conditional probability distributions.

Langevin equation The time evolution of a variable $\mathbf{x}(t)$ is described by Langevin equation if for $\mathbf{x}(t)$ it holds:

$$\frac{\mathrm{d}}{\mathrm{d}t}\mathbf{x(t)} = \mathbf{D}^{(1)}(\mathbf{x}, \mathbf{t}) \cdot \tau + \sqrt{\mathbf{D}^{(2)}(\mathbf{x}, \mathbf{t})} \cdot \boldsymbol{\Gamma}(\mathbf{t}_i)\,.$$

Using Itô's interpretation the deterministic part of the differential equation is equal to the drift term, the noise amplitude is equal to the square root of the diffusion term of a corresponding Fokker–Planck equation. Note, for vanishing noise a purely deterministic dynamics is included in this description.

Stochastic process in scale For the description of complex system with spatial or scale disorder usually a measure of disorder on different scales $q(l, x)$ is used. A stochastic process in scale is now a description of the l evolution of $q(l, x)$ by means of stochastic equations. As a special case the single event $q(l, x)$ follows a Langevin equation, whereas the probability $p(q(l))$ follows a Fokker–Planck equation.

Definition of the Subject

Measurements of time signals of complex systems of the inanimate and the animate world like turbulent fluid motions, traffic flow or human brain activity yield fluctuating time series. In recent years, methods have been devised which allow for a detailed analysis of such data. In particular methods for parameter free estimations of the underlying stochastic equations have been proposed. The present article gives an overview on the achievements obtained so far for analyzing stochastic data and describes results obtained for a variety of complex systems ranging from electrical nonlinear circuits, fluid turbulence, to traffic flow and financial market data. The systems will be divided into two classes, namely systems with complexity in time and systems with complexity in scale.

Introduction

The central theme of the present article is exhibited in Fig. 1. Given a fluctuating, sequentially measured set of experimental data one can pose the question whether it is possible to determine underlying trends and to assess
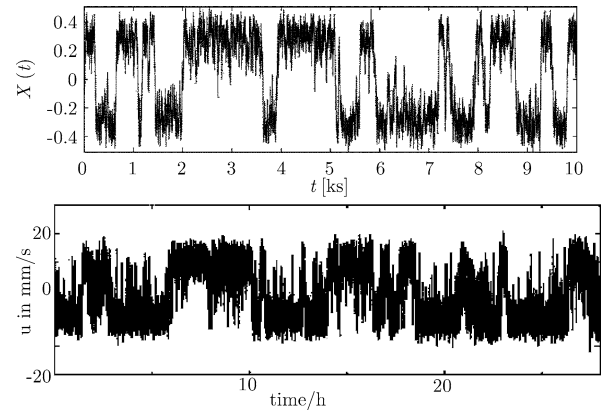
Fluctuations, Importance of: Complexity in the View of Stochastic Processes, Figure 1
Stochastic time series, data generated numerically and measured in a Rayleigh–Bénard experiment: Is it possible to disentangle turbulent trends from chances?

the characteristics of the fluctuations generating the experimental traces. This question becomes especially important for nonlinear systems, which can only partly be analyzed by the evaluation of the powerspectra obtained from a Fourier representation of the data.

In recent years it has become evident that for a wide class of stochastic processes the posed question can be answered in an affirmative way. A common method has been developed which can deal with stochastic processes (Langevin processes, Lévy processes) in time as well as in scale. In the first case, one faces the analysis of temporal disorder, whereas in the second case one considers scale disorder, which is an inherent feature of turbulent fluid motion and, quite interestingly, can also be detected in financial time series. This scale disorder is often linked to fractal scaling behavior and can be treated by a stochastic ansatz in a more general way.

In the present article we shall give an overview on the developed methods for analyzing stochastic data in time and scale. Furthermore, we list complex systems ranging from electrical nonlinear circuits, fluid turbulence, finance to biological data like heart beat data or data of human tremor, for which a successful application of the data analysis method has been performed. Furthermore, we shall focus on results obtained from some exemplary applications of the method to electronics, traffic flow, turbulence, and finance.

Complexity in Time

Complex systems are composed of a large number of subsystems behaving in a collective manner. In systems far

from equilibrium collectivity arises due to selforganization [1,2,3]. It results in the formation of temporal, spatial, spatio-temporal and functional structures.

The investigation of systems like lasers, hydrodynamic instabilities and chemical reactions has shown that selforganization can be described in terms of order parameters $u_i(t)$ which obey a set of stochastic differential equations of the form

$$\mathrm{d}u_i = N_i[u_1, .., u_n]\mathrm{d}t + \sum_j g_{ij}(u_1, .., u_n)\mathrm{d}W_j, \quad (1)$$

where W_j are independent Wiener processes. Although the state vector $\mathbf{q}(t)$ of the complex system under consideration is high dimensional, its long time behavior is entirely governed by the dynamics of typically few order parameters:

$$\mathbf{q}(t) = \mathbf{Q}(u_1, \ldots u_n). \quad (2)$$

This fact allows one to perform a macroscopic treatment of complex systems on the basis of the order parameter concept [1,2,3].

For hydrodynamic instabilities in the laminar flow regime like Rayleigh–Bénard convection or the Taylor–Couette experiment thermal fluctuations are usually small and can be neglected. However, in nonlinear optics and, especially, in biological systems the impact of noise has been shown to be of great importance. In principle, the order parameter equations (1) can be derived from basic equations characterizing the system under consideration close to instabilities [1,2]. However if the basic equations are not available, as is the case e. g. for systems considered in biology or medicine, the order parameter concept yields a top-down approach to complexity [3].

In this top-down approach the analysis of experimental time series becomes a central issue. Methods of nonlinear time series analysis (c.f. the monograph of Kantz and Schreiber [4]) have been widely applied to analyze complex systems. However, the developed methods aim at the understanding of deterministic systems and can only be successful if the stochastic forces are weak. Apparently, these methods have to be extended to include stochastic processes.

Complexity in Scale

In the case of selfsimilar structures complexity is commonly investigated by a local measure $q(l, x)$ characterizing the structure on the scale l at x. Selfsimilarity means that in a certain range of l the processes

$$q(l, x), \quad \lambda^\xi q(\lambda l, \lambda^\gamma x) \quad (3)$$

should have the same statistics. More precisely, the probability distribution of the quantity q takes the form

$$f(q, l) = \frac{1}{l^\xi} F\left(\frac{q}{l^\xi}\right) \quad (4)$$

with a universal function $F(Q)$. Furthermore, the moments exhibt scaling behavior

$$\langle q^n(l)\rangle = \int \mathrm{d}q\, q^n \frac{1}{l^\xi} F\left(\frac{q}{l^\xi}\right) = Q_n l^{n\xi} \quad (5)$$

Such type of behavior has been termed *fractal* scaling behavior.

There are many experimental examples of systems, like turbulent fields or surface roughness, just to mention two, that such a simple picture of a complex structure is only a rough, first approximation. In fact, especially for turbulence where $q(l, x)$ is taken as a velocity increment, it has been argued that *multifractal behavior* is more appropriate, where the nth order moments scale according to

$$\langle q^n(l)\rangle = Q_n l^{\zeta(n)}, \quad (6)$$

where the scaling indices $\zeta(n)$ are nonlinear functions of the order n:

$$\zeta(n) = n\xi_0 + n^2\xi_1 + n^3\xi_2 + \cdots. \quad (7)$$

Such a behavior can formally be obtained by the assumption that the probability distribution $f(q, l)$ has the following form

$$f(q, l) = \int \mathrm{d}\xi\, p(\xi, l) \frac{1}{l^\xi} F\left(\frac{q}{l^\xi}\right). \quad (8)$$

This formula is based on the assumption that in a turbulent flow regions with different scaling indices ξ exist, where $p(\xi, l)$ gives a measure of the scaling indices ξ at scale l. The major shortcoming of the fractal and multifractal approach to complexity in scale is the fact that it only addresses the statistics of the measure $q(l, x)$ at a single scale l. In general one has to expect dependencies of the measures $q(l, x)$ and $q(l', x)$ from different scales. Thus the question, which we will address in the following, can be posed, are there methods, which lead to a more comprehensive characterization of the scale disorder by general joint statistics

$$f(q_N, l_N; q_{N-1}, l_{N-1}; \ldots; q_1, l_1; q_0, l_0) \quad (9)$$

of the local measure q at multiple scales l_i.

Stochastic Data Analysis

Processes in time and scale can be analyzed in a similar way, if we generalize the time process given in (1) to a stochastic process equation evolving in scale

$$q(l + dl) = q(l) + N(q, l)dl + g(q, l)dW(l) \quad (10)$$

where dW belongs to a random process. The aim of data analysis is to disentangle deterministic dynamics and the impact of fluctuations. Loosely speaking, this amounts to detect *trends* and *chances* in data series of complex systems.

A complete analysis of experimental data, which is generated by the interplay of deterministic dynamics and dynamical noise, has to address the following issues

- Identification of the order parameters
- Extracting the deterministic dynamics
- Evaluating the properties of fluctuations

The outline of the present article is as follows. First, we shall summarize the description of stochastic processes focusing mainly on Markovian processes. Second, we discuss the approach developed to analyze stochastic processes. The last parts are devoted to applications of the data analysis method to processes in time and processes in scale.

Stochastic Processes

In the following we consider the class of systems which are described by a multivariate state vector $\mathbf{X}(t)$ contained in a d-dimensional state space $\{\mathbf{x}\}$. The evolution of the state vector $\mathbf{X}(t)$ is assumed to be governed by a deterministic part and by noise:

$$\frac{d}{dt}\mathbf{X}(t) = \mathbf{N}(\mathbf{X}(t), t) + \mathbf{F}(\mathbf{X}(t), t) . \quad (11)$$

$\mathbf{N}$ denotes a nonlinear function depending on the stochastic variable $\mathbf{X}(t)$ and additionally, may explicitly depend on time t (Note, time t can also be considered as a general variable and replaced for example by a scale variable l like in (86)). Because the function $\mathbf{N}$ can be nonlinear, also systems exhibiting chaotic time evolution in the deterministic case are included in the class of stochastic processes (11).

The second part, $\mathbf{F}(\mathbf{X}(t), t)$, fluctuates on a fast time scale. We assume that the d components F_i can be represented in the form

$$F_i(\mathbf{X}(t), t) = \sum_{j=1}^{d} g_{ij}(\mathbf{X}(t), t)\Gamma_j(t) . \quad (12)$$

The quantities $\Gamma_j(t)$ are considered to be random functions, whose statistical characteristics are well-defined. It is evident that these properties significantly determine the dynamical behavior of the state vector $\mathbf{X}(t)$. Formally, our approach also includes purely deterministic processes taking $F = 0$.

Discrete Time Evolution

It is convenient to consider the temporal evolution (11) of the state vector $\mathbf{X}(t)$ on a time scale, which is large compared to the time scale of the fluctuations $\Gamma_j(t)$.

As we shall briefly indicate below, a stochastic process related to the evolution Equation (11) can be modeled by stochastic evolution laws relating the state vectors $\mathbf{X}(t)$ at times $t_i, t_{i+1} = t_i + \tau, t_{i+2} = t_i + 2\tau, \ldots$ for small but finite values of τ.

In the present article we shall deal with the class of proper Langevin processes and generalized Langevin processes, which are defined by the following discrete time evolutions.

a) Proper Langevin Equations: White Noise The discrete time evolution of a proper Langevin process is given by

$$\mathbf{X}(t_{i+1}) = \mathbf{X}(t_i) + \mathbf{N}(\mathbf{X}(t_i), t_i) \cdot \tau$$
$$+ g(\mathbf{X}(t_i), t_i) \cdot \sqrt{\tau}\boldsymbol{\eta}(t_i) \quad (13)$$

where the stochastic increment $\boldsymbol{\eta}(t_i)$ is a fluctuating quantity characterized by a Gaussian distribution with zero mean, $\langle \boldsymbol{\eta}(t_i) \rangle = 0$

$$h(\boldsymbol{\eta}) = \frac{1}{(\sqrt{2\pi})^d}e^{-\frac{\eta^2}{2}} = \frac{1}{(\sqrt{2\pi})^d}e^{-\sum_{\alpha=1}^{d}\frac{\eta_\alpha^2}{2}} \quad (14)$$

Furthermore, the increments are statistically independent for different times

$$\langle \boldsymbol{\eta}(t_i)\boldsymbol{\eta}(t_j) \rangle = \delta_{ij} \quad (15)$$

b) Generalized Langevin Equations: Lévy Noise
A more general class is formed by the discrete time evolution laws [5,6],

$$\mathbf{X}(t_{i+1}) = \mathbf{X}(t_i) + \mathbf{N}(\mathbf{X}(t_i), t_i) \cdot \tau$$
$$+ g(\mathbf{X}(t_i), t_i) \cdot \tau^{1/\alpha}\boldsymbol{\eta}_{\alpha,\beta}(t_i) \quad (16)$$

where the increment $\boldsymbol{\eta}_{\alpha,\beta}(t_i)$ is a fluctuating quantity distributed according to the Lévy stable law characterized by the Lévy parameters α, β, [7].

As is well-known, only the Fourier-transform of this distribution can be given:

$$h_{\alpha,\beta} = \frac{1}{2\pi} \int dk\, Z(k, \alpha, \beta)\, e^{-ikx}$$

$$Z(k, \alpha, \beta) = e^{-i|k|^{\alpha}(1 - i\beta\, \text{sign}(k)\Phi)}$$

$$\Phi = \tan \pi \frac{\alpha}{2} \quad \alpha \neq 1\,, \tag{17}$$

$$\Phi = -\frac{2}{\pi} \ln|k| \quad \alpha = 1\,.$$

The Gaussian distribution is contained in the class of Lévy stable distributions ($\alpha = 2$, $\beta = 0$). Formally, α can be taken from the interval $0 < \alpha \leq 2$. However, for applications it seems reasonable to choose $1 < \alpha \leq 2$ in order that the first moment of the noise source exists.

The consideration of this type of statistics for the noise variables η is based on the central limit theorem, as discussed in a subsection below.

The discrete Langevin (13) and generalized Langevin Equations (16) have to be considered in the limit $\tau \to 0$. They are the basis of all further treatments. A central point is that if one assumes the noise sources to be independent of the variable $\mathbf{X}(t_i)$ the discrete time evolution equations define a Markov process, whose generator, i. e. the conditional probability distribution or short time propagator can be established on the basis of (13), (16).

In the following we shall discuss, how the discrete time processes can be related to the stochastic evolution equation (11).

Discrete Time Approximation of Stochastic Evolution Equations In order to motivate the discrete time approximations (13), (16) we integrate the evolution law (11) over a finite but small time increment τ:

$$\mathbf{X}(t + \tau) = \mathbf{X}(t) + \int_t^{t+\tau} dt'\, \mathbf{N}(\mathbf{X}(t'), t')$$

$$+ \int_t^{t+\tau} dt'\, g(\mathbf{X}(t'), t')\boldsymbol{\Gamma}(t')$$

$$\approx \mathbf{X}(t) + \tau \mathbf{N}(\mathbf{X}(t), t) \tag{18}$$

$$+ \int_t^{t+\tau} dt'\, g(\mathbf{X}(t'), t')\boldsymbol{\Gamma}(t')\,.$$

The time interval τ is chosen to be larger than the time scale of the fluctuations of $\Gamma_j(t)$. It involves the rapidly fluctuating quantities $\Gamma_j(t)$ and is denoted as a stochastic integral [8,9,10,11].

If we assume the matrix g to be independent on time t and state vector $\mathbf{X}(t)$, we arrive at the integrals

$$d\mathbf{W}(t, \tau) = \int_t^{t+\tau} dt'\, \boldsymbol{\Gamma}(t')\,. \tag{19}$$

These are the quantities, for which a statistical characterization can be given. We shall pursue this problem in the next subsection.

However, looking at (18) we encounter the difficulty that the integrals over the noise forces may involve functions of the state vector within the time interval $(t, t + \tau)$. The interpretation of such integrals for wildly fluctuating, stochastic quantities $\boldsymbol{\Gamma}(t)$ is difficult. The simplest way is to formulate an interpretation of these terms leading to different interpretations of the stochastic evolution equation (11). We formulate the widely used definitions due to Itô and Stratonovich.

In the *Itô* sense, the integral is interpreted as

$$\int_t^{t+\tau} dt'\, g(\mathbf{X}(t'), t')\boldsymbol{\Gamma}(t') = g(\mathbf{X}(t), t) \int_t^{t+\tau} dt'\, \boldsymbol{\Gamma}(t')$$

$$= g(\mathbf{X}(t), t)\, d\mathbf{W}(t, \tau)\,. \tag{20}$$

The *Stratonovich* definition is

$$\int_t^{t+\tau} dt'\, g(\mathbf{X}(t'), t')\Gamma(t')$$

$$= g\left(\frac{\mathbf{X}(t + \tau) + \mathbf{X}(t)}{2}, t + \frac{\tau}{2}\right) \int_t^{t+\tau} dt'\, \boldsymbol{\Gamma}(t') \tag{21}$$

$$= g\left(\frac{\mathbf{X}(t + \tau) + \mathbf{X}(t)}{2}, t + \frac{\tau}{2}\right) d\mathbf{W}(t, \tau)\,.$$

Since from experiments one obtains probability distributions of stochastic processes which are related to stochastic Langevin equations, we are free to choose a certain interpretation of the process. In the following we shall always adopt the Itô interpretation. In this case, the drift vector $\mathbf{D}^1(\mathbf{x}, t) = \mathbf{N}(\mathbf{x}, t)$ coincides with the nonlinear vector field $\mathbf{N}(\mathbf{x}, t)$.

Limit Theorems, Wiener Process, Lévy Process In the following we shall discuss possibilities to characterize the stochastic integrals

$$W(t + \tau) - W(t) = \int_t^{t+\tau} dt'\, \Gamma(t') \tag{22}$$

$\Gamma(t)$ is a rapidly fluctuating quantity of zero mean. In order to characterize the properties of this force one can resort to the limit theorems of statistical mechanics [7].

The central limit theorem states that if the quantities Γ_j, $j = 1, \ldots, n$ are statistically independent variables of zero mean and variance σ^2 then the sum

$$\frac{1}{\sqrt{n}} \sum_{j=1}^n \Gamma_j = \eta \tag{23}$$

tends to a Gaussian random variable with variance σ^2 for large values of n. The limiting probability distribution $h(\eta)$ is then a Gaussian distribution with variance σ^2:

$$h(\eta) = \frac{1}{\sqrt{2\pi\sigma^2}} e^{-\frac{\eta^2}{2\sigma^2}} . \tag{24}$$

As is well known, there is a generalization of the central limit theorem, which applies to random variables whose second moment does not exist. It states that the distribution of the sum over identically distributed random variables Γ_j

$$\frac{1}{n^{1/\alpha}} \sum_{j=1}^{n} \Gamma_j = \eta_{\alpha,\beta} \tag{25}$$

tends to a random variable $\eta_{\alpha,\beta}$, which is distributed according to the Lévy-stable distribution $h_{\alpha,\beta}(\eta)$. The Lévy stable distributions can only be given by their Fourier – transforms, cf. Eq. (17).

In order to evaluate the integral (22) using the limit theorems, it is convenient to represent the stochastic force $\Gamma(t)$ as a sum over N_τ δ-kicks occuring at discrete times t_j

$$\Gamma(t) = \sum_j \Gamma_j (\Delta t)^{1/\alpha} \delta(t - t_j) . \tag{26}$$

Thereby, Δt is the time difference between the occurrence of two kicks. Then, we obtain

$$\begin{aligned} dW(t, \tau) &= \sum_{j, t_j \in \tau} \Gamma_j (\Delta t)^{1/\alpha} \\ &= (N_\tau \Delta t)^{1/\alpha} \frac{1}{N_\tau^{1/\alpha}} \sum_{j, t_j \in \tau} \Gamma_j = \tau^{1/\alpha} \eta(t) . \end{aligned} \tag{27}$$

An application of the central limit theorem shows that if the quantities Γ_j are identically distributed independet variables the integral

$$\frac{1}{\tau^{1/\alpha}} \int_t^{t+\tau} = \eta(t) \tag{28}$$

can be considered to be a random variable $\eta(t)$ which in the limit $N_\tau \to \infty$ tends to a stable random variable.

Thus, for the case $\alpha = 2$, i. e. for the case where the second moments of the random kicks exist, the stochastic variable $dW(t, \tau)$ can be represented by the increments

$$dW(t, \tau) = \sqrt{\tau} \eta(t) \tag{29}$$

where $\eta(t)$ is a Gaussian distributed random variable.

For the more general case, $dW(t, \tau)$ is a stochastic variable

$$dW(t, \tau) = \tau^{1/\alpha} \eta_{\alpha,\beta}(t) \tag{30}$$

where the distribution of $\eta_{\alpha,\beta}$ is the Lévy distribution (17).

Statistical Description of Stochastic Processes

In the previous subsection we have discussed processes described by stochastic equations. In the present subsection we shall summarize the corresponding statistical description. Such a description is achieved by introducing suitable statistical averages. We shall denote these averages by the brackets $\langle \ldots \rangle$. For stationary processes the averages can be viewed as a time average. For nonstationary processes averages are defined as ensemble averages, i. e. averages over an ensemble of experimental (or numerical) realizations of the stochastic process (11). For stationary processes in time, one usually deals with time averages. For processes in scale, the average is an ensemble average.

Probability Distributions The set of stochastic evolution equations (11), or it's finite time representations (13), (16) define a Markov process. We consider the joint probability density

$$f(\mathbf{x}_n, t_n; \ldots; \mathbf{x}_1, t_1; \mathbf{x}_0, t_0) \tag{31}$$

which is related to the joint probability, to find the system at times t_i in the volume ΔV_i in phase space. If we take times t_i which are separated by the small time increment $\tau = t_{i+1} - t_i$, then the probability density can be related to the discrete time representation of the stochastic process (13), (16) according to

$$\begin{aligned} f(\mathbf{x}_n, t_n; \ldots; \mathbf{x}_1, t_1; \mathbf{x}_0, t_0) \\ = \langle \delta(\mathbf{x}_n - \mathbf{X}(t_n)) \ldots \delta(\mathbf{x}_0 - \mathbf{X}(t_0)) \rangle , \end{aligned} \tag{32}$$

where the brackets indicate the statistical average, which may be a time average (for stationary processes) or an ensemble average.

Markov Processes An important subclass of stochastic processes are Markov processes. For these processes the joint probability distribution $f(\mathbf{x}_n, t_n; \ldots; \mathbf{x}_1, t_1; \mathbf{x}_0, t_0)$ can be constructed from the knowledge of the conditional probability distributions

$$p(\mathbf{x}_{i+1}, t_{i+1} | \mathbf{x}_i, t_i) = \frac{f(\mathbf{x}_{i+1}, t_{i+1}; \mathbf{x}_i, t_i)}{f(\mathbf{x}_i, t_i)} \tag{33}$$

according to

$$\begin{aligned} f(\mathbf{x}_n, t_n; \ldots; \mathbf{x}_1, t_1; \mathbf{x}_0, t_0) \\ = p(\mathbf{x}_n, t_n | \mathbf{x}_{n-1}, t_{n-1}) \ldots p(\mathbf{x}_1, t_1 | \mathbf{x}_0, t_0) f(\mathbf{x}_0, t_0) , \end{aligned} \tag{34}$$

Here the Markov property of a process for multiple conditioned probabilities

$$p(\mathbf{x}_i, t_i | \mathbf{x}_{i-1}, t_{i-1}; \ldots; \mathbf{x}_0, t_0) = p(\mathbf{x}_i, t_i | \mathbf{x}_{i-1}, t_{i-1}) \tag{35}$$

is used. As a consequence, the knowledge of the transition probabilities together with the initial probability distribution $f(\mathbf{x}_0, t_0)$ suffices to define the N-times probability distribution.

It is straightforward to prove the Chapman–Kolmogorov equation

$$p(\mathbf{x}_j, t_j \,|\, \mathbf{x}_i, t_i) = \int d\mathbf{x}_k\, p(\mathbf{x}_j, t_j \,|\, \mathbf{x}_k, t_k)\, p(\mathbf{x}_k, t_k \,|\, \mathbf{x}_i, t_i). \tag{36}$$

This relation is valid for all times $t_i < t_k < t_j$. In the following we shall show that the transition probabilities $p(\mathbf{x}_{j+1}, t + \tau \,|\, \mathbf{x}_j, t)$ can be determined for small time differences τ. This defines the so-called short time propagators.

Short Time Propagator of Langevin Processes It is straightforward to determine the short time propagator from the finite time approximation (13) of the Langevin equation. We shall denote these propagators by $p(\mathbf{x}_{j+1}, t + \tau \,|\, \mathbf{x}_j, t)$, in contrast to the finite time propagators (33), for which the time interval $t_{i+1} - t_i$ is large compared to τ.

We first consider the case of Gaussian noise. The variables $\boldsymbol{\eta}(t_i)$ are Gaussian random vectors with probability distribution

$$h[\boldsymbol{\eta}] = \frac{1}{\sqrt{(2\pi)^d}} \exp\left[-\frac{\boldsymbol{\eta} \cdot \boldsymbol{\eta}}{2}\right]. \tag{37}$$

The finite time intepretation of the Langevin equation can be rewritten in the form

$$\boldsymbol{\eta}(t_i) = \frac{1}{\tau^{1/2}}[g(\mathbf{X}(t_i), t_i)]^{-1}[\mathbf{X}(t_{i+1}) - \mathbf{X}(t_i) - \tau\mathbf{N}(\mathbf{X}(t_i))]. \tag{38}$$

This relation, in turn, defines the transition probability distribution

$$p(\mathbf{x}_{i+1}, t_{i+1} \,|\, \mathbf{x}_i, t_i)\,d\mathbf{x}_{i+1} = h[\boldsymbol{\eta} = \boldsymbol{\eta}(t_i)]J(\mathbf{x}_i, t_i)\,d\mathbf{x}_{i+1}, \tag{39}$$

where J is the determinant of the Jacobian

$$J_{\alpha\beta} = \frac{\partial \eta_\alpha(t_i)}{\partial \mathbf{x}_{i+1,\beta}}, \tag{40}$$

and $[g]^{-1}$ denotes the inverse of the matrix g (which is assumed to exist).

For the following it will be convenient to define the so-called diffusion matrix $D^{(2)}(\mathbf{x}_i, t_i)$

$$D^{(2)}(\mathbf{x}_i, t_i) = g^T(\mathbf{x}_i, t_i)\, g(\mathbf{x}_i, t_i) \tag{41}$$

We are now able to explicitly state the short time propagator of the process (13):

$$p(\mathbf{x}_i + \tau, t_{i+1} \,|\, \mathbf{x}_i, t_i) = \frac{1}{\sqrt{(2\pi\tau)^d \, \mathrm{Det}[D^{(2)}]}} e^{-S(\mathbf{x}_{i+1}, \mathbf{x}_i, t_i, \tau)} \tag{42}$$

We have defined the quantity $S(\mathbf{x}_{i+1}, \mathbf{x}_i, t_i, \tau)$ according to

$$S(\mathbf{x}_{i+1}, \mathbf{x}_i, t_i, \tau) = \tau \left[\frac{\mathbf{x}_{i+1} - \mathbf{x}_i}{\tau} - \mathbf{D}^{(1)}(\mathbf{x}_i, t)\right]$$
$$\cdot [D^{(2)}(\mathbf{x}_i, t_i)]^{-1}\left[\frac{\mathbf{x}_{i+1} - \mathbf{x}_i}{\tau} - \mathbf{D}^{(1)}(\mathbf{x}_i, t_i)\right]. \tag{43}$$

As we see, the short time propagator, which yields transition probability density from state $\mathbf{x}_i$ to state $\mathbf{x}_{i+1}$ in the finite but small time interval τ is a normal distribution.

Short Time Propagator of Lévy Processes It is now straightforward to determine the short time propagator for Lévy processes. We have to replace the Gaussian distribution by the (multivariate) Lévy distribution $h_{\alpha,\beta}(\boldsymbol{\eta})$. As a consequence, we obtain the conditional probability, i. e. the short time propagator, for Lévy processes:

$$p(\mathbf{x}_{i+1}, t_i + \tau \,|\, \mathbf{x}_i, t_i) = \frac{1}{\mathrm{Det}[g(\mathbf{x}_i, t_i)]} h_{\alpha,\beta}$$
$$\cdot \left\{\frac{1}{\tau^{1/\alpha}}[g(\mathbf{x}(t_i), t_i)]^{-1}[\mathbf{x}(t_{i+1}) - \mathbf{x}(t_i) - \tau\mathbf{N}(\mathbf{x}(t_i))]\right\} \tag{44}$$

Joint Probability Distribution and Markovian Properties Due to statistical independence of the random vectors $\boldsymbol{\eta}(t_i)$, $\boldsymbol{\eta}(t_j)$ for $i \neq j$ we obtain the joint probability distribution as a product of the distributions $h(\boldsymbol{\eta})$:

$$h(\boldsymbol{\eta}_N, \ldots, \boldsymbol{\eta}_0) = h(\boldsymbol{\eta}_N)\, h(\boldsymbol{\eta}_{N-1}) \ldots h(\boldsymbol{\eta}_0) \tag{45}$$

Furthermore, we observe that under the assumption that the random vector $\boldsymbol{\eta}(t_i)$ is independent on the variables $\mathbf{X}(t_j)$ for all $j \leq i$ we can construct the N-time probability distribution

$$f(\mathbf{x}_n, t_n; \ldots; \mathbf{x}_1, t_1; \mathbf{x}_0, t_0)$$
$$= p(\mathbf{x}_n, t_n \,|\, \mathbf{x}_{n-1}, t_{n-1}) \ldots p(\mathbf{x}_1, t_1 \,|\, \mathbf{x}_0, t_0)f(\mathbf{x}_0, t_0) \tag{46}$$

However, this is the definition of a Markov chain. Thereby, the transition probabilities are the short time propagators, i. e. the representation (46) is valid in the short time limit $\tau \to 0$. The probability distribution (46) is the discrete approximation of the *path integral representation* of the stochastic process under consideration [8].

Let us summarize: The statistical description of the Langevin equation based on the n-times joint probability distribution leads to the representation in terms of the conditional probability distribution. This representation is the definition of a Markov process.

Due to the assumptions on the statistics of the fluctuating forces different processes arise. If the fluctuating forces are assumed to be Gaussian the short time propagator is Gaussian and, as a consequence, solely defined by the drift vector and the diffusion matrix.

If the fluctuating forces are assumed to be Lévy distributed, more complicated short time propagators arise.

Let us add the following remarks:

a) The assumption of Gaussianity of the statistics is not necessary. One can consider fluctuating forces with non-Gaussian probability distributions. In this case the probability distributions have to be characterized by higher order moments, or, more explicitly, by its cumulants. At this point we remind the reader that for non-Gaussian distributions, infinitely many cumulants exist.

b) The Markovian property, i. e. the fact that the propagator $p(\mathbf{x}_i, t_i | \mathbf{x}_{i-1} t_{i-1})$ does not depend on states $\mathbf{x}_k$ at times $t_k < t_{i-1}$ is usually violated for physical systems due to the fact that the noise sources become correlated for small time differences T_{mar}. This point already has been emphasized in the famous work of A. Einstein on Brownian motion, which is one of the first works on stochastic processes [12]. This time scale is denoted as Markov–Einstein scale. It seems to be a highly interesting quantity especially for nonequilibrium systems like turbulence [13,14] and earthquake signals [15].

Finite Time Propagators

Up to now, we have considered the short time propagators $p(\mathbf{x}_i, t_i | \mathbf{x}_{i-1}, t_{i-1})$ for inifinitesimal time differences $t_i - t_{i-1} = \tau$. However, one is interested in the conditional probability distributions for finite time intervals, $p(\mathbf{x}, t | \mathbf{x}', t'), t - t' \gg \tau$.

Fokker–Planck Equation The conditional probability distribution $p(\mathbf{x}, t | \mathbf{x}', t'), t - t' \gg \tau$ can be obtained from the solution of the Fokker–Planck equation (also known as second Kolmogorov equation [16]):

$$\frac{\partial}{\partial t} p(\mathbf{x}, t \mid \mathbf{x}', t') = -\sum_{i=1}^{d} \frac{\partial}{\partial x_i} D_i^{(1)}(\mathbf{x}, t)\, p(\mathbf{x}, t \mid \mathbf{x}', t')$$
$$+ \frac{1}{2} \sum_{i,j=1}^{d} \frac{\partial^2}{\partial x_i \partial x_j} D_{ij}^{(2)}(\mathbf{x}, t)\, p(\mathbf{x}, t \mid \mathbf{x}', t') , \quad (47)$$

$\mathbf{D}^{(1)}$ and $\mathbf{D}^{(2)}$ are drift vector and diffusion matrix.

Under consideration of Itô's definitions of stochastic integrals the coefficients $\mathbf{D}^{(1)}$, $\mathbf{D}^{(2)}$ of the Fokker–Planck equation (47) and the functionals $\mathbf{N}$, $\mathbf{g}$ of the Langevin equation (11), (12) are related by

$$D_i^{(1)}(\mathbf{x}, t) = N_i(\mathbf{x}, t) , \quad (48)$$

$$D_{ij}^{(2)}(\mathbf{x}, t) = \sum_{l=1}^{d} g_{il}(\mathbf{x}, t)\, g_{jl}(\mathbf{x}, t) . \quad (49)$$

They are defined according to

$$D_i^{(1)}(\mathbf{x}, t) = \lim_{\tau \to 0} \frac{1}{\tau} \langle X_i(t + \tau) - x_i \rangle|_{\mathbf{X}(t) = \mathbf{x}}$$
$$= \lim_{\tau \to 0} \frac{1}{\tau} \int d\mathbf{x}'\, p(\mathbf{x}', t + \tau \mid \mathbf{x}, t)(x_i' - x_i) \quad (50)$$

$$D_{ij}^{(2)}(\mathbf{x}, t)$$
$$= \lim_{\tau \to 0} \frac{1}{\tau} \langle (X_i(t + \tau) - x_i)(X_j(t + \tau) - x_j) \rangle|_{\mathbf{X}(t) = \mathbf{x}}$$
$$= \lim_{\tau \to 0} \frac{1}{\tau} \int d\mathbf{x}'\, p(\mathbf{x}', t + \tau | \mathbf{x}, t)(x_i' - x_i)(x_j' - x_j) . \quad (51)$$

These expressions demonstrate that drift vector and diffusion matrix can be determined as the first and second moments of the conditional probability distributions $p(\mathbf{x}', t + \tau | \mathbf{x}, t)$ in the small time limit.

Fractional Fokker–Planck Equations The finite time propagators or conditional probability distributions of stochastic processes containing Lévy-noise lead to fractional diffusion equations. For a discussion of this topic we refer the reader to [5,6].

Master Equation The most general equation specifying a Markov process for a continuous state vector $\mathbf{X}(t)$ takes the form

$$\frac{\partial}{\partial t} p(\mathbf{x}, t \mid \mathbf{x}_0, t_0) = \int d\mathbf{x}'\, w(\mathbf{x}, \mathbf{x}', t)\, p(\mathbf{x}', t \mid \mathbf{x}_0, t_0)$$
$$- \int d\mathbf{x}'\, w(\mathbf{x}', \mathbf{x}, t)\, p(\mathbf{x}, t \mid \mathbf{x}_0, t_0) \quad (52)$$

here w denote transition probabilities.

Measurement Noise

We can now go a step ahead and include measurement noise. Due to measurement noise, the observed state vector, which we shall now denote by $\mathbf{Y}(t_i)$, is related to the stochastic variable $\mathbf{X}(t_i)$ by an additional noise term $\epsilon(t_i)$:

$$\mathbf{Y}(t_i) = \mathbf{X}(t_i) + \epsilon(t_i) \quad (53)$$

We assume that the stochastic variables $\epsilon(t_i)$ have zero mean, are statistically independent and obey the probability density $h(\epsilon)$. Then the probability distribution for the variable $\mathbf{Y}_{t_i}$ is given by

$$g(\mathbf{y}_n, t_n; \ldots; \mathbf{y}_0, t_0) = \int \ldots \int d\epsilon_n \ldots d\epsilon_0 \quad (54)$$
$$\cdot f(\mathbf{y}_n - \epsilon_n, t_n; \ldots; \mathbf{y}_0 - \epsilon_0, t_0) \, h(\epsilon_n) \ldots h(\epsilon_0) \, .$$

The short time propagator recently has been determined for the Ornstein–Uhlenbeck process, a process with linear drift term and constant diffusion. Analysis of data sets spoilt by measurement noise is currently under investigation [17,18].

Stochastic Time Series Analysis

The ultimate goal of nonlinear time series analysis applied to deterministic systems is to extract the underlying nonlinear dynamical system directly from measured time series in the form of a system of differential equations, cf. [4]. The role played by dynamic fluctuations has not been fully appreciated. Mostly, noise has been considered as a random variable additively superimposed on a trajectory generated by a deterministic dynamical system. Noise has been usually considered as extrinsic or measurement noise. The problem of dynamical noise, i. e. fluctuations which interfere with the deterministic dynamical evolution, has not been addressed in full details.

The natural extension of the nonlinear time series analysis to (continuous) Markov processes is the estimation of *short time propagators* from time series. During recent years, it has become evident that such an approach is feasible. In fact, noise may help in the estimation of the deterministic ingredients of the dynamics. Due to dynamical noise the system explores a larger part of phase space and thus measurements of signals yield considerably more information about the dynamics as compared to the purely deterministic case, where the trajectories fastly converge to attractors providing only limited information.

The analysis of data set's of stochastic systems exhibiting Markov properties has to be performed along the following lines:

- Disentangling Measurement and Dynamical Noise
- Evaluating Markovian Properties
- Determination of Short Time Propagators
- Reconstruction of Data

Since the methods for disentangling measurement and dynamical noise are currently under intense investigation, see Subsect. "Measurement Noise", our focus is on the three remaining issues.

Evaluating Markovian Properties

In principle it is a difficult task to decide on Markovian properties by an inspection of experimental data. The main point is that Markovian properties usually are violated for small time increments τ, as it already has been pointed out above and in [12]. There are at least two reasons for this fact.

First, the dynamical noise sources become correlated at small time differences. If we consider Gaussian noise sources one usually observes an exponential decay of correlations

$$\langle \Gamma_i(t)\Gamma_j(t') \rangle = \delta_{ij} \frac{\mathrm{e}^{-|t-t'|/T_{\mathrm{mar}}}}{T_{\mathrm{mar}}} \quad (55)$$

Markovian properties can only expected to hold for time increments $\tau > T_{\mathrm{mar}}$.

Second, measurement noise can spoil Markovian properties [19].

Thus, the estimation of the Markovian time scale T_{mar} is a necessary step for stochastic data analysis. Several methods have been proposed to test Markov properties.

Direct Evaluation of Markovian Properties   A direct way is to use the definition of a Markov process (35) and to consider the higher order conditional probability distributions

$$p(\mathbf{x}_3, t_3 \mid \mathbf{x}_2 t_2; \mathbf{x}_1, t_1) = \frac{f(\mathbf{x}_3, t_3; \mathbf{x}_2 t_2; \mathbf{x}_1, t_1)}{f(\mathbf{x}_2 t_2; \mathbf{x}_1, t_1)}$$
$$= p(\mathbf{x}_3, t_3 \mid \mathbf{x}_2 t_2) \quad (56)$$

This procedure is feasible if large data sets are available. Due to the different conditioning both probabilities are typically based on different number of events. As an appropriate method to statistically show the similarity of (56) the Wilcoxon Test has been proposed, for details see [20,21].

In principle, higher order conditional probability distributions should be considered in a similar way. However, the validity of relation (56) is a strong hint for Markovianity.

Evaluation of Chapman–Kolmogorov Equation   An indirect way is to use the Chapman–Kolmogorov equation (36), whose validity is a necessary condition for Markovianity. The method is based on a comparison between the conditional pdf

$$p(\mathbf{x}_k, t_k \mid \mathbf{x}_i, t_i) \quad (57)$$

taken from experiment and the one calculated by the Chapman–Kolmogorov equation

$$\tilde{p}(\mathbf{x}_k, t_k | \mathbf{x}_i, t_i) = \int d\mathbf{x}_j \, p(\mathbf{x}_k, t_k | \mathbf{x}_j, t_j) \, p(\mathbf{x}_j, t_j | \mathbf{x}_i, t_i)$$

(58)

where t_j is an intermediate time $t_i < t_j < t_k$. A refined method can be based on an iteration of the Chapman–Kolmogorov equation, i. e. considering several intermediate times. If the Chapman–Kolmogorov equation is not fulfilled, deviations are enhanced by each iteration.

Direct Estimation of Stochastic Forces Probably the most direct way is the determination of the stochastic forces from data. If the drift vector field $\mathbf{D}^{(1)}(\mathbf{x}, t)$ has been established, as discussed below, the fluctuating forces can be estimated according to

$$g(\mathbf{x}, t)\boldsymbol{\Gamma}(t) = \frac{d\mathbf{x} - \tau \mathbf{D}^{(1)}(\mathbf{x}, t)}{\sqrt{\tau}} \ .$$

(59)

The correlations of this force can then be examined directly, see also [22] and Subsect. "Noisy Circuits".

Differentiating Between Stochastic Process and Noise Data Looking at the joint statistics of increments extracted from given data, it could be shown that the nesting of increments and the resulting statistics can be used to differentiate between noise – like data sets and those resulting from stochastic time processes [24].

Estimating the Short Time Propagator

A crucial step in the stochastic analysis is the assessment of the short time propagator of the continuous Markov process. This gives access to the deterministic part of the dynamics as well as to the fluctuations.

Gaussian Noise We shall first consider the case of Gaussian white noise. As we have already indicated, drift vector and diffusion matrix are defined as the first and second moments of the conditional probability distribution or short time propagator, Eq. (50).

We shall now describe an operational approach, which allows one to estimate drift vector and diffusion matrix from data and has been successfully applied to a variety of stochastic processes. We shall discuss the case, where averages are taken with respect to an ensemble of experimental realizations of the stochastic process under consideration in order to include nonstationary processes. Replacing the ensemble averages by time averages for statistically stationary processes is straightforward.

a

b
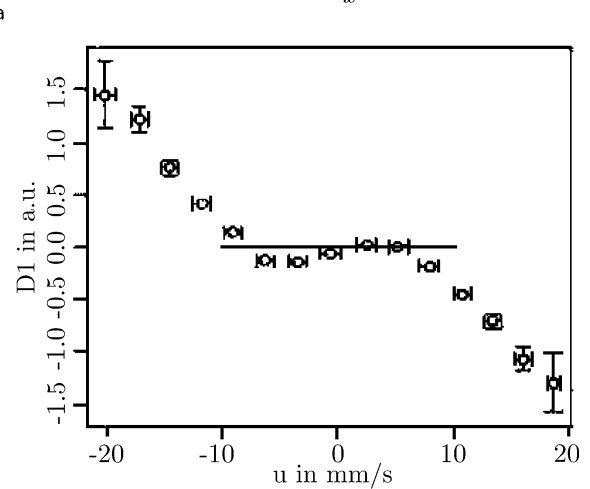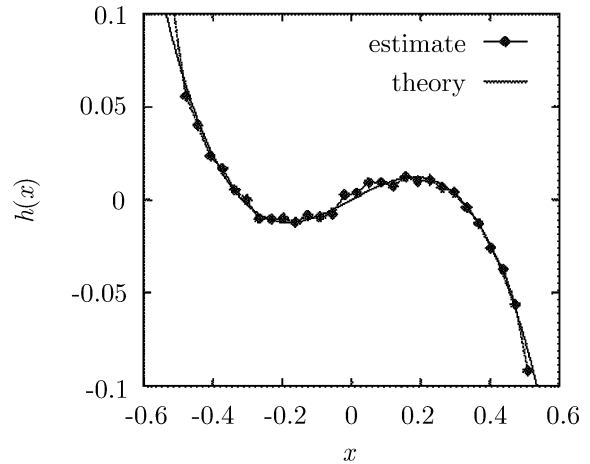
Fluctuations, Importance of: Complexity in the View of Stochastic Processes, Figure 2
Estimated drift terms from the numerical and experimental data of Fig. 1. In part **a** also the exact function of the numerical model is shown as solid curve, in part **b** the line $D^{(1)} = 0$ is shown to visualize the multiple fixed points; after [23]

For illustration we show in Fig. 2 the estimated functions of the drift terms obtained from the analysis of the data of Fig. 1.

The procedure is as follows:

- The data is represented in a d-dimensional phase space
- The phase space is partitioned into a set of finite, but small d-dimensional volume elements
- For each bin (denoted by α), located at the point $\mathbf{x}_\alpha$ of the partition we consider the quantity

$$\mathbf{x}(t_j + \tau) = \mathbf{x}(t_j) + \mathbf{D}^{(1)}(\mathbf{x}(t_j), t_j)\tau + \sqrt{\tau}g(\mathbf{x}(t_j)\boldsymbol{\Gamma}(t_j)$$

(60)

Thereby, the considered points $\mathbf{x}(t_i)$ are taken from the bin located at $\mathbf{x}_\alpha$. Since we consider time dependent processes this has to be done for each time step t_j separately

- **Estimation of the drift vector:**
  The drift vector assigned to the bin located at $\mathbf{x}_\alpha$ is determined as the small τ- limit

  $$\mathbf{D}^{(1)}(\mathbf{x}, t) = \lim_{\tau \to 0} \frac{1}{\tau} \mathbf{M}(\mathbf{x}, t, \tau) \tag{61}$$

  of the conditional moment

  $$\mathbf{M}^{(1)}(\mathbf{x}_\alpha, t_j, \tau) = \frac{1}{N_\alpha} \sum_{\mathbf{x}(t_j) \in \alpha} [\mathbf{x}(t_j + \tau) - \mathbf{x}(t_j)] \tag{62}$$

  The sum is over all N_α points contained in the bin α.
  Proof: The drift vector assigned to the bin α located at $\mathbf{x}_\alpha$ is approximated by the conditional expectation value

  $$\mathbf{M}^{(1)}(\mathbf{x}_\alpha, t, \tau) = \tau \frac{1}{N_\alpha} \sum_{\mathbf{x}(t_j) \in \alpha} \mathbf{D}^{(1)}(\mathbf{x}_j, t_j)$$
  $$+ \sqrt{\tau} \frac{1}{N_\alpha} \sum_{\mathbf{x}(t_j) \in \alpha} g(\mathbf{x}(t_j) \boldsymbol{\Gamma}(t_j) \tag{63}$$

  Thereby, the sum is over all points $\mathbf{x}(t_j)$ located in the bin assigned to $\mathbf{x}_\alpha$. Assuming that $\mathbf{D}^{(1)}(\mathbf{x}, t)$ and $g(\mathbf{x}, t)$ do not vary significantly over the bin, the second contribution drops out since

  $$\frac{1}{N_\alpha} \sum_{\mathbf{x}_j \in \alpha} \boldsymbol{\Gamma}(t_j) \to 0 \tag{64}$$

  $\square$

- **Estimation of the Diffusion matrix**
  The diffusion matrix can be estimated by the small τ-limit

  $$\mathbf{D}^{(2)}(\mathbf{x}, t) = \lim_{\tau \to 0} \frac{1}{\tau} \mathbf{M}^{(2)}(\mathbf{x}, t, \tau) \tag{65}$$

  of the conditional second moment

  $$\mathbf{M}^{(2)}(\mathbf{x}_\alpha, t, \tau)$$
  $$= \frac{1}{N_\alpha} \sum_j \{[\mathbf{x}(t_j + \tau) - \mathbf{x}(t_j)] - \tau \mathbf{D}^{(1)}(\mathbf{x}_j, t_j)\}^2 \tag{66}$$

  Proof: We consider the quantity

  $$\mathbf{M}^{(2)}(\mathbf{x}_\alpha, t, \tau)$$
  $$= \tau \frac{1}{N_\alpha} \sum_{\mathbf{x}(t_k) \in \alpha} \sum_k g(\mathbf{x}(t_k), t_k) \boldsymbol{\Gamma}(t_k) g(\mathbf{x}(t_j) \boldsymbol{\Gamma}(t_j) \tag{67}$$

If the bin size is small comparable to the scale, where the matrix $g(\mathbf{x}, t)$ varies significantly, we can replace $g(\mathbf{x}(t_k), t_k)$ by $g(\mathbf{x}_\alpha, t_k)$ such that

$$\mathbf{M}^{(2)}(\mathbf{x}_\alpha, t, \tau)$$

$$= g(\mathbf{x}_\alpha, t_k) \left[\frac{1}{N_\alpha} \sum_{\mathbf{x}_j \in \alpha} \sum_{\mathbf{x}_k \in \alpha} \boldsymbol{\Gamma}(t_k) \boldsymbol{\Gamma}(t_j) \right] g^T(\mathbf{x}_\alpha, t_k)$$

$$= \tau g(\mathbf{x}_\alpha, t_k) g^T(\mathbf{x}_\alpha, t_k)$$

$$\tag{68}$$

Thereby, we have used the assumption of the statistical independence of the fluctuations

$$\frac{1}{N_\alpha} \sum_{\mathbf{x}(t_j) \in \alpha} \sum_{\mathbf{x}(t_k) \in \alpha} \boldsymbol{\Gamma}(t_k) \boldsymbol{\Gamma}(t_j) = \delta_{kj} E \tag{69}$$

$\square$

- **Higher order cumulants**
  In a similar way one may estimate higher cumulants M^n, which in the small time limit converge to the so-called Kramers–Moyal coefficients. The estimation of these quantities allows to answer the question whether the noise sources actually are Gaussian distributed random variables

Technical Aspects   The above procedure of estimating drift vector and diffusion matrix explicitly shows the properties, which limit the accuracy of the determined quantities.

First of all, the bin size influences the results. The bin size should allow for a resonable number of events such that the sums converge, however, it should be reasonable small in order to allow for a accurate resolution of drift vector and diffusion matrix.

Second, the data should allow for the estimation of the conditional moments in the limit $\lim_{\tau \to 0}$ [25,26]. Here, a finite Markov–Einstein coherence length may cause problems. Furthermore, measurement noise can spoil the possibility of performing this limit.

From the investigation of the Fokker–Planck equation much is known on the τ dependence of the conditional moments. This may be used for further improved estimations, as has been discussed in [27]. Furthermore, as we shall discuss below, extended estimation procedures have been devised, which overcome the problems related with the small τ-limit.

Lévy Processes   A procedure to analyze Lévy processes along the same lines has been proposed in [28]. An important point here is the determination of the Lévy parameter α.

Selfconsistency   After determining the drift vector and the characteristics of the noise sources from data, it is straightforward to synthetically generate data sets by iterating the corresponding stochastic evolution equations. Subsequently, their statistical properties can be compared with the properties of the real world data. This yields a selfconsistent check of the obtained results.

Estimation of Drift and Diffusion from Sparsely Sampled Time Series

As we have discussed, the results from an analysis of data sets can be reconsidered selfconsistently. This fact can be used to extend the procedure to data sets with insufficient amount of data or sparsely sampled time series, for which the estimation of conditional moments $M^{(i)}(\mathbf{x}, t, \tau)$ and the subsequent limiting procedure $\tau \to 0$ can not be performed accurately.

In this case, one may proceed as follows. In a first step one obtains a zeroth order approximation of drift vector $\mathbf{D}^{(1)}(\mathbf{x})$ and diffusion matrix $\mathbf{D}^{(2)}(\mathbf{x})$. Based on this estimate one performes, in a second step, a suitable ansatz for the drift vector and the diffusion matrix containing a set of free parameters σ

$$\mathbf{D}^{(1)}(\mathbf{x}, \sigma), \mathbf{D}^{(2)}(\mathbf{x}, \sigma) \tag{70}$$

defining a class of Langevin-equations. Each Langevin equation defines a joint probability distribution

$$f(\mathbf{x}_n, t_n; \ldots \mathbf{x}_1, t_1; \sigma) \tag{71}$$

This joint probability distribution can be compared with the experimental one, $f(\mathbf{x}_n, t_n; \ldots \mathbf{x}_1, t_1; \text{exp})$. The best representative of the class of Langevin equations for the reconstruction of experimental data is then obtained by minimizing a suitably defined distance between the two distributions:

$$\text{Dist}\{f(\mathbf{x}_n, t_n; \ldots \mathbf{x}_1, t_1; \sigma) - f(\mathbf{x}_n, t_n; \ldots \mathbf{x}_1, t_1; \text{exp})\}$$
$$= \text{Min} \tag{72}$$

A reasonable choice is the so-called Kullback–Leibler distance between two distributions, defined as

$$K = \int \mathrm{d}f(\mathbf{x}_n, t_n; \ldots; \mathbf{x}_1, t_1; \text{exp})$$
$$\cdot \ln \frac{f(\mathbf{x}_n, t_n; \ldots; \mathbf{x}_1, t_1; \text{exp})}{f(\mathbf{x}_n, t_n; \ldots; \mathbf{x}_1, t_1; \sigma)} \tag{73}$$

Recently, it has been shown how the iteration procedure can be obtained from maximum likelihood arguments.

For more details, we refer the reader to [29,30]. A technical question concerns the determination of the minimum. In [31] the limited-memory Broyden–Fletcher–Goldfarb–Shanno algorithm for constraint problems has been used for the solution of the optimization problem.

Applications: Processes in Time

The method outlined in the previous section has been used for revealing nonlinear deterministic behavior in a variety of problems ranging from physics, meteorology, biology to medicine. In most of these cases, alternative procedures with strong emphasis on deterministic features have only been partly successful due to their inappropriate treatment of dynamical fluctuations. The following list (with some exemplary citations) gives an overview on the investigated phenomena, which range from technical applications over many particle systems to biological and geophysical systems.

- Chatter in cutting processes [32,33]
- Identification of bifurcations towards drifting solitary structures in a gas-discharge system [34,35,36,37,38]
- Electric circuits [17,39,40]
- Wind energy convertors [41,42,43,44]
- Traffic flow [45]
- Inverse grading of granular flows [46]
- Heart rythms [47,48,49,50]
- Tremor data [39]
- Epileptic brain dynamics [51]
- Meteorological data like El NINO [52,53,54]
- Earth quake prediction [15]

The main advantage of the stochastic data analysis method is its independence on modeling assumptions. It is purely data driven and is based on the mathematical features of Markov processes. As mentioned above these properties can be verified and validated selfconsistently.

Before we proceed to consider some exemplary applications we would like to add the following comment. The described analysis method cleans data from dynamical and measurement noise and provides the drift vector field, i. e. one obtains the underlying deterministic dynamical system. In turn, this system can be analyzed by the methods of nonlinear time series analysis: One can determine proper embedding, Ljapunov-exponents, dimensions, fixed points, stable and unstable limit cycles etc. [55]. We want to point out that the determination of these quantities from uncleaned data usually is flawed by the presence of dynamical noise.

Synthetic Data: Potential Systems, Limit Cycles, Chaos

The above method of disentangling drift and dynamical noise has been tested for synthetically generated data. The investigated systems include the pitchfork bifurcation, the Van der Pol oscillator as an example of a nonlinear oscillator, as well as the noisy Lorenz equations as an example of a system exhibiting chaos [56,57]. Furthermore, it has been shown how one can analyze processes which additionally contain a time periodic forcing [58]. This is of high interest for analyzing systems exhibiting the phenomena of stochastic resonance. Quite recently, stochastic systems with time delay have been considered [59,60].

The results of these investigations may be summarized as follows: Provided there is enough data and the data is well sampled, it is possible to extract the underlying deterministic dynamics and the strength of the fluctuations accurately. Figure 3 summarizes what can be achieved for the example of the noisy Lorenz model. For a detailed discussion we refer the reader to [57].

Noisy Circuits

Next, we present the application of the method to data sets from experimental systems. As first example, a chaotic electric circuit has been chosen. Its dynamics is formed by a damped oscillator with nonlinear energy support and additional dynamic noise terms. In this case, well defined electric quantities are measured for which the dynamic equations are known. The measured time series are analyzed according to the numerical algorithm described

above. Afterwards, the numerically determined results and the expected results according to the system's equations are compared.

The dynamic equations of the electric circuit are given by the following equations, where the deterministic part is known as Shinriki oscillator [61]:

$$\dot{X}_1 = \left(-\frac{1}{R_{NIC}C_1} - \frac{1}{R_1 C_1} \right) X_1$$
$$- \frac{f(X_1 - X_2)}{C_1} + \frac{1}{R_{NIC}C_1} \Gamma(t) \tag{74}$$
$$= g_1(X_1, X_2) + h_1 \Gamma(t) \tag{75}$$

$$\dot{X}_2 = \frac{f(X_1 - X_2)}{C_2} - \frac{1}{R_3 C_2} X_3 = g_2(X_1, X_2, X_3) \tag{76}$$

$$\dot{X}_3 = -\frac{R_3}{L}(X_2 - X_3) = g_3(X_2, X_3) \tag{77}$$

X_1, X_2 and X_3 denote voltage terms, R_i are values of resistors, L and C stand for inductivity and capacity values. The function $f(X_1 - X_2)$ denotes the characteristic of the nonlinear element. The quantities X_i, characterizing the stochastic variable of the Shinriki oscillator with dynamical noise, were measured by means of a 12 bit A/D converter. Our analysis is based on the measurement of 100.000 data points [39]. The attractor of the noise free dynamics is shown in Fig. 4.

The measured 3-dimensional time series were analyzed as outlined above. The determined deterministic dynamics – expressed by the deterministic part of the evolution equations – corresponds to a vector field in the

Lorenz systems – Reconstructed trajectories
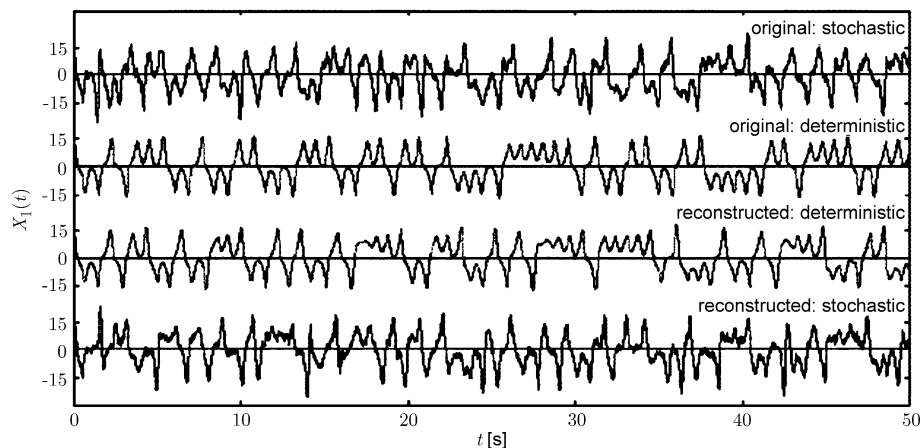
Fluctuations, Importance of: Complexity in the View of Stochastic Processes, Figure 3
Time series of the stochastic Lorenz equation from top to bottom: a) Original time series b) Deterministic time series c) Time series obtained from an integration of the reconstructed vector field d) Reconstructed time series including noise. For details cf. [57]
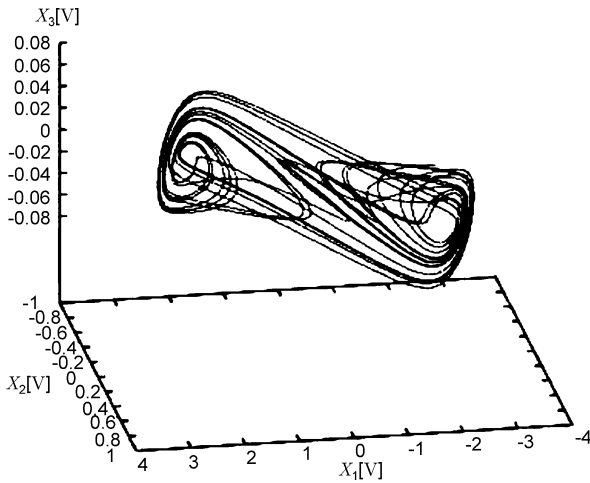
Fluctuations, Importance of: Complexity in the View of Stochastic Processes, Figure 4

Trajectory for the Shinriki oscillator in the phase space without noise; after [40]

Fluctuations, Importance of: Complexity in the View of Stochastic Processes, Figure 5

Cuts of the function $D^{(1)}(x)$ reconstructed from experimental data of the electric circuit in comparison with the expected functions according to the known differential (Eq. (75), (76)). In part a the cut $g_1(X_1, X_2), g_2(X_1, X_2, X_3 = 0)$ is shown as a two-dimensional vector field. *Thick arrows* represent values determined by data analysis, *thin arrows* represent the theoretically expected values. In areas of the state space where the trajectory did not show up during the measurement no estimated values for the functions are obtained. Figure b shows the one dimensional cut $g_1(X_1, X_2 = 0)$. *Crosses* represent values estimated numerically by data analysis. Additionally, the affiliated theoretically curve is printed as well; after [39]

three dimensional state space. For presentation of the results, cuts of lower dimension have been generated. Part a) of Fig. 5 illustrates the vector field $(g_1(x_1, x_2), g_2(x_1, x_2, x_3 = 0))$ of the reconstructed deterministic parts affiliated with (75), (76). Furthermore, the one-dimensional curve $g_1(x_1, x_2 = 0)$ is drawn in part b). Additionally to the numerically determined results found by data analysis the expected vector field and curve (75), (76) are shown for comparison. A good agreement can be recognized.

Based on the reconstructed deterministic functions it is possible to reconstruct also the noisy part from the data set, see Subsect. "Direct Estimation of Stochastic Forces". This has been performed for the three dimensionally embedded data, as well for the case of two dimensional embedding. From these reconstructed noise data, the autocorrelation was estimated. As shown in Fig. 6 correlated noise is obtained for wrong embedding indicating the violation of Markovian properties. In fact such an approach can be used to verify the correct embedding of nonlinear noisy dynamical systems. We emphasize that, provided sufficient data is available, this check of correct embedding can also be performed locally in phase space to find out where for the corresponding deterministic system crossing of trajectories take place. This procedure can be utilized to find the correct local embedding of data.

The electronic circuit of the Shinriki oscillator has also been investigated with two further perturbations. In [17] the reconstruction of the deterministic dynamics by the presence of additional measurement noise has been ad-

dressed. In [40] the Langevin noise has been exchanged by a high frequency periodic source, as shown in Fig. 7. Even for this case reasonable (correct) estimations of the deterministic part can be achieved.

Many Particle Physics – Traffic Flow

Far from equilibrium interacting many particle systems exhibit collective macroscopic phenomena like pattern formation, which can be described by the concept of order parameters. In the following we shall exemplify for the case of traffic flow how complex behavior can be analyzed by means of the proposed method leading to stochastic equations for the macroscopic description of the system.

a

b
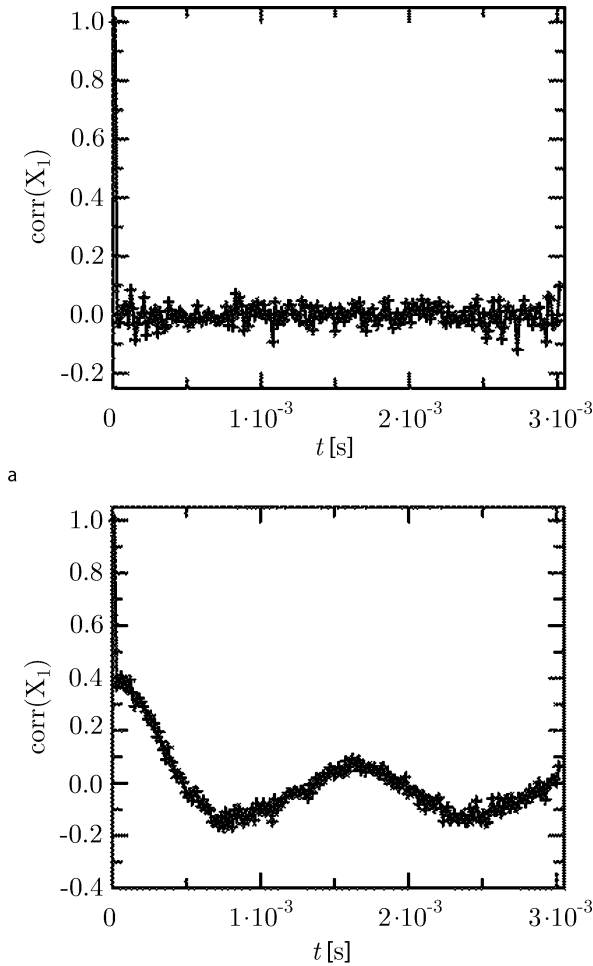
Fluctuations, Importance of: Complexity in the View of Stochastic Processes, Figure 6
Autocorrelation function of reconstructed dynamical noise, **a** correctly three dimensional embedded showing δ-correlated noise, **b** the projected dynamics in the twodimensional phase space $X_1(t)$ and $X_3(t)$, showing finite time correlations; after [17]

Traffic flow certainly is a collective phenomena, for which a huge amount of data is available. Measured quantities are velocity v and current $q = \rho v$ of cars passing a fixed line on the highway. Theoretical models of traffic flow are based on the so-called fundamental diagram, which is a type of material law for traffic flow relating current and velocity of the traffic flow

$$q = Q(v) \tag{78}$$

The special form of this relation has been much under debate in recent years.

a

b
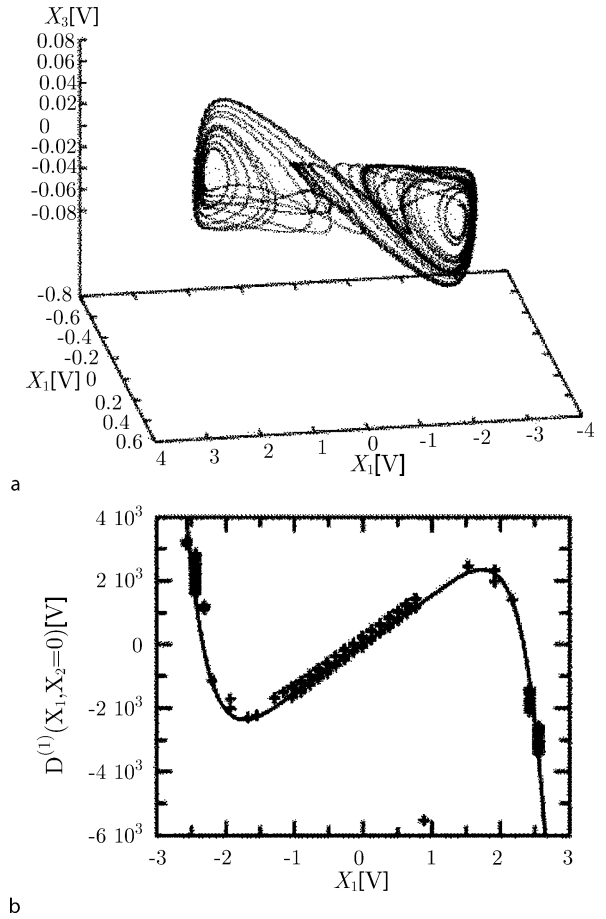
Fluctuations, Importance of: Complexity in the View of Stochastic Processes, Figure 7
a Trajectory for the Shinriki oscillator in the phase space with a sinusoidal force. **b** The corresponding trajectory in the phase space, compare Fig. 5b; after [40]

It is tempting to describe the dynamics by the following set of stochastic difference equation

$$v_{N+1} = G(v_N, q_N) + \xi_N$$
$$q_{N+1} = F(v_N, q_N) + \eta_N \tag{79}$$

Here, v_{N+1}, $q_N + 1$ are velocity and current of the $N + 1$ car traversing the line. ξ_N and η_N are noise terms with zero mean, which may depend on the variables u and q. The drift vector field $\mathbf{D}^1 = [D_1^1(v, q), D_2^1(v, q)]$ has been determined in [45]. We point out that the meaning of the drift vector field does not depend on the assumption of ideal noise sources, as has been discussed above Subsect. "Estimating the Short Time Propagator" and Subsect. "Noisy Circuits". The obtained drift vector field is depicted in Fig. 8.

a

b
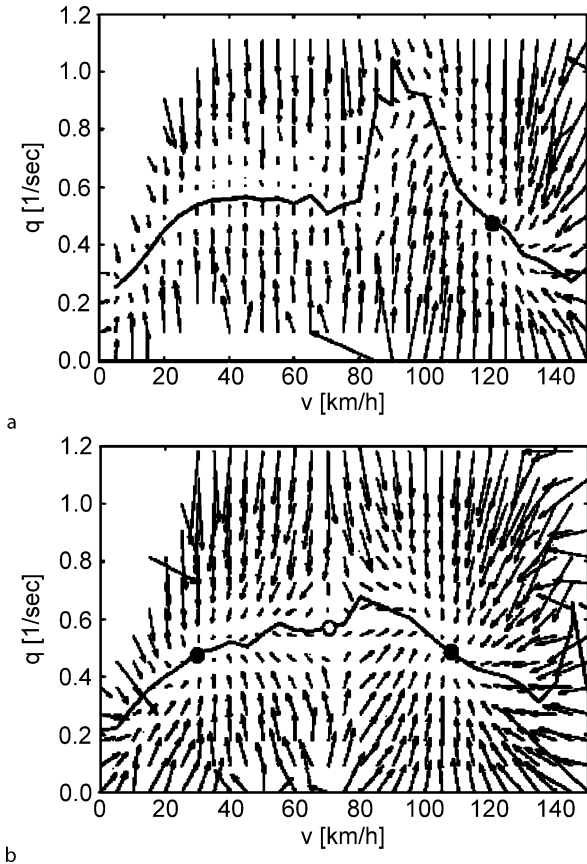
Fluctuations, Importance of: Complexity in the View of Stochastic Processes, Figure 8
Deterministic dynamics of the twodimensional traffic states (q, v) given by the drift vector for **a** right lane of freeway with vans and **b** all three lanes. Bold dots indicate stable fixed points and open dots saddle points, respectively; after [45]

a

b

Fluctuations, Importance of: Complexity in the View of Stochastic Processes, Figure 9
Corresponding potentials for the deterministic dynamics given by the drift coefficients of a one dimensional projection of the results in Fig. 8 [45]

The phase space yields interesting behavior. For the traffic data involving all three lanes there are three fixed points, two sinks separated by a saddle point. Furthermore, the arrows representing the drift vector field indicate the existence of an invariant manifold,

$$q = Q(v) \qquad (80)$$

which has to be interpreted as the above mentioned fundamental diagram of traffic flow. It is interesting to see differences caused by separation of the traffic dynamics into that of cars and that of vans. This has been roughly achieved by considering different lanes of the highway. In Figs. 8a and 9a the dynamics of the right lane caused by vans is shown. It can clearly be seen that up to a speed of about 80 km/h a meta stable plateau is present corresponding to a quasi interaction free dynamics of vans. Up to now the

information contained in the noise terms has not led to clear new insights.

For the discussion of further examples we refer the reader to the literature. In particular, we want to mention the analysis of segregational dynamics of single particles with different sizes in avalanches [46], which can be treated along similar lines .

Applications: Processes in Scale

In the following we shall consider complex behavior in scale as discussed in the introduction by the method of stochastic processes.

In order to statistically describe scale complexity in a comprehensive way, one has to study joint probability

distributions

$$f(q_N, l_N; q_{N-1}, l_{N-1}; \ldots; q_1, l_1; q_0, l_0) \tag{81}$$

of the local measure q at multiple scales l_i. To grasp relations across several scales (like given by "coherent" structures) all q_i tuples are taken at common locations x. In the case of statistical independency of q at different scales l_i, this joint probability density factorizes:

$$f(q_N, l_N; q_{N-1}, l_{N-1}; \ldots; q_1, l_1; q_0, l_0)$$
$$= f(q_N, l_N) \ldots f(q_0, l_0) . \tag{82}$$

In this case multifractal scaling (6) is a sufficient characterization of systems with scale complexity, provided the scaling property is given. If there is no factorization, one can use the joint probability distribution Eq. (81) to define a stochastic process in scale. Thus, one can identify the scale l with time t and try to obtain a representation of the spatial disorder in the form of a stochastic process similar to a stochastic process in time (see Subsect. "Statistical Description of Stochastic Processes" and Subsect. "Finite Time Propagators").

For such problems our method has been used as an alternative description of multifractal behavior. The present method has the advantage to relate the random variables across different scales by introducing a conditional probability distribution or, in fact, a two scale probability distribution. Scaling properties are no prerequisite. Provided that the Markovian property holds, which can be tested experimentally, a complete statistical characterization of the process across scale is obtained. The method has been used to characterize the complexity of data sets for the following systems (with some exemplary citations):

- Turbulent flows [20,62,63,64]
- Passive scalar in turbulent flows [21]
- Financial data [65,66,67,68]
- Surface roughness [69,70,71,72,73,74]
- Earthquakes [15]
- Cosmic background radiation [75]

The stochastic analysis of scale dependent complex system aims to achieve a n-scale characterization, from which the question arose, whether it will be possible to derive from these stochastic processes methods for generating synthetic time series.

Based on the knowledge of the n-scale statistics a way to estimate the n-point statistics has been proposed, which enables to generate synthetic time series with the same n-scale statistics [76]. It is even possible to extend given data sets, an interesting subject for forecasting. Other approaches has been proposed in [70,77].

Turbulence and Financial Market Data

In the following we present results from investigations of fully developed turbulence and from data of the financial market. The complexity of turbulent fields still has not been understood in detail. Although the basic equations, namely the Navier Stokes equations, are known for more than 150 years, a general solution of these equations for high Reynolds numbers, i. e. for turbulence, is not known. Even with the use of powerful computers no rigorous solutions can be obtained. Thus for a long time there has been the challenge to understand at least the complexity of an idealized turbulent situation, which is taken to be isotropic and homogeneous. The main problem is to formulate statistical laws based on the treatment of the deterministic evolution laws of fluid dynamics. A first approach is due to Kolmogorov [78], who formulated a phenomenological theory characterizing properties of turbulent fluid motions by statistical quantities. The central observable of Kolmogorov's theory is the so-called longitudinal velocity increment (which we label here with q) of a turbulent velocity field $\mathbf{u}(\mathbf{x}, t)$ defined according to:

$$\mathbf{q_x}(\mathbf{l}, t) = \frac{\mathbf{l}}{l} \cdot [\mathbf{u}(\mathbf{x} + \mathbf{l}, t) - \mathbf{u}(\mathbf{x}, t)] . \tag{83}$$

A statistical description is given in terms of the probability distribution

$$f(q, \mathbf{l}, t, \mathbf{x}) = \langle \delta(q - q_\mathbf{x}(\mathbf{l}, t)) \rangle . \tag{84}$$

For stationary, homogeneous and isotropic turbulence, this probability distribution is independent of the reference point $\mathbf{x}$, time t, and, due to isotropy, only depends on the scale $l = |\mathbf{l}|$. As a consequence, the central statistical quantity is the probability distribution $f(q, l)$. Turbulent fields have been considered from the viewpoint of selfsimilarity addressing scaling behavior of the probability distribution $f(q, l)$ and their nth order moments, the so-called structure functions $\langle q^n \rangle$. Multifractal scaling properties, mentioned already in Eq. (7), of the velocity increments for turbulence are identical to the well known intermittency problem[1], which manifests itself in the occurrence of heavy tailed statistics, i. e. an unexpected high probability

---

[1] Here it should be noted that the term "intermittency" is used frequently in physics for different phenomena, and may cause confusions. This turbulent intermittency is not equal to the intermittency of chaos. There are also different intermittency phenomena introduced for turbulence. There is this intermittency due to the nonlinear scaling, there is the intermittency of switches between turbulent and laminar flows for non local isotropic fully developed turbulent flows, there is the intermittency due to the statistics of small scale turbulence which we discuss here as heavy tailed statistics.

of extreme events. Kolmogorov and Oboukhov proposed the so-called intermittency correction [79]

$$\langle q(l, x)^n \rangle = l^{\xi_n} \text{ with } \xi_n = \frac{n}{3} - \mu \frac{n(n-3)}{18} \text{ and } n > 2$$

(85)

with $0.25 < \mu < 0.5$ (for further details see [80]). The form of ξ_n has been heavily debated during the last decades. For isotropic turbulence the central issue is to reveal the mechanism, which leads to this anomalous statistics (see [80,81]).

A completely different point of view of the properties of turbulent fields is gained by interpreting the probability distribution as the distribution of a random process in scale l [62,63]. It is tempting to hypothesize the existence of a stochastic process, see Subsect. "Stochastic Data Analysis"

$$q(l + dl) = q(l) + N(q, l)dl + g(q, l)dW(l),$$

(86)

where $dW(l)$ is an increment of a random process. This type of stochastic equation indicates how the velocity increment of a snapshot of the turbulent field changes as a function of scale l. In this respect, the process $q(l)$ can be considered to be a stochastic cascade process in "*time l*".

This concept of complexity in scale can be carried over to other systems like the roughness of surfaces or financial data. In the latter case the scale variable l is replaced by the time distance or time scale τ.

Anomalous Statistics

A direct consequence of multifractal scaling related with nonlinear behavior of the scaling exponents ξ_n is the fact that the shape of the probability distribution $f(q, l)$ has to change its shape as a function of scale. A selfsimilar behavior of the form (4) would lead to fractal scaling behavior, as outlined in Eq. (5). Using experimental data from a turbulent flow this change of the shape of the pdf becomes obvious. In Fig. 10 we present $f(q, l)$ for a data set measured in the central line of a free jet with Reynolds number of $2.7 \cdot 10^4$, see [20]. Note for large scales ($l \approx L_0$) the distributions become nearly Gaussian. As the scale is decreased the probability densities become more and more heavy tailed.

Quite astonishingly, the anomalous statistical features of data from the financial market are similar to the just discussed intermittency of turbulence [82]. The following analysis is based on a data set $Y(t)$, which consists of 1 472 241 quotes for US dollar-German Mark exchange
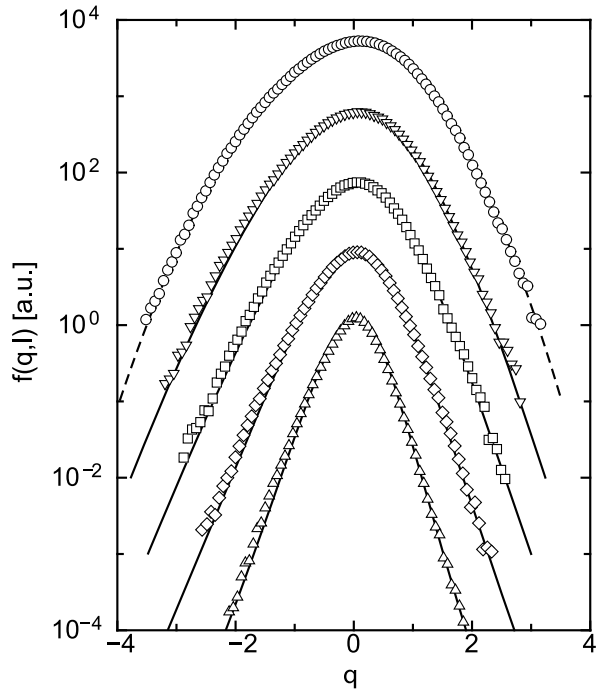
Fluctuations, Importance of: Complexity in the View of Stochastic Processes, Figure 10
Comparison of the numerical solution of the Fokker–Planck equation (*solid lines*) for the pdfs $f(q(x), l)$ with the pdfs obtained directly from the experimental data (bold symbols). The scales l are (from top to bottom): $l = L_0$, $0.6L_0$, $0.35L_0$, $0.2L_0$ and $0.1L_0$. The distribution at the largest scale L_0 was parametrized (*dashed line*) and used as initial condition for the Fokker–Planck equation (L_0 is the correlation length of the turbulent velocity signal). The pdfs have been shifted in vertical direction for clarity of presentation and all pdfs have been normalized to their standard deviations; after [20]

rates from the years 1992 and 1993. Many of the features we will discuss here are also found in other financial data like for instance quotes of stocks, see [83]. A central issue is the understanding of the statistics of price changes over a certain time interval τ which determines losses and gains. The changes of a time series of quotations $Y(t)$ are commonly measured by returns $r(\tau, t) := Y(t + \tau)/Y(t)$, logarithmic returns or by increments $q(\tau, t) := Y(t + \tau) - Y(t)$ [84]. The moments of these quantities often exhibit power law behavior similar to the just discussed Kolmogorov scaling for turbulence, cf. [85,86,87]. For the probability distributions one additionally observes an increasing tendency to heavy tailed probability distributions for small τ (see Fig. 11). This represents the high frequency dynamics of the financial market. The identification of the underlying process leading to these heavy tailed probability density functions of price
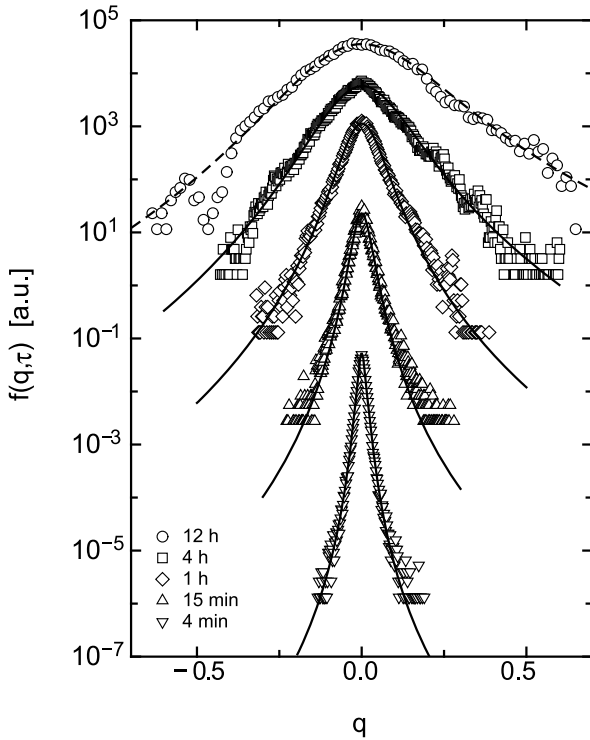
Fluctuations, Importance of: Complexity in the View of Stochastic Processes, Figure 11

Probability densities (*pdf*) $f(q(t), \tau)$ of the price changes $q(\tau, t) = Y(t + \tau) - Y(t)$ for the time delays $\tau = 5120, 10240, 20480, 40960\,s$ (from bottom to top). Symbols: results obtained from the analysis of middle prices of bit-ask quotes for the US dollar-German Mark exchange rates from October 1st, 1992 until September 30th, 1993. *Full lines*: results from a numerical iteration of the Fokker–Planck equation (95); the probability distribution for $\tau = 40960s$ (*dashed lines*) was taken as the initial condition. The pdfs have been shifted in vertical direction for clarity of presentation and all pdfs have been normalized to their standard deviations; after [66]

changes is a prominent puzzle (see [86,87,88,89,90,91]), like it is for turbulence.

Stochastic Cascade Process for Scale

The occurrence of the heavy tailed probability distributions on small scales will be discussed as a consequence of stochastic process evolving in scale, using the above mentioned methods. Guided by the finding that the statistics changes with scale, as shown in Figs. 10 and 11, we consider the evolution of the quantity $q(l, x)$, or $q(\tau, t)$ with the scale variable l or τ, respectively.

For a single fixed scale l we get the scale dependent disorder by the statistics of $q(l, x)$. The complete stochastic information of the disorder on all length scales is given by

the joint probability density function

$$f(q_1, \ldots, q_n), \qquad (87)$$

where we set $q_i = q(l_i, x)$. Without loss of generality we take $l_i < l_{i+1}$. This joint probability may be seen in analogy to joint probabilities of statistical mechanics (thermodynamics), describing in the most general way the occupation probabilities of the microstates of n particles, where q is a six-dimensional phase state vector (space and momentum).

Next, the question is, whether it is possible to simplify the joint probability by conditional probabilities:

$$f(q_1, \ldots, q_n) = p(q_1 | q_2, \ldots, q_n) \\ \cdot p(q_2 | q_3, \ldots, q_n) \ldots p(q_{n-1} | q_n) f(q_n), \qquad (88)$$

where the multiple conditioned probabilities are given by

$$p(q_i \,|\, q_{i+1}, \ldots, q_n) = p(q_i \,|\, q_{i+1}). \qquad (89)$$

Eq. (89) is nothing else than the condition for a Markov process evolving from state q_{i+1} to the state q_i, i.e. from scale l_{i+1} to l_i as it has been introduced above, see Eq. (35).

A "single particle approximation" would correspond to the representation:

$$f(q_1, \ldots, q_n) = f(q_1) f(q_2) \ldots f(q_n). \qquad (90)$$

According to Eqs. (89) and (90), Eq. (91) holds if $p(q_i | q_{i+1}) = f(q_i)$. Only for this case, the knowledge of $f(q_i)$ is sufficient to characterize the complexity of the whole system. Otherwise an analysis of the single scale probability distribution $f(q_i)$ is an incomplete description of the complexity of the whole system. This is the deficiency of the approach characterizing complex structures by means of fractality or multiaffinity (cf. for multiaffinity [92], for turbulence [80,81], for financial market [85]). The scaling analysis of moments as indicated for turbulence in Eq. (85) provides a complete knowledge of any joint n-scale probability density only if Eq. (90) is valid.

These remarks underline the necessity to investigate these conditional probabilities, which can be done in a straightforward manner from given experimental or numerical data. For the case of turbulence as well as for financial data we see that $p(q_i | q_{i+1})$ does not coincide with $f(q_i)$, as it is shown for turbulence data in Fig. 12. If $p(q_i | q_{i+1}) = f(q_i)$ no dependency on q_{i+1} could be detected.

The next point of interest is whether the Markov properties are fulfilled. Therefore doubly conditioned probabil-

a

b
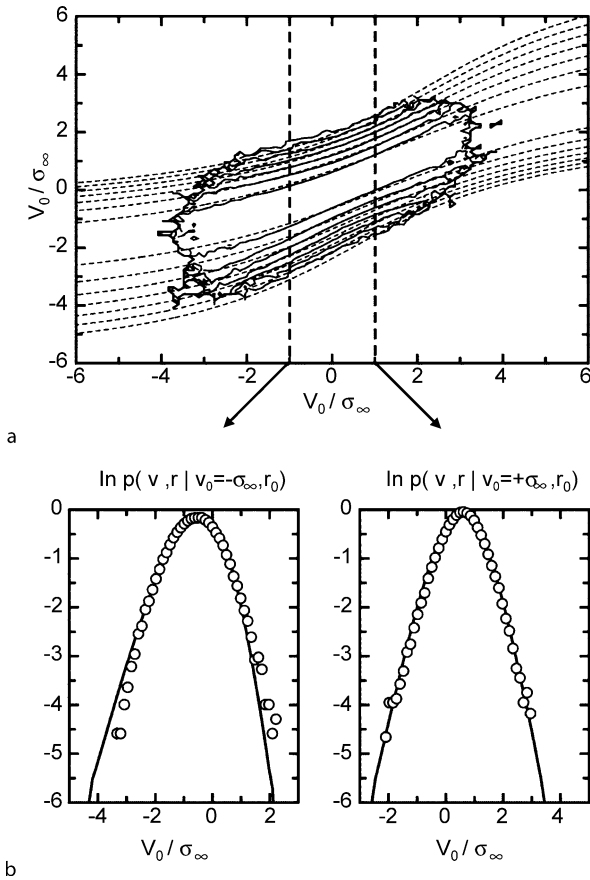
Fluctuations, Importance of: Complexity in the View of Stochastic Processes, Figure 12

Comparison of the numerical solution of the Fokker–Planck equation for the conditional pdf $p(q, l|q_0, l_0)$ denoted in this figure as $p(v, r|v_0, r_0)$ with the experimental data. a: Contour plots of $p(v, r|v_0, r_0)$ for $r_0 = L$ and $r = 0.6L$, where L denotes the integral length. *Dashed lines*: numerical solution of (94), *solid lines*: experimental data. b and c: Cuts through $p(v, r|v_0, r_0)$ for $v_0 = +\sigma_\infty$ and $v_0 = -\sigma_\infty$ respectively. *Open symbols*: experimental data, solid lines: numerical solution of the Fokker–Planck equation; after [20]

ities were extracted from data and compared to the single conditioned ones. For financial data as well as for turbulence we found evidence that the Markov property is fulfilled if the step size is larger than a Markov–Einstein length [20,66]. For turbulence it could be shown that the Markov–Einstein length coincides with the Taylor length marking the small scale end of the inertial range [14]. (The extensive discussion of the analysis of financial and turbulent data can be found in [20,65,66,93]).

Based on the fact that the multiconditioned probabilities are equal to the single conditioned probabilities, and taking this as a verification of Markovian properties we can proceed according to Subsect. "Estimating the Short Time

Propagator" and estimate from given data sets the stochastic equations underlying the cascade process. The evolution of the conditional probability density $p(q, l|q_0, l_0)$ starting from a selected scale l_0 follows

$$-l\frac{\partial}{\partial l}\, p(q, l|q_0, l_0)$$
$$= \sum_{k=1}^{\infty} \frac{1}{k!}\left(-\frac{\partial}{\partial q}\right)^k D^{(k)}(q, l)\, p(q, l|q_0, l_0)\,. \tag{91}$$

(The minus sign on the left side is introduced, because we consider processes running to smaller scales l, furthermore we multiply the stochastic equation by l, which leads to a new parametrization of the cascade by the variable $\ln(1/l)$, a simplification for a process with scaling law behavior of its moments.) This equation is known as the Kramers–Moyal expansion [8]. As outlined in Subsect. "Finite Time Propagators" and Subsect. "Estimating the Short Time Propagator", the Kramers–Moyal coefficients $D^{(k)}(q, l)$ are now defined as the limit $\Delta l \to 0$ of the conditional moments $M^{(k)}(q, l, \Delta l)$:

$$D^{(k)}(q, l) = \lim_{\Delta l \to 0} M^{(k)}(q, l, \Delta l)\,, \tag{92}$$

$$M^{(k)}(q, l, \Delta l) := \frac{l}{\Delta l}\int_{-\infty}^{+\infty}(\tilde{q} - q)^k\, p\,(\tilde{q}, l - \Delta l\,|\,q, l)\,\mathrm{d}\,\tilde{q}. \tag{93}$$

Thus, for the estimation of the $D^{(k)}$ coefficients it is only necessary to estimate the conditional probabilities $p(\tilde{q}, l - \Delta l|q, l)$. For a general stochastic process, all Kramers–Moyal coefficients are different from zero. According to Pawula's theorem, however, the Kramers–Moyal expansion stops after the second term, provided that the fourth order coefficient $D^{(4)}(q, l)$ vanishes. In that case the Kramers–Moyal expansion is reduced to a Fokker–Planck equation:

$$-l\frac{\partial}{\partial l}\, p(q, l|q_0, l_0)$$
$$= \left\{-\frac{\partial}{\partial q} D^{(1)}(q, l) + \frac{1}{2}\frac{\partial^2}{\partial q^2} D^{(2)}(q, l)\right\} p(q, l|q_0, l_0)\,. \tag{94}$$

$D^{(1)}$ is denoted as drift term, $D^{(2)}$ as diffusion term now for the cascade process. The probability density function $f(q, l)$ has to obey the same equation:

$$-l\frac{\partial}{\partial l}\, f(q, l)$$
$$= \left\{-\frac{\partial}{\partial q} D^{(1)}(q, l) + \frac{1}{2}\frac{\partial^2}{\partial q^2} D^{(2)}(q, l)\right\} f(q, l)\,. \tag{95}$$

The novel point of our analysis here is that knowing the evolution equation (91), the n-increment statistics $f(q_1, \ldots, q_n)$ can be retrieved as well. Definitely, information like scaling behavior of the moments of $q(l, x)$ can also be extracted from the knowledge of the process equations. Multiplying (91) by q^n and successively integrating over q, an equation for the moments is obtained:

$$-l \frac{\partial}{\partial l} \langle q^n \rangle = \sum_{k=1}^{n} \left(-\frac{\partial}{\partial q} \right)^k \frac{n!}{k!(n-k)!} \langle D^{(k)}(q, l) q^{n-k} \rangle.$$

(96)

Scaling, i. e. multi affinity as described in Eq. (7), is obtained if $D^{(k)}(q, l) \propto q^k$, see [71,94].

We summarize: By the described procedure we were able to reconstruct stochastic processes in scale directly from given data. Knowing these processes one can perform numerical solutions in order to obtain a selfconsistent check of the procedure(see [20,66]). In Figs. 10, 11 and 12 the numerical solutions are shown by solid (dashed) curves. The heavy tailed structure of the single scale probabilities as well as the conditional probabilities are well described by this approach based on a Fokker–Planck equation. Further improvements can be achieved by optimization procedures mentioned in Subsect. "Estimation of Drift and Diffusion from Sparsely Sampled Time Series".

New Insights

The fact that the complexity of financial market data as well as turbulent data can be expressed by a Markovian process in scale, has the consequence that the conditioned probabilities involving only two different scales are sufficient for the general n scale statistics. This indicates that three or four point correlation are sufficient for the formulation of n-point statistics.

The finding of a finite length scale above which the Markov properties are fulfilled [14] has lead to a new interpretation of the Taylor length for turbulence, which so far had no specific physical meaning.

For financial, as well as, for the turbulent data it has been found that the diffusion term is quadratic in the state space of the scale resolved variable. With respect to the corresponding Langevin equation, the *multiplicative nature of the noise* term becomes evident, which causes heavy tailed probability densities and multifractal scaling. The scale dependency of drift and diffusion terms corresponds to a non-stationary process in scale variables τ and l, respectively. From this point we conclude that a Levy statistics for one fixed scale, i. e. for the statistics of $q(l, x)$ for

fixed l can not be an adequate class for the statistical description of financial and turbulent data.

Comparing the maximum of the distributions at small scales in Figs. 10 and 11 one finds a less sharp tip for the turbulence data. This finding is in accordance with a comparably larger additive contribution in the diffusion term $D^{(2)} = a + bq^2$ for turbulence data. Knowing that $D^{(2)}$ has an additive term and quadratic q dependence it is clear that for small q values, i. e. for the tips of the distribution, the additive term dominates. Taking this result in combination with the Langevin equation, we see that for small q values Gaussian noise is present, which leads to a *Gaussian tip* of the probability distribution, as found in Fig. 10.

A further consequence of the additive term in $D^{(2)}$ is that the structure functions as given by Eq. (96) are not independent and a general scaling solution does not exist. This fact has been confirmed by optimizing the coefficients [31]. This *nonscaling behavior of turbulence* seems to be present also for higher Reynolds numbers. It has been found that the additive terms becomes smaller but still remains relevant [96]. The cascade processes encoded in the functions $D^{(1)}$ and $D^{(2)}$ seem to depend on the Reynolds number, which might indicate that turbulence is less universal as commonly thought.

As has been outlined above it is straight forward to extend the analysis to higher dimensions. For the case of *longitudinal and transversal increments* a symmetry for these two different directions of the complex velocity field has been found in the way that the cascade process runs with different speeds. The factor is 3/2 [64,96] and again is not in accordance with the proposed multifractal scaling property of turbulence.

The investigation of the advection of passive scalar in a turbulent flow along the present method has revealed the interesting result that the Markovian properties are fulfilled but that higher order Kramers–Moyal coefficients are nonnegligible [21]. This indicates that for passive scalars non-Gaussian noise is present, which can be attributed to the existence of shock like structures in the distribution of the passive scalar.

Future Directions

The description of complex systems on the basis of stochastic processes, which include nonlinear dynamics, seems to by a promising approach for the future. The challenge will be to extend the understanding to more complicated processes, like Levy processes, processes with no white noise or higher dimensional processes, just to mention some. As it has been shown in this contribution, for these cases it should be possible to derive from precise

mathematical results general methods of data series analysis, too.

Besides the further improve of the method, we are convinced that there is still a wide range of further applications. Advanced sensor techniques enables scientists to collect huge data sets measured with high precision. Based on the stochastic approach we have presented here it is not any more the question to put much efforts into noise reduction, but in contrary the involved noise can help to derive a better characterization and thus a better understanding of the system considered. Thus there seem to be many applications in the inanimate and the animate world, ranging form technical applications over socio-econo systems to biomedical applications. An interesting feature will be the extraction of of higher correlation aspects, like the question of the cause and effect chain, which may be unfolded by asymmetric determinism and noise terms reconstructed from data.

Further Reading

For further reading we suggest the publications [1,2,3,4, 8,9,10,11].

Acknowledgment

The scientific results reported in this review have been worked out in close collaboration with many colleagues and students. We mention St. Barth, F. Böttcher, F. Ghasemi, I. Grabec, J. Gradisek, M. Haase, A. Kittel, D. Kleinhans, St. Lück, A. Nawroth, Chr. Renner, M. Siefert, and S. Siegert.

Bibliography

1. Haken H (1983) Synergetics, An Introduction. Springer, Berlin
2. Haken H (1987) Advanced Synergetics. Springer, Berlin
3. Haken H (2000) Information and Self-Organization: A Macroscopic Approach to Complex Systems. Springer, Berlin
4. Kantz H, Schreiber T (1997) Nonlinear Time Series Analysis. Cambridge University Press, Cambridge
5. Yanovsky VV, Chechkin AV, Schertzer D, Tur AV (2000) Physica A 282:13
6. Schertzer D, Larchevéque M, Duan J, Yanovsky VV, Lovejoy S (2001) J Math Phys 42:200
7. Gnedenko BV, Kolmogorov AN (1954) Limit distributions of sums of independent random variables. Addison-Wesley, Cambridge
8. Risken H (1989) The Fokker-Planck Equation. Springer, Berlin
9. Gardiner CW (1983) Handbook of Stochastic Methods. Springer, Berlin
10. van Kampen NG (1981) Stochastic processes in physics and chemistry. North-Holland Publishing Company, Amsterdam
11. Hänggi P, Thomas H (1982) Stochastic processes: time evolution, symmetries and linear response. Phys Rep 88:207
12. Einstein A (1905) Über die von der molekularkinetischen Theorie der Wärme geforderte Bewegung von in ruhenden Flüssigkeiten suspendierten Teilchen. Ann Phys 17:549
13. Friedrich R, Zeller J, Peinke J (1998) A Note in Three Point Statistics of Velocity Increments in Turbulence. Europhys Lett 41:153
14. Lück S, Renner Ch, Peinke J, Friedrich R (2006) The Markov Einstein coherence length a new meaning for the Taylor length in turbulence. Phys Lett A 359:335
15. Tabar MRR, Sahimi M, Ghasemi F, Kaviani K, Allamehzadeh M, Peinke J, Mokhtari M, Vesaghi M, Niry MD, Bahraminasab A, Tabatabai S, Fayazbakhsh S, Akbari M (2007) Short-Term Prediction of Mediumand Large-Size Earthquakes Based on Markov and Extended Self-Similarity Analysis of Seismic Data. In: Bhattacharyya P, Chakrabarti BK (eds) Modelling Critical and Catastrophic Phenomena in Geoscience. Lecture Notes in Physics, vol 705. Springer, Berlin, pp 281–301
16. Kolmogorov AN (1931) Über die analytischen Methoden in der Wahrscheinlichkeitsrechnung. Math Ann 140:415
17. Siefert M, Kittel A, Friedrich R, Peinke J (2003) On a quantitative method to analyze dynamical and measurement noise. Europhys Lett 61:466
18. Böttcher F, Peinke J, Kleinhans D, Friedrich R, Lind PG, Haase M (2006) On the proper reconstruction of complex dynamical systems spoilt by strong measurement noise. Phys Rev Lett 97:090603
19. Kleinhans D, Friedrich R, Wächter M, Peinke J (2007) Markov properties under the influence of measurement noise. Phys Rev E 76:041109
20. Renner C, Peinke J, Friedrich R (2001) Experimental indications for Markov properties of small scale turbulence. J Fluid Mech 433:383
21. Tutkun M, Mydlarski L (2004) Markovian properties of passive scalar increments in grid-generated turbulence. New J Phys 6:49
22. Marcq P, Naert A (2001) A Langevin equation for turbulent velocity increments. Phys Fluids 13:2590
23. Langner M, Peinke J, Rauh A (2008) A Langevin analysis with application to a Rayleigh-Bénard convection experiment. Exp Fluids (submitted)
24. Wächter M, Kouzmitchev A, Peinke J (2004) Increment definitions for sale-dependent analysis of stochastic data. Phys Rev E 70:055103(R)
25. Ragwitz M, Kantz H (2001) Indispensable finite time corrections for Fokker-Planck equations from time series. Phys Rev Lett 87:254501
26. Ragwitz M, Kantz H (2002) Comment on: Indispensable finite time correlations for Fokker-Planck equations from time series data-Reply. Phys Rev Lett 89:149402
27. Friedrich R, Renner C, Siefert M, Peinke J (2002) Comment on: Indispensable finite time correlations for Fokker-Planck equations from time series data. Phys Rev Lett 89:149401
28. Siegert S, Friedrich R (2001) Modeling nonlinear Lévy processes by data analysis. Phys Rev E 64:041107
29. Kleinhans D, Friedrich R, Nawroth AP, Peinke J (2005) An iterative procedure for the estimation of drift and diffusion coefficients of Langevin processes. Phys Lett A 346:42
30. Kleinhans D, Friedrich R (2007) Note on Maximum Likeli-

hood estimation of drift and diffusion functions. Phys Lett A 368:194

31. Nawroth AP, Peinke J, Kleinhans D, Friedrich R (2007) Improved estimation of Fokker-Planck equations through optimisation. Phys Rev E 76:056102

32. Gradisek J, Grabec I, Siegert S, Friedrich R (2002). Stochastic dynamics of metal cutting: Bifurcation phenomena in turning. Mech Syst Signal Process 16(5):831

33. Gradisek J, Siegert S, Friedrich R, Grabec I (2002) Qualitative and quantitative analysis of stochastic processes based on measured data-I. Theory and applications to synthetic data. J Sound Vib 252(3):545

34. Purwins HG, Amiranashvili S (2007) Selbstorganisierte Strukturen im Strom. Phys J 6(2):21

35. Bödeker HU, Röttger M, Liehr AW, Frank TD, Friedrich R, Purwins HG (2003) Noise-covered drift bifurcation of dissipative solitons in planar gas-discharge systems. Phys Rev E 67:056220

36. Purwins HG, Bödeker HU, Liehr AW (2005) In: Akhmediev N, Ankiewicz A (eds) Dissipative Solitons. Springer, Berlin

37. Bödeker HU, Liehr AW, Frank TD, Friedrich R, Purwins HG (2004) Measuring the interaction law of dissipative solitions. New J Phys 6:62

38. Liehr AW, Bödeker HU, Röttger M, Frank TD, Friedrich R, Purwins HG (2003) Drift bifurcation detection for dissipative solitons. New J Phys 5:89

39. Friedrich R, Siegert S, Peinke J, Lück S, Siefert M, Lindemann M, Raethjen J, Deuschl G, Pfister G (2000) Extracting model equations from experimental data. Phys Lett A 271:217

40. Siefert M Peinke J (2004) Reconstruction of the Deterministic Dynamics of Stochastic systems. Int J Bifurc Chaos 14:2005

41. Anahua E, Lange M, Böttcher F, Barth S, Peinke J (2004) Stochastic Analysis of the Power Output for a Wind Turbine. DEWEK 2004, Wilhelmshaven, 20–21 October 2004

42. Anahua E, Barth S, Peinke J (2006) Characterization of the wind turbine power performance curve by stochastic modeling. EWEC 2006, BL3.307, Athens, February 27–March 2

43. Anahua E, Barth S, Peinke J (2007) Characterisation of the power curve for wind turbines by stochastic modeling. In: Peinke J, Schaumann P, Barth S (eds) Wind Energy – Proceedings of the Euromech Colloquium. Springer, Berlin, p 173–177

44. Anahua E, Barth S, Peinke J (2008) Markovian Power Curves for Wind Turbines. Wind Energy 11:219

45. Kriso S, Friedrich R, Peinke J, Wagner P (2002) Reconstruction of dynamical equations for traffic flow. Phys Lett A 299:287

46. Kern M, Buser O, Peinke J, Siefert M, Vulliet L (2005) Stochastic analysis of single particle segregational dynamics. Phys Lett A 336:428

47. Kuusela T (2004) Stochastic heart-rate model can reveal pathologic cardiac dynamics. Phys Rev E 69:031916

48. Ghasemi F, Peinke J, Reza Rahimi Tabar M, Muhammed S (2006) Statistical properties of the interbeat interval cascade in human subjects. Int J Mod Phys C 17:571

49. Ghasemi F, Sahimi M, Peinke J, Reza Rahimi Tabar M (2006) Analysis of Non-stationary Data for Heart-rate Fluctuations in Terms of Drift and Diffusion Coefficients. J Biological Phys 32:117

50. Tabar MRR, Ghasemi F, Peinke J, Friedrich R, Kaviani K, Taghavi F, Sadghi S, Bijani G, Sahimi M (2006) New computational approaches to analysis of interbeat intervals in human subjects. Comput Sci Eng 8:54

51. Prusseit J, Lehnertz K (2007) Stochastic Qualifiers of Epileptic Brain Dynamics. Phys Rev Lett 98:138103

52. Sura P, Gille ST (2003) Interpreting wind-driven Southern Ocean variability in a stochastic framework. J Marine Res 61:313

53. Sura P (2003) Stochastic Analysis of Southern and Pacific Ocean Sea Surface Winds. J Atmospheric Sci 60:654

54. Egger J, Jonsson T (2002) Dynamic models for islandic meteorological data sets. Tellus A 51(1):1

55. Letz T, Peinke J, Kittel A (2008) How to characterize chaotic time series distorted by interacting dynamical noise. Preprint

56. Siegert S, Friedrich R, Peinke J (1998) Analysis of data sets of stochastic systems. Phys Lett A 234:275–280

57. Gradisek J, Siegert S, Friedrich R, Grabec I (2000) Analysis of time series from stochastic processes. Phys Rev E 62:3146

58. Gradisek J, Friedrich R, Govekar E, Grabec I (2002) Analysis of data from periodically forced stochastic processes. Phys Lett A 294:234

59. Frank TD, Beek PJ, Friedrich R (2004) Identifying noise sources of time-delayed feedback systems. Phys Lett A 328:219

60. Patanarapeelert K, Frank TD, Friedrich R, Beek PJ, Tang IM (2006) A data analysis method for identifying deterministic components of stable and unstable time-delayed systems with colored noise. Phys Lett A 360:190

61. Shinriki M, Yamamoto M, Mori S (1981) Multimode Oscillations in a Modified Van-der-Pol Oscillator Containing a Positive Nonlinear Conductance. Proc IEEE 69:394

62. Friedrich R, Peinke J (1997). Statistical properties of a turbulent cascade. Physica D 102:147

63. Friedrich R, Peinke J (1997) Description of a turbulent cascade by a Fokker-Planck equation. Phys Rev Lett 78:863

64. Siefert M, Peinke J (2006) Joint multi-scale statistics of longitudinal and transversal increments in small-scale wake turbulence. J Turbul 7:1

65. Friedrich R, Peinke J, Renner C (2000) How to quantify deterministic and random influences on the statistics of the foreign exchange market. Phys Rev Lett 84:5224

66. Renner C, Peinke J, Friedrich R (2001) Markov properties of high frequency exchange rate data. Physica A 298:499–520

67. Ghasemi F, Sahimi M, Peinke J, Friedrich R, Reza Jafari G, Reza Rahimi Tabar M (2007) Analysis of Nonstationary Stochastic Processes with Application to the Fluctuations in the Oil Price. Phys Rev E (Rapid Commun) 75:060102

68. Farahpour F, Eskandari Z, Bahraminasab A, Jafari GR, Ghasemi F, Reza Rahimi Tabar M, Muhammad Sahimi (2007) An Effective Langevin Equation for the Stock Market Indices in Approach of Markov Length Scale. Physica A 385:601

69. Wächter M, Riess F, Kantz H, Peinke J (2003) Stochastic analysis of raod surface roughness. Europhys Lett 64:579

70. Jafari GR, Fazeli SM, Ghasemi F, Vaez Allaei SM, Reza Rahimi Tabar M, Iraji zad A, Kavei G (2003) Stochastic Analysis and Regeneration of Rough Surfaces. Phys Rev Lett 91:226101

71. Friedrich R, Galla T, Naert A, Peinke J, Schimmel T (1998) Disordered Structures Analyzed by the Theory of Markov Processes. In: Parisi J, Müller S, Zimmermann W (eds) A Perspective Look at Nonlinear Media. Lecture Notes in Physics, vol 503. Springer, Berlin

72. Waechter M, Riess F, Schimmel T, Wendt U, Peinke J (2004) Stochastic analysis of different rough surfaces. Eur Phys J B 41:259

73. Sangpour P, Akhavan O, Moshfegh AZ, Jafari GR, Reza Rahimi

Tabar M (2005) Controlling Surface Statistical Properties Using Bias Voltage: Atomic force microscopy and stochastic analysis. Phys Rev B 71:155423

74. Jafari GR, Reza Rahimi Tabar M, Iraji zad A, Kavei G (2007) Etched Glass Surfaces, Atomic Force Microscopy and Stochastic Analysis. J Phys A 375:239

75. Ghasemi F, Bahraminasab A, Sadegh Movahed M, Rahvar S, Sreenivasan KR, Reza Rahimi Tabar M (2006) Characteristic Angular Scales of Cosmic Microwave Background Radiation. J Stat Mech P11008

76. Nawroth AP, Peinke J (2006) Multiscale reconstruction of time series. Phys Lett A 360:234

77. Ghasemi F, Peinke J, Sahimi M, Reza Rahimi Tabar M (2005) Regeneration of Stochastic Processes: An Inverse Method. Eur Phys J B 47:411

78. Kolmogorov AN (1941) Dissipation of energy in locally isotropic turbulence. Dokl Akad Nauk SSSR 32:19

79. Kolmogorov AN (1962) A refinement of previous hypotheses concerning the local structure of turbulence in a viscous incompressible fluid at high Reynolds number. J Fluid Mech 13:82

80. Frisch U (1995) Turbulence. Cambridge University Press, Cambridge

81. Sreenivasan KR, Antonia RA (1997) The phenomenology of small-scale turbulence. Annu Rev Fluid Mech 29:435–472

82. Ghashghaie S, Breymann W, Peinke J, Talkner P, Dodge Y (1996) Turbulent Cascades in Foreign Exchange Markets. Nature 381:767–770

83. Nawroth AP, Peinke J (2006) Small scale behavior of financial data. Eur Phys J B 50:147

84. Karth M, Peinke J (2002) Stochastic modelling of fat-tailed probabilities of foreign exchange rates. Complexity 8:34

85. Bouchaud JP, Potters M, Meyer M (2000) Apparent multifractality in financial time series. Eur Phys J B 13:595–599

86. Bouchaud JP (2001) Power laws in economics and finance: some ideas from physics Quant Finance 1:105–112

87. Mandelbrot BB (2001) Scaling in financial prices: I. Tails and dependence. II. Multifractals and the star equation. Quant Finance 1:113–130

88. Embrechts P, Klüppelberg C, Mikosch T (2003) Modelling extremal events. Springer, Berlin

89. Mantegna RN, Stanley HE (1995) Nature 376:46–49

90. McCauley J (2000) The Futility of Utility: how market dynamics marginalize Adam Smith. Physica A 285:506–538

91. Muzy JF, Sornette D, Delour J, Areneodo A (2001) Multifractal returns and hierarchical portfolio theory. Quant Finance 1: 131–148

92. Viscek T (1992) Fractal Growth Phenomena. World Scientific, Singapore

93. Renner C, Peinke J, Friedrich R (2000) Markov properties of high frequency exchange rate data. Int J Theor Appl Finance 3:415

94. Davoudi J, Reza Rahimi Tabar M (1999) Theoretical Model for Kramers-Moyal's description of Turbulence Cascade. Phys Rev Lett 82:1680

95. Renner C, Peinke J, Friedrich R, Chanal O, Chabaud B (2002) Universality of small scale turbulence. Phys Rev Lett 89: 124502

96. Siefert M, Peinke J (2004) Different cascade speeds for longitudinal and transverse velocity increments of small-scale turbulence. Phys Rev E 70:015302R

Fluctuation Theorems, Brownian Motors and Thermodynamics of Small Systems

FELIX RITORT
Department de Fisica Fonamental, Faculty of Physics, Universitat de Barcelona, Barcelona, Spain

Article Outline

Glossary

Trajectory or path A crucial concept in the statistical description of small system is that of a trajectory or path. A path is the time sequence of configurations followed by the system as it is driven to a non-equilibrium state by the action of an external perturbation.

Control parameter External perturbations are usually described in terms of the control parameter λ. These are a set of external parameters (e. g. an electric field, magnetic field, optical force, ...) that can be experimentally controlled and do not fluctuate. Experimentally, control parameters are produced by macroscopic systems that are used to manipulate the small system under study and which are insensitive to thermal fluctuations (but that produce other sorts of uncontrolled instrumental noises and drift effects).

Single molecule experiments (SME) Recent technological developments have provided the tools to design and build scientific instruments of high enough sensitivity and precision to manipulate and visualize individual molecules and measure microscopic forces. Using SME it is possible to manipulate molecules one at a time and measure distributions describing molecular properties, characterize the kinetics of bio-molecular reactions, and detect molecular intermediates. SME provide the additional information about thermodynamics and kinetics of bio-molecular processes. This complements information obtained in traditional bulk assays. In SME it is also possible to measure small energies and detect large Brownian deviations in

bio-molecular reactions, thereby offering new methods and systems to scrutinize the basic foundations of statistical mechanics. Common single molecule experimental techniques are: atomic-force microscopy, laser optical tweezers, magnetic tweezers and single-molecule fluorescence.

Free energy The natural or spontaneous evolution of any thermodynamic process is determined by the free energy. The free energy in thermodynamics is the equivalent of the mechanical energy in classical mechanics. Spontaneous transformations take place by a decrease of the free energy in the system. In addition, mechanical work must be exerted by an external agent upon the system to increase its free energy. For reversible processes the amount of work is equal to the free energy change. However, in general, processes are irreversible and the work must be always larger than the free energy difference (a statement of the second law of thermodynamics). Free energies in small systems are typically expressed in either work (pN·nm) or energy units (kJ/mol, kcal/mol or $k_B T$ where k_B is the Boltzmann constant and T is a reference temperature – usually 298 K or 25 degrees Celsius). The conversion factors are ($T = 298$ K): $1\,k_B T = 4.11\,\text{pN} \cdot \text{nm} = 4.11 10^{-21}$ J, $1\,k_B T = 0.6\,\text{kcal/mol} = 2.4\,\text{kJ/mol}$.

ATP Acronym for adenosine triphosphate, the molecule that carries the energy necessary to sustain life processes. ATP is made of one adenosine base weakly bonded to three phosphate groups. Upon conversion (by hydrolysis) to ADP (adenosine diphosphate) and inorganic phosphate or AMP (adenosine monophosphate) and pirophosphate (P-P), ATP delivers a considerable amount of free energy (in the range 8–12 kcal/mol, depending on buffer conditions). By coupling to other reactions, ATP hydrolysis supplies the energy necessary to carry out unfavorable transformations.

RNA RNA (ribonucleic acid) is a very important player in molecular biology that shows biological functions in between those attributed to DNA and proteins. For the biophysicist and the statistical physicist RNA is also a fascinating molecule. Primarily found in nature in single stranded form, RNA folds into a three dimensional structure mainly stabilized by stacking interactions and hydrogen bonds between complementary bases (A-U,G-C). Full complementarity between different RNA segments is often impossible so, at difference with DNA, RNA structure includes also mismatches between bases as well other structural defects (bulges, loops, junctions, …). In addition to Watson–Crick base pairing, RNA forms a com-

pact structure through specific interactions mediated by magnesium ions that bring together distal RNA segments.

Definition of the Subject

The thermodynamics of small systems describes energy exchange processes between a system and its environment in the low energy range of a few $k_B T$ where Brownian fluctuations are dominant [1]. The main goal of this discipline is to identify the building blocks of a general theory describing energy fluctuations in non-equilibrium processes occurring in systems ranging from condensed matter physics to biophysics.

Thermodynamics, a scientific discipline inherited from the 18th century, is facing new challenges in the description of non-equilibrium small (sometimes also called mesoscopic) systems. Thermodynamics is a discipline built in order to explain and interpret energetic processes occurring in macroscopic systems made out of a large number of molecules on the order of the Avogadro number. The subsequent development of statistical mechanics has provided a solid probabilistic basis to thermodynamics and increased its predictive power at the same time. The development of statistical mechanics goes together with the establishment of the molecular hypothesis. Matter is made out of interacting molecules in motion. Heat, energy and work are measurable quantities that depend on the motion of molecules. The laws of thermodynamics operate at all scales.

However, thermodynamics, a science inherited in the 18th century from the times of the industrial revolution, has been inspired by motors and steam engines that proved to be indispensable during that time. It is fair then to question the relevance and applicability of all this knowledge when scientists immerse into the realm of the very small, far from the initial context that inspired Carnot and others.

What are the novel features of thermodynamics when applied to small (also called mesoscopic) systems? Is it necessary to revisit some of the main concepts that we learned from standard thermodynamics? How are energy dissipation and efficiency related for non-equilibrium small systems where energy fluctuations are dominant? Finally, what are the implications in quantum systems already governed by quantum fluctuations? Answering such questions is one of the main goals of this new discipline.

Introduction

The *non-equilibrium thermodynamics of small systems* is becoming quite popular among statistical physicists who

recognize there new aspects of thermodynamics where large Brownian fluctuations are of pivotal importance as compared to fluctuations in macroscopic (or large) systems. In macroscopic systems, fluctuations represent just small deviations respect to the average behavior. For example, an ideal gas of N molecules in thermal contact with a bath at temperature T has an average total kinetic energy of $(3/2)Nk_BT$. However, the total energy is not a conserved quantity but fluctuates, its spectrum being a Gaussian distribution of variance $(3/2)N(k_BT)^2$ according to the law of equipartition. Therefore, relative deviations of the energy are on the order $1/\sqrt{N}$ respect to the average value. For macroscopic systems such deviations are very small: for $N = 10^{12}$ (this is the typical number of molecules in a 1ml test tube at nanomolar concentration) relative deviations are on the order of 10^{-6}, hence experimentally unobservable by calorimetry methods. For a few molecules, $N \sim \mathcal{O}(1)$, relative deviations are on the same order. Fluctuations are then measurable by direct observation of individual molecules.

Small systems share the property that energy fluctuations are much larger than $\sim \sqrt{E}$ (the prediction by the law of large numbers) where E is the average total energy. Large deviations from average values are normally observed in mesoscopic systems where non-equilibrium fluctuations are governed by a few degrees of freedom. Examples abound in physics and biology: the Brownian motion of a micron-size silica bead captured in an optical trap; the unfolding of a bio-molecule (e. g. a nucleic acid hairpin or a protein); the movement of molecular motors inside the cell; the cooperative rearrangement of a nanosized region containing a few molecules inside an amorphous material such as a glass.

As a rule of thumb we can say that small systems are those where the typical energy content of the system is a few times k_BT, maybe from 1 to 1000 but not much more. As often happens, there is no well defined frontier separating the small-size regime from the large-size regime. The name thermodynamics of small systems was first coined by T. L. Hill [2] who showed the importance of the statistical ensemble in thermodynamic relations. A main result of statistical mechanics is the independence of the equation of state on the statistical ensemble in the thermodynamic limit. Such independence breaks down in small systems due to the contribution of fluctuations which depend on the type of statistical ensemble considered. In biology, the most important aspect of these tiny machines is that they operate far from equilibrium; its consequences and importance in their biological function are still unknown. The combination of small size and non-equilibrium behavior is the playground for the strik-

ing behavior observed in condensed matter physics and biophysics.

Prominent in the field is the study of the so called work and heat fluctuations in systems driven to a non-equilibrium state. Fluctuation theorems are mathematical relations that quantify the relative probability of trajectories that release and absorb a given amount of work/heat to and from the environment. Taken individually, the work and heat along these trajectories can violate some of the inequalities of thermodynamics, leading to what is commonly referred as *transient violations of the second law*. This name has raised strong objections among some groups of physicists. Of course the second law remains inviolate. The name just stresses the fact that Brownian fluctuations are big enough for such deviations from the average value to be observed. For macroscopic systems these trajectories are known to be irrelevant and unobservable, however at the level of small system sizes, when the energies involved are of order of several times k_BT, these trajectories become important. Although thermodynamic inequalities are known to describe the behavior of average values, it is important to explore the implications and relevance of these deviations in our understanding of energy transformation processes at the molecular level.

The quantitative experimental observation and measurement of large energy fluctuations has become possible only recently with the development of new micro-manipulation tools. Particularly important are the application of single-molecule techniques to explore physical theories in systems out of equilibrium. The use of new micro-manipulation tools in the exploration of the behavior of tiny objects (such as bio-molecules and motors) embedded in a thermal environment opens the possibility to investigate how these systems exchange energy with their environment. This question is of great interest both at a fundamental and practical level. From a fundamental point of view, the comprehension of how bio-molecules operating very far from equilibrium are so efficient raises the question whether such tiny systems exploit rare and large deviations from their average behavior by rectifying thermal fluctuations from the bath. From a practical point of view, this might help in the design of efficient nanomotors in the future.

In the next sections we overview a selection of topics in this fascinating area of research. We will start by discussing molecular motors in Sect. "Molecular Motors". In Sect. "Non-equilibrium Thermodynamics of Small Systems" we explain what are the main types of non-equilibrium states and introduce the microscopic definitions of heat and work. In Subsect. "The Jarzynski Equality" we

derive the Jarzynski equality by using the master equation approach. This part is a bit technical and the reader who is not interested in the details can jump to Eq. (17) and continue reading from there. Finally, in Sect. "Fluctuation Theorems" we discuss fluctuation theorems and review some of their applications to condensed matter physics and biophysics. The article finishes with a discussion of future directions in research.

Molecular Motors

Molecular motors are proteins that use the energy extracted from the hydrolysis of ATP to exert mechanical work (Fig. 1) [3]. The mechanism by which motors utilize the energy stored in the high energy bonds of the ATP molecules to perform mechanical work is based on two hypothesized mechanisms: 1) Power stroke generation or 2) Brownian ratchet mechanism. In the first mechanism the release of the pyrophosphate during the ATP hydrolysis cycle is tightly coupled to the generation of force which

drives the motor. In the second mechanism, the motor diffuses reversibly along the substrate. Movement is then produced by the hydrolysis of ATP that induces a conformational change in the protein. This change is then rectified by thermal fluctuations that induce the release of ADP. By steady repetition of this mechanochemical cycle (one ATP molecule is hydrolyzed per cycle) the motor carries out important cellular functions. Motors are characterized by the so called processivity or number of turnover cycles the motor does until detaching from the substrate. Processivities of molecular motors can vary by several orders of magnitude depending on the type of motor and the presence of other regulating factors. For example, the muscle myosin II motors work in large assemblies, each myosin having a processivity around 1, meaning that each myosin performs one mechanochemical cycle on average before detaching from the substrate. In the other extreme of the scale there are DNA polymerases in eukaryotes which show processivities that range from 1 (adding approximately one nucleotide before detaching) up to sev-
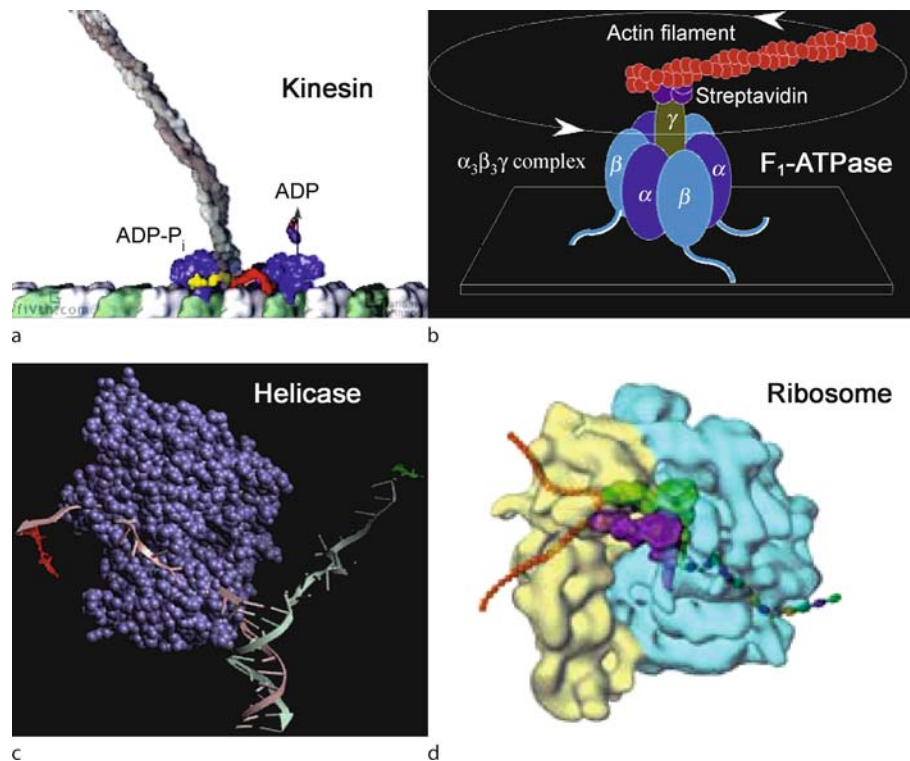
Fluctuation Theorems, Brownian Motors and Thermodynamics of Small Systems, Figure 1
Examples of molecular machines: **a** Kinesin walking a long micro-tubule and transporting a cargo. **b** F_1-ATP synthase is the proton pump responsible of producing ATP in the mitochondria of eukaryotic cells. **c** Helicases are forerunners of the DNA polymerase that unwind DNA by transforming dsDNA into two strands of ssDNA. **d** The ribosome is one among the largest molecular machines inside the cytoplasm of the cell in charge of manufacturing proteins

eral thousands or even millions. However, in the presence of sliding clamps (proteins with the shape of a doughnut that encircle the DNA and tightly bind DNA polymerases) processivities go up to 10^9. Molecular motors are magnificent objects from the point of view of their efficiency. If we define the efficiency rate as the ratio between the useful work performed by the motor and the energy released in the hydrolysis of one ATP molecule in one mechanochemical cycle, then typical values for the efficiencies are around several tens per cent, reaching the value of 90% in some cases (like in the rotary motor F_1-ATPase). For example, out of the $20\,k_\mathrm{B}T$ obtained from the hydrolysis of one molecule of ATP, kinesin exerts a mechanical work of $12\,k_\mathrm{B}T$ at every step, having an efficiency of around 60%. Such large efficiencies are rarely found in macroscopic systems (motors of cars have efficiencies below 20%) meaning that molecular motors have been designed by evolution to efficiently operate in a highly noisy environment. Molecular motors are expected to be essential constituents of future nanodevices.

What is the relation between molecular motors and the non-equilibrium thermodynamics of small systems? It is a well established fact that the typical amounts of energy obtained from chemical sources (e. g. ATP or GTP hydrolysis) used by most molecular machines are a few kcal per mol (at $T \sim 300\,\mathrm{K}$ this corresponds to a few units of $k_\mathrm{B}T$, $1\,k_\mathrm{B}T \simeq 0.6\,\mathrm{kcal/mol}$). Let us consider the example of RNA transcription. The process by which RNA nucleotides (A,U,G,C) are added to the newly synthesized RNA strand during the transcription process involves the hydrolysis of the different nucleoside-phosphate complexes as they are added to the 3' end of the growing chain. The overall process by which one base is added to the newly synthesized strand is a highly favorable reaction (mainly driven by the hydrolysis of the pyrophosphate) with a free energy release, ΔG, in the range between 7 and 12 kcal/mol mainly depending on the magnesium concentration in the environment. Effectively this is an irreversible process that generates an amount of available energy between 10 and 20 $k_\mathrm{B}T$ at room temperature ($\sim 300\,\mathrm{K}$, $1k_\mathrm{B}T \simeq 0.6\,\mathrm{kcal/mol}$) per base pair added. This energy would be lost to the environment in the form of heat were it not for the fact that a big part of the energy is used by the RNA polymerase to locally unwind the double DNA helix and pull apart the two DNA strands to produce a bubble a few bases long of denatured DNA. This bubble is then used by the DNA/RNA/polymerase ternary complex as a substrate to polymerize the RNA. As transcription proceeds the bubble moves downstream together with the RNA polymerase and the RNA transcript is synthesized.

For this process to occur, the RNA polymerase must move against the Stokes friction produced by water as well as other roadblocks that hamper its motion. In particular, the RNA polymerase must exert force and torque on the DNA. Typical forces to unzip DNA are on the order of 15 pN meaning that the minimum mechanical work necessary to unzip one base pair is around 15 pN times 12 Angstroms (the typical extension gained after pulling apart two bases at the fork of a DNA hairpin), which is equal to $18\,\mathrm{pN \cdot nm}$ or equivalently $4.4\,k_\mathrm{B}T$ ($1\,k_\mathrm{B}T \simeq 4.11\mathrm{pN \cdot nm}$ at room temperature). We can define the efficiency of the RNA motor as the ratio between the mechanical work needed to unzip one base pair and the amount of energy obtained from hydrolysis upon the addition of a nucleotide (however this is not the only way to define mechanical efficiencies, e. g. see [4,5]). The efficiency of the transcription process is then about 40 per cent, a quite remarkable feat if we compare this number with the ones obtained in man made machines (cars have efficiencies below 20 per cent). The motion of a single RNA polymerase has been studied in several prokaryotic systems using optical and magnetic tweezers [6,7]. In these experiments a DNA/polymerase complex is tethered between a trapped bead and an streptavidin coated immobilized bead or surface. To initiate transcription nucleotides are allowed to flow inside the chamber and the elongation of the transcript can be followed in real time while force is applied on the tether. The extension of the RNA transcript as a function of time reveals a complex intermittent motion of the polymerase with pauses (temporary stops), arrests (permanent stops) and even backtracking events [8,9].

Non-equilibrium Thermodynamics of Small Systems

Non-equilibrium States

An important concept in thermodynamics is the state variable. State variables are those that, once determined, uniquely specify the thermodynamic state of the system. Examples are the temperature, the pressure, the volume and the mass of the different components in a given system. To specify the state variables of a system it is common to put the system in contact with a bath. The bath is any set of sources (of energy, volume, mass, etc.) large enough to remain unaffected by the interaction with the system under study. The bath ensures that a system can reach a given temperature, pressure, volume and mass concentrations of the different components when put in thermal contact with the bath (i. e. with all the relevant sources). Equilibrium states are then generated by putting the system in contact with a bath and waiting until the system proper-

ties relax to the equilibrium values. Under such conditions the system properties do not change with time and the average heat/work/mass exchanged between the system and the bath is zero.

Non-equilibrium states can be produced in many ways, either by continuously changing the parameters of the bath or by preparing the system in an initial non-equilibrium state that slowly relaxes toward equilibrium. In general a non-equilibrium state is produced whenever the system properties change with time and/or the net heat/work/mass exchanged by the system and the bath is non zero. We can distinguish at least three different classes of non-equilibrium states:

- **Non-equilibrium transient state (NETS)** The system is initially prepared in an equilibrium state and later driven out of equilibrium by switching on an external perturbation. The system quickly returns to a new equilibrium state once the external perturbation stops changing.

  A classic example of NETS is the case of a protein in its initial native state that is mechanically pulled (e. g. using AFM) by exerting force on the ends of the molecule. The protein is initially folded and in thermal equilibrium with the surrounding aqueous solvent. Upon pulling the protein is driven away from equilibrium into a transient state until it finally settles into the unfolded and extended equilibrium state. Another example of a NETS is a bead immersed in water and trapped in an optical well generated by a focused laser beam. When the trap is moved to a nearby new position (e. g. by moving the laser beams) the bead is driven to a NETS. After some time the bead reaches equilibrium again at the new position of the trap. In another experiment the trap is suddenly put in motion at a speed v so the bead is transiently driven away from its equilibrium average position until it settles into a non-equilibrium steady-state (NESS, see below) characterized by the speed of the trap. The average position of the bead lags behind the position of the center of the trap.

- **Non-equilibrium steady-state (NESS)** The system is driven by external forces (either time dependent or non-conservative) in a stationary non-equilibrium state where its properties do not change with time. The steady state cannot be described by the Boltzmann–Gibbs distribution and the average net heat that is dissipated by the system (equal to the entropy production of the bath) is positive.

  A classic example of a NESS is an electrical circuit made out of a battery and a resistance. The current flows through the resistance and the chemical energy stored in the battery is dissipated to the environment in the form of heat; the average dissipated power, $\mathcal{P}_{\mathrm{dis}} = VI$, is equal to the power supplied by the battery. Another example is a sheared fluid between two plates or coverslips and one of them is moved relative to the other at a constant velocity v. To sustain such state a mechanical power equal to $\mathcal{P} \propto \eta v^2$ has to be exerted upon the moving plate where η is the viscosity of water. The mechanical work produced is then dissipated in the form of heat through the viscous friction between contiguous fluid layers. Further examples of NESS are chemical reactions in metabolic pathways that are sustained by activated carrier molecules such as ATP. In such cases, hydrolysis of ATP is strongly coupled to specific oxidative reactions. For example, ionic channels use ATP hydrolysis to transport protons against the electromotive force.

- **Non-equilibrium aging state (NEAS)** The system is initially prepared in a non-equilibrium state and put in contact with the sources. However, at difference with NETS, the system fails to reach thermal equilibrium in observable or laboratory time scales. In this case the system is in a non-stationary slowly relaxing non-equilibrium state called *aging state* and characterized by a very small entropy production of the sources. In the aging state two-time correlations decay slower as the system becomes older. Two-time correlation functions depend on both times and not just on their difference. The classic example of a NEAS is a supercooled liquid cooled below its glass transition temperature [10]. The liquid solidifies into an amorphous slowly relaxing state characterized by huge relaxational times and anomalous low frequency response. Other systems are colloids that can be prepared in a NEAS by the sudden reduction/increase of the volume fraction of the colloidal particles or by putting the system under a strain/stress [11].

The classes of non-equilibrium states previously described do not make distinctions whether the system is macroscopic or small. In small systems, however, it is common to speak about the control parameter to emphasize the importance of the constraints imposed by the bath that are externally controlled and do not fluctuate. The control parameter (λ) represents a value (in general, a set of values) that defines the state of the bath. Its value determines the equilibrium properties of the system, e. g. the equation of state. In macroscopic systems it is unnecessary to discern which value is externally controlled because fluctuations are small and all equilibrium ensembles give the same equivalent thermodynamic description, i. e. the same

equation of state. Differences arise when taking into account fluctuations. The non-equilibrium behavior of small systems is then strongly dependent on the specific non-equilibrium protocol. Figure 2 shows a representation of a few examples of NESS and control parameters.

Microscopic Definitions of Work and Heat

Microscopic definitions of work and heat can be given using Markov processes. Let us consider a general system described by an energy function $E(C)$ where C is a generic configuration in contact with a bath at temperature T. For instance, in a gas of N molecules, C would stand for their positions and momenta. The dynamics are assumed to be discrete in time with elementary time-step Δt. A trajectory or path of the system is characterized by the sequence of configurations

$$\Gamma \equiv \{C_k; 0 \leq k \leq M\} \equiv \{C_0, C_1, C_2, \ldots, C_M\} \quad (1)$$

where k is the index for the discrete time step and M is the total number of time steps. The time corresponding to step k is then given by $t = k\Delta t$ with $t = 0 \, (k = 0)$ and $t_f \, (k = M/\Delta t)$ denoting the initial and final times respec-

tively. The continuous-time limit is recovered by taking $\Delta t \to 0, M \to \infty$ with t_f finite.

Now we will treat the case where the system is perturbed in a prescribed way. Because the dynamics is stochastic, it will generate an ensemble of trajectories when the same experiment is repeated many times. In addition to the configuration C, and in order to characterize the perturbation protocol, we need to specify the temporal sequence of values $\{\lambda_k; 0 \leq k \leq M\}$. The control parameter λ shifts the energy levels of the system according to the relation,

$$E_\lambda(C) = E(C) - \lambda A(C) \quad (2)$$

where $A(C)$ is the observable coupled to the external perturbation λ (e. g., if λ is a magnetic or gravitational field then A stands for the magnetization or the height of the center of mass respectively).

We now consider the variation of energy along a given path $\Delta E(\Gamma) = E_{\lambda_f}(C_f) - E_{\lambda_0}(C_0)$ where C_0, C_f are the initial and final configurations for that path and λ_0, λ_f are the initial and final values of the control parameter as defined by the protocol (path independent). From Eq. (2) the
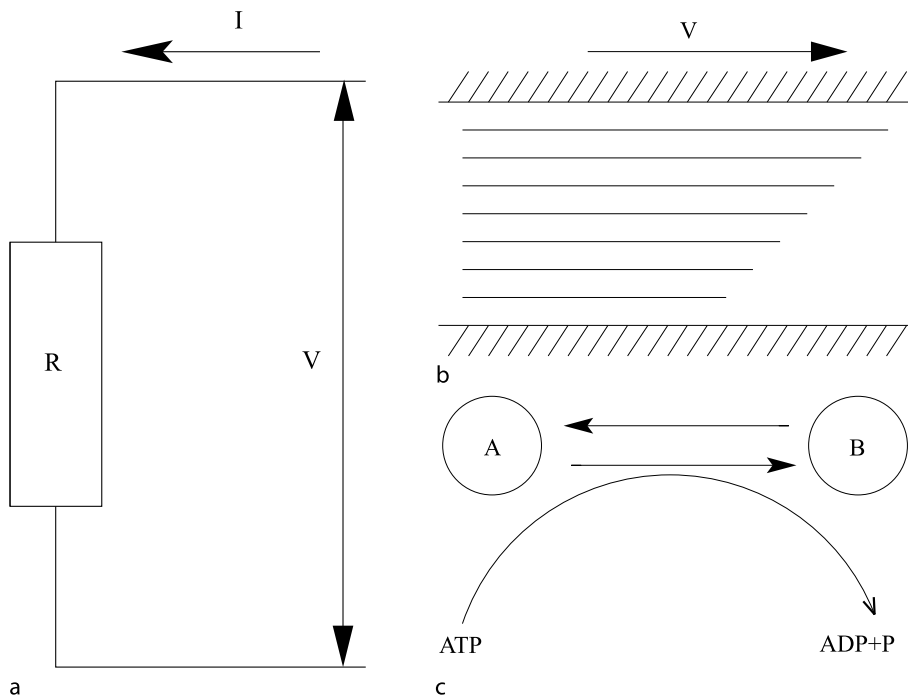
Fluctuation Theorems, Brownian Motors and Thermodynamics of Small Systems, Figure 2
Examples of NESS: **a** An electric current *I* flowing through a resistance *R* and maintained by a voltage source or control parameter *V*. **b** A fluid sheared between two plates that move at speed *v* (the control parameter) relative to each other. **c** A chemical reaction $A \to B$ coupled to ATP hydrolysis. The control parameters are the concentrations of ATP and ADP

energy variation is given by,

$$\Delta E(\Gamma) = E_{\lambda_M}(C_M) - E_{\lambda_0}(C_0)$$

$$= \sum_{k=0}^{M-1} \left(E_{\lambda_{k+1}}(C_{k+1}) - E_{\lambda_k}(C_{k+1}) \right)$$

$$- \sum_{k=0}^{M-1} \left(E_{\lambda_k}(C_k) - E_{\lambda_k}(C_{k+1}) \right) \quad (3)$$

with $\Delta\lambda_k = \lambda_{k+1} - \lambda_k$. This decomposition identifies work and heat by using the first law of thermodynamics, $\Delta E = W - Q$. The first term in Eq. (3) is identified as the work W exerted upon the system whereas the second corresponds to the heat Q transferred from the system to the bath,

$$W(\Gamma) = \sum_{k=0}^{M-1} \left(E_{\lambda_{k+1}}(C_{k+1}) - E_{\lambda_k}(C_{k+1}) \right) \quad (4)$$

$$Q(\Gamma) = \sum_{k=0}^{M-1} \left(E_{\lambda_k}(C_k) - E_{\lambda_k}(C_{k+1}) \right) . \quad (5)$$

We concentrate our attention on the work exerted upon the system along a given path Γ. Inserting (2) in (4) we get

$$W(\Gamma) = \sum_{k=0}^{M-1} \left(\frac{\partial E_\lambda(C_k)}{\partial\lambda} \right)_{\lambda=\lambda_k} \Delta\lambda_k$$

$$= -\sum_{k=0}^{M-1} A(C_k)\Delta\lambda_k \equiv -\int_0^t \mathrm{d}s\,\dot\lambda(s)A(C(s))\mathrm{d}s$$

$$\quad (6)$$

where we have applied the continuous-time limit in the last term in the r.h.s. of (6).

As the path is stochastic the work is a fluctuating quantity that can be characterized by its probability distribution $\mathcal{P}(W)$ defined as,

$$\mathcal{P}(W) = \sum_\Gamma P(\Gamma)\delta(W - W(\Gamma)) \quad (7)$$

where Γ stands for the path and $P(\Gamma)$ indicates the probability of that path. The importance of $\mathcal{P}(W)$ relies upon the fact that it is a quantity that is experimentally measurable and therefore is suitable to quantitatively characterize work fluctuations with the aid of recently developed micro-manipulation tools.

The Jarzynski Equality

In 1997 Chris Jarzynski derived a remarkable equality describing work fluctuations in non-equilibrium isothermal systems [12]. This relation was somewhat unexpected because it related the free energy change in a reversible process with exponential averages of the work measured in irreversible processes. The equality applies to all non-equilibrium systems under general assumptions of local detailed balance and ergodicity. In what follows we show a derivation of the equality based on a master equation approach. The following calculation intends to show the simplicity of the algebraic math used in the derivation of this general result. The reader not interested in math details can jump directly to Eq. (17) and continue reading from there.

Let us consider a system in contact with a thermal bath at temperature T that is initially in thermal equilibrium. The system is then driven to a NETS under the action of an external perturbation described by the temporal sequence of values $\{\lambda_k; 0 \leq k \leq M\}$.

We consider the ensemble of all possible trajectories that start from an initial state characterized by the distribution $P_{\lambda_0}(C)$. Dynamics of the system are then given by the set of probabilities $P_{\lambda_k}(C)$ for the system to be found at configuration C at time-step k. These probabilities satisfy a master equation. For a Markov process the time evolution of the $P_\lambda(C)$ depends on the quantities, $\mathcal{W}_{\lambda_k}(C \to C')$, defined as the transition probability per unit time to go from configuration C to C' at time-step k. The $\mathcal{W}$'s are assumed to lead to an ergodic dynamics (where any pair of configurations are always connected by at least one trajectory, i. e. dynamics is *irreducible*) and satisfy the detailed balance condition,

$$\frac{\mathcal{W}_{\lambda_k}(C \to C')}{\mathcal{W}_{\lambda_k}(C' \to C)} = \frac{P_{\lambda_k}^{\mathrm{eq}}(C')}{P_{\lambda_k}^{\mathrm{eq}}(C)}$$

$$= \exp\left(-\beta(E_{\lambda_k}(C') - E_{\lambda_k}(C))\right) \quad (8)$$

where $\beta = 1/k_B T$, k_B is the Boltzmann constant and T is the temperature. The $P_{\lambda_k}^{\mathrm{eq}}(C)$ is the Boltzmann–Gibbs distribution,

$$P_\lambda^{\mathrm{eq}}(C) = \exp(-\beta E_\lambda(C_0))/Z_\lambda ;$$

$$Z_\lambda = \sum_C \exp(\beta E_\lambda(C)) \quad (9)$$

where $Z_\lambda = \exp(-\beta F_\lambda)$ is the partition function and F_λ is the free energy.

The energy function $E_{\lambda_k}(C)$ is given in (2). Under very general conditions these dynamics guarantee that the system reaches a stationary state where configurations are populated according to the Boltzmann–Gibbs weight. The solution to the master equation gives the time evolution for the system.

For a generic path-dependent observable $\mathcal{A}(\Gamma)$, the ensemble average value is given by,

$$\langle \mathcal{A} \rangle = \sum_{\Gamma} P(\Gamma)\mathcal{A}(\Gamma) . \tag{10}$$

Using the fact that the dynamics are Markovian together with the definition (1) we can write,

$$P(\Gamma) = P_{\lambda_0}^{\text{eq}}(C_0) \prod_{k=0}^{M-1} \mathcal{W}_{\lambda_k}(C_k \to C_{k+1}) \tag{11}$$

where the system initially starts in equilibrium at λ_0. By inserting (11) into (10) we obtain,

$$\langle \mathcal{A} \rangle = \sum_{\Gamma} \mathcal{A}(\Gamma)P_{\lambda_0}^{\text{eq}}(C_0) \prod_{k=0}^{M-1} \mathcal{W}_{\lambda_k}(C_k \to C_{k+1}) . \tag{12}$$

Using the detailed balance condition (8) this expression reduces to,

$$\langle \mathcal{A} \rangle = \sum_{\Gamma} P_{\lambda_0}^{\text{eq}}(C_0)\mathcal{A}(\Gamma) \prod_{k=0}^{M-1} \Bigg[\mathcal{W}_{\lambda_k}(C_{k+1} \to C_k)$$
$$\exp\left[-\beta(E_{\lambda_k}(C_{k+1}) - E_{\lambda_k}(C_k))\right] \Bigg] \tag{13}$$

$$= \sum_{\Gamma} \mathcal{A}(\Gamma)P_{\lambda_0}^{\text{eq}}(C_0)$$
$$\exp\left[-\beta \sum_{k=0}^{M-1} \left(E_{\lambda_k}(C_{k+1}) - E_{\lambda_k}(C_k)\right)\right]$$
$$\prod_{k=0}^{M-1} \mathcal{W}_{\lambda_k}(C_{k+1} \to C_k) . \tag{14}$$

This equation can not be worked out further for a general observable $\mathcal{A}$. However, let us consider the observable, $\mathcal{A}(\Gamma) = \exp(-W(\Gamma))$, where $W(\Gamma)$ stands for the work defined in (4). By inserting this expression in (14) we obtain the Jarzynski equaltity:

$$\langle \exp(-W) \rangle$$
$$= \frac{1}{Z_{\lambda_0}} \sum_{\Gamma} \prod_{k=0}^{M-1} \mathcal{W}_{\lambda_k}(C_{k+1} \to C_k) \exp(-\beta E_{\lambda_M}(C_0))$$
$$= \frac{Z_{\lambda_M}}{Z_{\lambda_0}} = \exp\left(-\beta F_{\lambda_M} - F_{\lambda_0}\right) = \exp(-\beta\Delta F)$$
$$\tag{15}$$

where we have applied a telescopic sum and used (9). To carry out the telescopic sums we first summed over

$C_0, C_1 \ldots$ by applying the normalization condition on the transition probabilities,

$$\sum_{C'} \mathcal{W}_k(C \to C') = 1 . \tag{16}$$

The second law of thermodynamics, $\langle W \rangle \geq \Delta F$, also follows naturally as a particular case of (15) by using the convexity inequality, $\langle \exp(x) \rangle \geq \exp\langle x \rangle$. The Jarzynski equality is often written in the form,

$$\langle \exp(-W_{\text{diss}}) \rangle = 1 \tag{17}$$

where $W_{\text{diss}} = W - \Delta F$ is called the dissipated work. The second law of thermodynamics puts bounds on the minimum amount of average work performed on the system: although W may strongly fluctuate from path to path its mean value (averaged over an infinite number of repeated experiments, i. e. the first moment of $P(W)$) is always greater than the reversible or quasi-static work, W_{rev}, which is also equal to the free energy difference ΔF between the initial and final equilibrium states. The reversible work is the value of the work that is obtained for protocols that are adiabatic or quasi-static, i. e. the control parameters are changed infinitely slowly. The difference between the actual work and the reversible work then corresponds to the dissipated work, $W_{\text{diss}} = W - \Delta F$. The second law establishes that in average a positive amount of heat is irreversibly lost to the environment, $\langle W_{\text{diss}} \rangle \geq 0$ (Fig. 3). The amount of dissipation in irreversible processes is then related to the asymmetry between the phase space densities obtained when the process is run forward and backward in time [13].

Gaussian work distributions exactly satisfy (17) provided the average dissipated work $\langle W_{\text{diss}} \rangle$ and the variance of the work, σ_W^2, are related by $\sigma_W^2 = 2k_{\text{B}}T\langle W_{\text{diss}} \rangle$. A fluctuation-dissipation parameter R can be introduced to quantify deviations from the Gaussian behavior [14],

$$R = \frac{\sigma_W^2}{2k_{\text{B}}T\langle W_{\text{diss}} \rangle} . \tag{18}$$

For Gaussian work distributions $R = 1$, corresponding to non-equilibrium processes in the linear regime where the fluctuation-dissipation theorem holds.

The validity of the Jarzynski equality extends to deterministic dynamics (e. g. Hamiltonian or thermostated). In Hamiltonian dynamics the set of phase space points then behaves as an incompressible fluid, a consequence of the Liouville theorem. The case of Hamiltonian dynamics was originally addressed by Jarzynski in his original derivation of the non-equilibrium work relation [12]. The
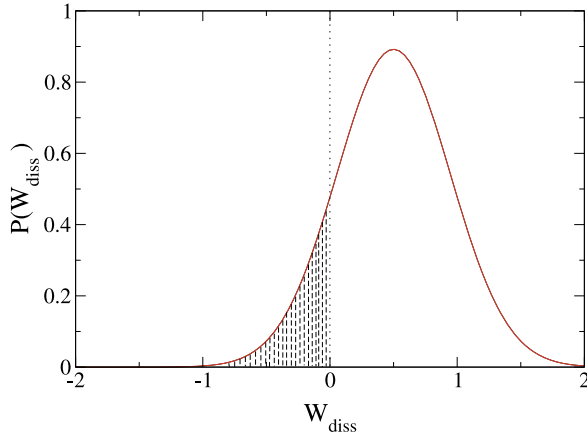
Fluctuation Theorems, Brownian Motors and Thermodynamics of Small Systems, Figure 3
Probability distribution of the dissipated work: According to the second law of thermodynamics the average dissipated work is always positive. However, because of fluctuations, the dissipated work of some paths can be negative (*shaded area*). These paths are sometimes referred to as "transient violations of the second law"

stochastic case has been analyzed also for general Markov processes by Crooks [15,16] and for Langevin dynamics by Kurchan [17] and Seifert [18]. Equation (15) has appeared in the past in the literature in the form of a generalized fluctuation-dissipation relation proposed by Bochkov and Kuzovlev [19] which is mathematically identical to the Jarzynski equality [12]. Related results to the Jarzynski's equality can be also traced back also in the free-energy perturbation identity derived by Zwanzig [20] and the Kirkwood formula [21].

Fluctuation Theorems

Since the beginning of the 90's some exact results in statistical mechanics have provided a mathematical description of energy fluctuations (in the form of heat and work) for non-equilibrium systems. This new class of results go under the name of fluctuation theorems (FTs) and provide a solid theoretical basis to quantify energy fluctuations in non-equilibrium systems. FTs describe energy fluctuations in systems while they execute transitions between different types of states. For these fluctuations to be observable the system has to be small enough and/or operate over short periods of time, otherwise the measured properties approach the macroscopic limit where fluctuations get masked by the dominant average behavior. Most fluctuation theorems are of the form,

$$\frac{P(+S)}{P(-S)} = \exp\left(\frac{S}{k_{\mathrm{B}}}\right) , \qquad (19)$$

where S has the dimensions of an entropy that may represent heat and/or work produced during a given time interval. The precise mathematical form of relations such as (19) (for instance, the precise definition of S or whether they are valid at finite time intervals or just in the limit where the time interval goes to infinity) depends on the particular non-equilibrium conditions (e. g. whether the systems starts in an equilibrium Gibbs state, or whether the system is in a non-equilibrium steady state, or whether the system executes transitions between steady states, etc.).

Generally speaking, FTs relate the amounts of work or heat exchanged between the system and its environment for a given non-equilibrium process and its corresponding time-reversed process. The time-reversed process is defined as follows. Let us consider a given non-equilibrium process (we call it forward, denoted by F) characterized by the protocol $\lambda_{\mathrm{F}}(t)$ of duration t_f. In the reverse process (denoted by R) the system starts at $t = 0$ in a stationary state at the value $\lambda_{\mathrm{F}}(t_f)$ and the control parameter is varied for the same duration t_f as in the forward process according to the protocol $\lambda_{\mathrm{R}}(t) = \lambda_{\mathrm{F}}(t_f - t)$. FTs depend on the type of initial state and the particular type of dynamics (deterministic versus stochastic) or thermostated conditions.

Despite of the fact that most of these theorems are treated as distinct they are in fact closely related. The main hypothesis for all theorems is the validity of some form of microscopic reversibility or local detailed balance (see however [22,23,24] for some controversy). Major classes of FTs include the transient FT (TFT) and the steady state FT (SSFT):

- The transient FT (TFT) In the TFT the system initially starts in an equilibrium (Boltzmann–Gibbs) state and is driven away from equilibrium by the action of time-dependent forces that derive from a time-dependent potential $V_{\lambda(t)}$. The potential depends on time through the value of the control parameter $\lambda(t)$. At any time during the process the system in an unknown transient non-equilibrium state. If the value of λ is kept fixed then the system relaxes into a new equilibrium state. The TFT was introduced by Evans and Searles [25] in thermos-tatted systems and later extended by Crooks to Markov processes [15].

- The steady state FT (SSFT) In the SSFT the system is in a non-equilibrium steady state where it exchanges net heat and work with the environment. The existence of the SSFT was numerically anticipated by Evans and collaborators for thermostated systems [26] and demonstrated for deterministic Anosov systems by Gallavotti and Cohen [27]. The entropy production S by the system (equal to the heat exchanged by the system divided

by the temperature of the environment) satisfies the relation (19) in the asymptotic limit of large times $t \to \infty$ and for bounded energy fluctuations, $\sigma = \frac{|S|}{t} < \sigma^*$ where σ^* is a model dependent quantity. Other classes of SSFTs include stochastic Langevin dynamics [17], Markov chains [28,29] or the case where the system initially starts in a steady state [30] and executes transitions between different steady states [31,32].

Particularly relevant to the single molecule context is the FT by Crooks [15,16] which relates the work distributions measured along the forward (F) and reverse (R) paths,

$$\frac{P_\mathrm{F}(W)}{P_\mathrm{R}(-W)} = \exp\left(\frac{W - \Delta F}{k_\mathrm{B} T}\right) , \qquad (20)$$

where $P_\mathrm{F}(W)$, $P_\mathrm{R}(-W)$ are the work distributions along the F and R processes respectively, and ΔF is the free energy difference between the equilibrium states corresponding to the final value of the control parameter $\lambda_f = \lambda(t_f)$ and the initial one $\lambda_i = \lambda(0)$. A particular result of (20) is the Jarzynski equality [12] described in Subsect. "The Jarzynski Equality" that is obtained from (20) by rewriting it as $P_\mathrm{R}(-W) = \exp\left(\frac{-W + \Delta G}{k_\mathrm{B} T}\right) P_\mathrm{F}(W)$ and integrating both sides of the equation between $W = -\infty$ and $W = \infty$. Because of the normalization property of probability distributions, the left hand side is equal to 1 and the Jarzynski equality reads,

$$\left\langle \exp\left(-\frac{W}{k_\mathrm{B} T}\right) \right\rangle_\mathrm{F} = \exp\left(-\frac{\Delta F}{k_\mathrm{B} T}\right) \text{ or}$$
$$\Delta F = -k_\mathrm{B} T \log\left(\left\langle \exp\left(-\frac{W}{k_\mathrm{B} T}\right) \right\rangle_\mathrm{F}\right) , \qquad (21)$$

where $< \dots >_\mathrm{F}$ denotes an average over an infinite number of paths, all generated by a given forward protocol $\lambda_\mathrm{F}(t)$.

Experimental Tests and Free Energy Recovery

Various categories of FTs have been introduced and experimentally validated. The first experimental tests of FTs were carried out by Ciliberto and coworkers for the Gallavoti–Cohen FT in Rayleigh–Bernard convection [33] and turbulent flows [34]. Later on FTs were tested for beads trapped in an optical potential and moved through water at low Reynolds numbers. The motion of the bead is then well described by a Langevin equation that includes a friction (non-conservative) force, a confining (conservative) potential and a source of stochastic noise. Experiments have been carried out by Evans and collaborators who have tested the validity of the TFT [35,36], and by

Liphardt and collaborators for a bead executing transitions between different steady states [37]. The validity of the TFT has been also recently tested for non-Gaussian optical trap potentials [38].

The Jarzynski equality and the FT by Crooks can be used to recover equilibrium free-energy differences between different molecular states by using non-equilibrium measurements in single molecule experiments [39,40,41]. In particular, by using the Jarzynski equality (21) it is possible to extract the value of ΔF from repeated measurements of the mechanical work along many trajectories. The idea is to repeat non-equilibrium experiments many times and evaluate the exponential average in the r.h.s of (21) to extract the work corresponding to the reversible process. Would it then not be easier to directly measure the work for a reversible process? Unfortunately many processes cannot be carried in quasistatic conditions (either simulations or experiments) and therefore, alternative methods are required to determine free-energy differences. There are practical difficulties in the applicability of (21) as the number of trajectories included in the exponential average must be actually infinite. This is unrealizable in practice as non-equilibrium experiments can be performed only a finite number of times and the finiteness of the number of trajectories introduces a bias. It is known that the number of trajectories required to evaluate the Jarzynski equality grows exponentially with the average value of the dissipated work. The dependence of the bias and error with the number of pulls has been estimated in some cases [42,43]. In general this dependence can be quite complicated as it depends on the behavior of the low-work tails of the distribution $P(W)$ which are difficult to analyze in general.

In 2002, the Jarzynski equality was experimentally tested by pulling the P5ab RNA hairpin, a derivative of the *Tetrahymena Termophila* L21 ribozyme [44] using optical tweezers. However, in that case the molecule was pulled not too far from equilibrium. The Jarzynski equality and related identities for athermal systems have been recently put under scrutiny in other systems [45,46,47]. The Jarzynski equality and the FT by Crooks have inspired several theoretical papers discussing other related exact results [48,49,50,51,52,53], free-energy recovery from numerical simulations [54,55,56,57,58,59], bias and error estimates for free-energy differences [42,43,60,61,62,63], enzyme kinetics [64,65] or solvable models [66,67,68,69,70]. In addition, analytical studies on small systems thermodynamics show that work/heat distributions display non-Gaussian tails describing large and rare deviations from the average and/or most probable behavior [71,72,73,74]. These theoretical studies open the way to investigate the

possible relevance of these large deviations in other non-equilibrium systems in condensed matter physics.

The FT by Crooks can be applied and tested by measuring the unfolding and refolding work distributions in single molecule pulling experiments (Fig. 4a). For example, let us consider the case of a molecule (e. g. a DNA or RNA hairpin or a protein) initially in thermal equilibrium in the folded (F) or native state. By applying mechanical force (e. g. using atomic force microscopy or optical tweezers) the molecule can be mechanically unfolded and the conformation of the molecule changed from the native to the unfolded (U) state. The unfolding event is observed by the presence of a rip in the force-extension curve of the molecule (Fig. 4b). During the unfolding process the tip of the cantilever or the bead in the trap exerts a mechanical work on the molecule that is given by,

$$W = \int_{x_0}^{x_f} F \mathrm{d}x \tag{22}$$

where x_0, x_f are the initial and final extension of the molecule. In (22) we are assuming that the molecular extension x is the externally controlled parameter (i. e. $\lambda \equiv x$) which is not necessarily the case. However the corrections introduced by using (22) are shown to be often small. The work (22) done upon the molecule along a given path corresponds to the area below the force-extension curve that is limited by the initial and final extensions, x_0 and x_f (Fig. 4b). Because the unfolding of the
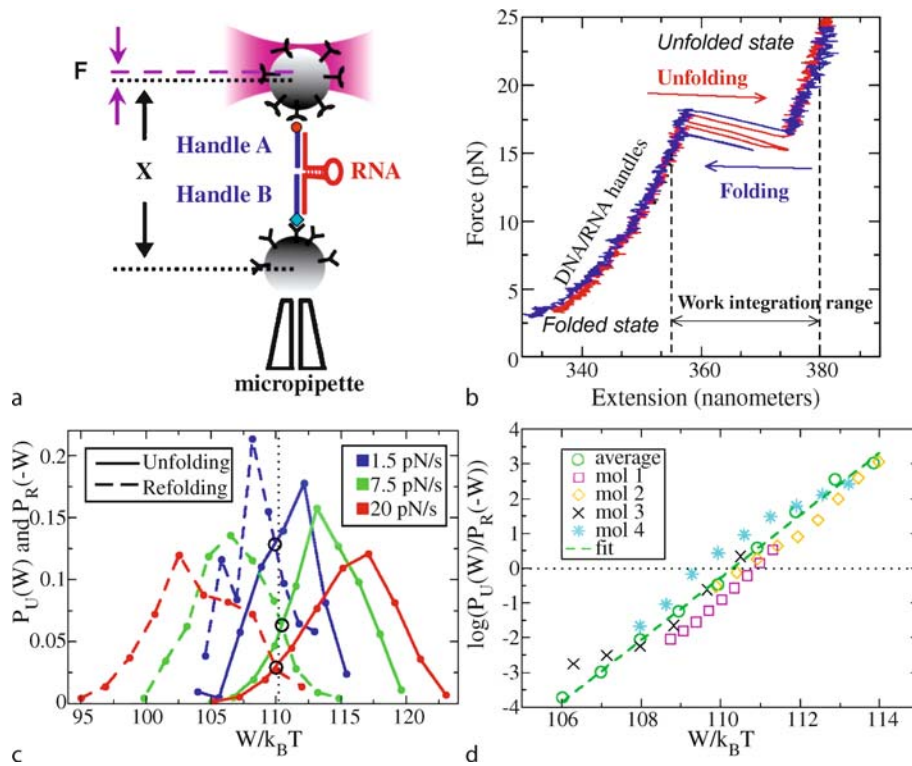
Fluctuation Theorems, Brownian Motors and Thermodynamics of Small Systems, Figure 4
Experimental measurement of work fluctuations in small systems using single molecules. a Experimental setup in RNA force pulling experiments using optical tweezers. An RNA hairpin of a few tens of base pairs is inserted between two molecular handles (A and B) that are attached to two beads. One bead is trapped in the optical well, the other bead is immobilized on the tip of a micropipette. As the optical trap is moved relative to the micropipette the RNA molecule is stretched. b Measured force-extension cycle showing a force rip around 15pN characteristic of the unfolding/refolding of the RNA molecule. The work along a given unfolding/refolding force-extension curve corresponds to the area below the curve integrated along a given range of the molecular extension (indicated by the two *vertical dashed lines*). c Work distributions for the unfolding and refolding process measured at three different pulling speeds. According to the FT by Crooks all curves should cross at a given value of the work, $W = \Delta F$, independently of the pulling speed. In this case $\Delta F \simeq 110 k_B T$, a number that includes also the stretching contributions from the hybrid handles and the ssRNA. d Experimental verification of the FT by Crooks for four different tethered molecules (all with identical sequence). The *black dashed line* is the best fit for all curves and has slope equal to 0.9 ± 0.1. Results taken from [75]

molecule is a stochastic (i. e. random) process, the value of the force at which the molecule unfolds changes from experiment to experiment and so does the value of the mechanical work required to unfold the molecule. Upon repetition of the experiment many times a distribution of unfolding work values for the molecule to go from the folded (F) to the unfolded (U) state is obtained, $P_{F \to U}(W)$. A related work distribution can be obtained if we reverse the pulling process by releasing the molecular extension at the same speed at which the molecule was previously pulled, to allow the molecule to go from the unfolded (U) to the folded (F) state. In that case the molecule refolds by performing mechanical work on the cantilever or the optical trap. Upon repetition of the folding process many times the work distribution, $P_{U \to F}(W)$ can be also measured. The unfolding and refolding work distributions can then be measured in stretching/releasing cycles (Fig. 4c). From (20) we observe that $P_F(\Delta F) = P_R(-\Delta F)$ so the forward and reverse work probability distributions cross each other at $W = \Delta F$. In Fig. 4c we observe that both distributions cross each other at a value (ΔF) that is independent of the pulling speed as expected. Figure 4d shows the experimental verification of (20).

The FT by Crooks has been tested in different types of RNA molecules and the method has been shown capable of recovering free-energies under strong non-equilibrium conditions [75]. The work probability distributions were measured along the unfolding and refolding pathways for a three-way junction RNA molecule and found to strongly deviate from a Gaussian distribution [75]. These experimental results pave the way for other related studies, for example in molecular dynamics simulations [76].

These kind of studies will expand in the future to cover more complex cases and other non-equilibrium situations such as the free-energy recovery of folding free energies in native states in proteins or free energies in misfolded structures and intermediate states in RNA molecules and proteins. Ultimately FTs, when combined with SME, will offer an excellent opportunity to characterize and understand the possible biological relevance of large deviations and extremal fluctuations in molecular systems.

Future Directions

The experimental and theoretical study of non-equilibrium small systems offers exciting possibilities for the statistical physicist and the biophysicist. This discipline aims to describe the novel properties observed in bio-molecules and molecular machines operating far from equilibrium, such as the folding of a nucleic acid or a protein or the trans-location motion of a molecular motor.

We are just starting to have a glance about how these small objects exchange energy with their environment. It is a well known fact in molecular biology and biochemistry that biological function at the molecular level is tightly related to structure. It might not be surprising that the link between molecular structure and biological function is encoded in the low frequency region of the spectrum of non-equilibrium energy fluctuations (the spectrum of energy fluctuations extending far at the most extreme tails of the distribution). It is difficult to imagine how bio-molecular processes, often carrying a lot of information, can operate solely from high frequency events describing the motion of a few number of atoms. Rather, these should somehow rely on the low frequency cooperative motion between different and distant parts of the molecule. Investigating fluctuations in non-equilibrium systems calls for a deeper theoretical understanding of large deviation functions in non-equilibrium systems as well as more systematic and accurate experiments identifying sources of large energy fluctuations in biological systems.

We are at the dawn of an interdisciplinary scientific discipline that will bring together scientists with expertises coming from very different branches of knowledge. This merging process might culminate with the future engineering of artificial mesoscopic structures capable of reproducing and even improving the behavior of the biological ones.

Bibliography

Primary Literature

1. Bustamante C, Liphardt J, Ritort F (2005) The nonequilibrium thermodynamics of small systems. Phys Today 58:43–48
2. Hill TL (1994) Thermodynamics of small systems. Dover Publications, New York
3. Spudich A (2002) How molecular motors work. Nature 372:515–518
4. Wang H, Oster G (2002) The Stokes efficiency for molecular motors and its applications. Europhys Lett 57:134–140
5. Bustamante C, Chemla YR, Forde NR, Izhaky D (2004) Mechanical processes in biochemistry. Annu Rev Biochem 73:705–748
6. Yin H, Wang MD, Svoboda K, Landick R, Block SM, Gelles J (1995) Transcription against an applied force. Science 270:1653–1657
7. Wang MD, Schnitzer MJ, Yin H, Landick R, Gelles J, Block SM (1998) Force and velocity measured for single molecules of RNA polymerase. Science 282:902–907
8. Davenport RJ, Wuite GJ, Landick R, Bustamante C (2000) Single-molecule study of transcriptional pausing and arrest by E. coli RNA polymerase. Science 287:2497–2500
9. Forde NR, Izhaky D, Woodcock GR, Wuite GJL, Bustamante C (2002) Using mechanical force to probe the mechanism of pausing and arrest during continuous elongation by Es-

cherichia coli RNA polymerase. Proceedings of the National Academy of Sciences 99:11682–11687

10. Ediger MD, Angell CA, Nagel SR (1996) Supercooled liquids and glasses. J Phys Chem 100:13200–13212

11. Cipelletti L, Ramos L (2005) Slow dynamics in glassy soft matter. J Phys 17:R253–R285

12. Jarzynski C (1997) Non-equilibrium equality for free-energy differences. Phys Rev Lett 78:2690–2693

13. Kawai R, Parrondo JMR, den Broeck CV (2007) Dissipation: the phase-space perspective. Phys Rev Lett 98:080602

14. Ritort F, Bustamante C, Tinoco I (2002) A two-state kinetic model for the unfolding of single molecules by mechanical force. Proceedings of the National Academy of Sciences 99:13544–13548

15. Crooks GE (1999) Entropy production fluctuation theorem and the non-equilibrium work relation for free-energy differences. Phys Rev E 60:2721–2726

16. Crooks GE (2000) Path-ensemble averages in systems driven far from equilibrium. Phys Rev E 61:2361–2366

17. Kurchan J (1998) Fluctuation theorem for stochastic dynamics. J Phys A 31:3719–3729

18. Seifert U (2005) Entropy production along a stochastic trajectory and an integral fluctuation theorem. Phys Rev Lett 95:040602

19. Bochkov GN, Kuzovlev JE (1981) Non-linear fluctuation relations and stochastic models in non-equilibrium thermodynamics. I. Generalized fluctuation-dissipation theorem. Physica A 106:443–479

20. Zwanzig RW (1954) High-temperature equation of state by a perturbation method. i. non-polar gases. J Chem Phys 22:1420–1426

21. Kirkwood JG (1935) Statistical mechanics of fluid mixtures. J Chem Phys 3:300–313

22. Cohen EGD, Mauzerall D (2004) A note on the Jarzynski equality. J Stat Mech P07006

23. Jarzynski C (2004) Non-equilibrium work theorem for a system strongly coupled to a thermal environment. J Stat Mech P09005

24. Astumian RD (2006) The unreasonable effectiveness of equilibrium-like theory for interpreting non-equilibrium experiments. Am J Phys 74:683

25. Evans DJ, Searles DJ (1994) Equilibrium micro-states which generate second law violating steady-states. Phys Rev E 50:1645–1648

26. Cohen EGD, Evans DJ, Morriss GP (1993) Probability of second law violations in shearing steady states. Phys Rev Lett 71:2401–2404

27. Gallavotti G, Cohen EGD (1995) Dynamical ensembles in non-equilibrium statistical mechanics. Phys Rev Lett 74:2694–2697

28. Lebowitz JL, Spohn H (1999) A Gallavotti-Cohen type symmetry in the large deviation functional for stochastic dynamics. J Stat Phys 95:333–365

29. Maes C (1999) The fluctuation theorem as a Gibbs property. J Stat Phys 95:367–392

30. Oono Y, Paniconi M (1998) Steady state thermodynamics. Prog Theor Phys Suppl 130:29–44

31. Hatano T, Sasa S (2005) Steady-state thermodynamics of Langevin systems. Phys Rev Lett 86:3463–3466

32. Speck T, Seifert U (2005) Integral fluctuation theorem for the housekeeping heat. J Phys A 38:L581–L588

33. Ciliberto S, Laroche C (1998) An experimental test of the

34. Ciliberto S, Garnier N, Hernandez S, Lacpatia C, Pinton JF, Ruiz-Chavarria G (2004) Experimental test of the Gallavotti–Cohen fluctuation theorem in turbulent flows. Physica A 340:240–250

35. Wang GM, Sevick EM, Mittag E, Searles DJ, Evans DJ (2002) Experimental demonstration of violations of the second law of thermodynamics for small systems and short timescales. Phys Rev Lett 89:050601

36. Carberry DM, Reid JC, Wang GM, Sevick EM, Searles DJ, Evans DJ (2004) Fluctuations and irreversibility: An experimental demonstration of a second-law-like theorem using a colloidal particle held in an optical trap. Phys Rev Lett 92:140601

37. Trepagnier EH, Jarzynski C, Ritort F, Crooks GE, Bustamante C, Liphardt J (2004) Experimental test of Hatano and Sasa's nonequilibrium steady state equality. Proceedings of the National Academy of Sciences 101:15038–15041

38. Blickle V, Speck T, Helden L, Seifert U, Bechinger C (2005) Thermodynamics of a colloidal particle in a time-dependent non-harmonic potential. Phys Rev Lett 93:158105

39. Hummer G, Szabo A (2001) Free-energy reconstruction from non-equilibrium single-molecule experiments. Proc Nat Acad Sci 98:3658–3661

40. Jarzynski C (2001) How does a system respond when driven away from thermal equilibrium? Proc Nat Acad Sci 98:3636–3638

41. Hummer G, Szabo A (2005) Free-energy surfaces from single-molecule force spectroscopy. Acc Chem Res 38:504–513

42. Zuckermann DM, Woolf TB (2002) Theory of systematic computational error in free energy differences. Phys Rev Lett 89:180602

43. Gore J, Ritort F, Bustamante C (2003) Bias and error in estimates of equilibrium free-energy differences from non-equilibrium measurements. Proc Nat Acad Sci 100:12564–12569

44. Liphardt J, Dumont S, Smith SB, Tinoco I, Bustamante C (2002) Equilibrium information from non-equilibrium measurements in an experimental test of the Jarzynski equality. Science 296:1833–1835

45. Schuler S, Speck T, Tierz C, Wrachtrup J, Seifert U (2005) Experimental test of the fluctuation theorem for a driven two-level system with time-dependent rates. Phys Rev Lett 94:180602

46. Douarche F, Ciliberto S, Petrosyan A (2005) An experimental test of the Jarzynski equality in a mechanical experiment. Europhys Lett 70:593–598

47. Douarche F, Ciliberto S, Petrosyan A (2005) Estimate of the free energy difference in mechanical systems from work fluctuations: experiments and models. J Stat Mech (Theor. Exp.) P09011

48. Van Zon R, Cohen EGD (2003) Extension of the fluctuation theorem. Phys Rev Lett 91:110601

49. Van Zon R, Cohen EGD (2003) Stationary and transient work-fluctuation theorems for a dragged Brownian particle. Phys Rev E 67:046102

50. Seifert U (2004) Fluctuation theorem for birth-death or chemical master equations with time-dependent rates. J Phys A 37:L517–L521

51. Jarzynski C, Wojcik DK (2004) Classical and quantum fluctuation theorems for heat exchange. Phys Rev Lett 92:230602

52. Seifert U (2005) Entropy production along a stochastic tra-

Gallavotti–Cohen fluctuation theorem. J Phys IV (France) 8:215–220

jectory and an integral fluctuation theorem. Phys Rev Lett 95:040602

53. Reid JC, Sevick EM, Evans DJ (2005) A unified description of two theorems in non-equilibrium statistical mechanics: The fluctuation theorem and the work relation. Europhys Lett 72: 726–730

54. Hummer G (2001) Fast-growth thermodynamic integration: Error and efficiency analysis. J Chem Phys 114:7330–7337

55. Isralewitz B, Gao M, Schulten K (2001) Steered molecular dynamics and mechanical functions of proteins. Curr Opin Struct Biol 11:224–230

56. Jensen MO, Park S, Tajkhorshid E, Schulten K (2002) Energetics of glycerol conduction through aquaglyceroporin Glpf. Proc Nat Acad Sci 99:6731–6736

57. Park S, Khalili-Araghi F, Tajkhorshid E, Schulten K (2003) Free-energy calculation from steered molecular dynamics simulations using Jarzynski's equality. J Phys Chem B 119: 3559–3566

58. Andriocioaei I, Dinner AR, Karplus M (2003) Self-guided enhanced sampling methods for thermodynamic averages. J Chem Phys 118:1074–1084

59. Park S, Schulten K (2004) Calculating potentials of mean force from steered molecular dynamics simulation. J Chem Phys 13:5946–5961

60. Bennett CH (1976) Efficient estimation of free-energy differences from Monte Carlo data. J Comput Phys 22:245–268

61. Wood RH, Mühlbauer WCF, Thompson PT (1991) Systematic errors in free energy perturbation calculations due to a finite sample of configuration space: sample-size hysteresis. J Phys Chem 95:6670–6675

62. Hendrix DA, Jarzynski C (2001) A "fast growth" method of computing free-energy differences. J Chem Phys 114:5974–5981

63. Shirts MR, Bair E, Hooker G, Pande VS (2003) Equilibrium free energies from non-equilibrium measurements using maximum-likelihood methods. Phys Rev Lett 91:140601

64. Qian H (2005) Cycle kinetics, steady state thermodynamics and motors – a paradigm for living matter physics. J Phys (Condensed Matter) 17:S3783–3794

65. Min W, Jiang L, Yu J, Kou SC, Qian H, Xie XS (2005) Non-equilibrium steady state of a nanometric biochemical system: Determining the thermodynamic driving force from single enzyme turnover time traces. Nanoletters 5:2373–2378

66. Mazonka O, Jarzynski C (1999) Exactly solvable model illustrating far-from-equilibrium predictions. Preprint arXiv:cond-mat/9912121

67. Lhua RC, Grossberg AY (2005) Practical applicability of the Jarzynski relation in statistical mechanics: a pedagogical example. J Phys Chem B 109:6805–6811

68. Bena I, Van den Broeck C, Kawai R (2005) Jarzynski equality for the Jepsen gas. Europhys Lett 71:879–885

69. Speck T, Seifert U (2005) Dissipated work in driven harmonic diffusive systems: general solution and application to stretching rouse polymers. Eur Phys J B 43:521–527

70. Cleuren B, Van den Broeck C, Kawai R (2007) Fluctuation and dissipation of work in a Joule experiment. Phys Rev Lett 98:080602

71. Van Zon R, Cohen EGD (2004) Extended heat-fluctuation theorems for a system with deterministic and stochastic forces. Phys Rev E 69:056121

72. Ritort F (2004) Work and heat fluctuations in two-state sys-

tems: a trajectory thermodynamics formalism. J Stat Mech (Theor. Exp.) P10016

73. Imparato A, Peliti L (2005) Work distribution and path integrals in mean-field systems. Europhys Lett 70:740–746

74. Imparato A, Peliti L (2005) Work probability distribution in systems driven out of equilibrium. Phys Rev E 72:046114

75. Collin D, Ritort F, Jarzynski C, Smith SB, Tinoco I, Bustamante C (2005) Verification of the Crooks fluctuation theorem and recovery of RNA folding free energies. Nature 437:231–234

76. Kosztin I, Barz B, Janosi L (2006) Calculating potentials of mean force and diffusion coefficients from non-equilibrium processes without Jarzynski equality. J Chem Phys 124:064106

Books and Reviews

Comptes Rendus Physique (2007) Work, dissipation and fluctuations in non-equilibrium physics, vol 8(5–6). Elsevier

Duplantier B, Dalibard J, Rivasseau V (eds)(2004) Bose-Einstein Condensation and Entropy 2, Birkhäuser, Basel. Contributions by Maes C, On the origin and use of fluctuation relations on the entropy, p 145–191 and Ritort F, Work fluctuations, transient violations of the second law and free-energy recovery methods, pp 193–226. Available also at: http://www.ffn.ub.es/ritort/publications.html

Evans DJ, Searles D (2002) The fluctuation theorem. Adv Phys 51:1529–1585

Garbaczewski P, Olkiewicz R (eds)(2002) Dynamics of dissipation. Springer, Berlin. Contribution by Jarzynski C, What is the microscopic response of a system driven far from equilibrium, pp 63–82

Mathews CK, van Holde KE, Ahern KG (2000) Biochemistry. Addison-Wesley, Longman, SF

Ritort F (2008) Nonequilibrium fluctuations in small systems: From physics to biology. In: Rice SA (ed) Advances in Chemical Physics, vol 137. Wiley, Hoboken, NJ, pp 31–123

Fluid Dynamics, Pattern Formation

MICHAEL BESTEHORN
Brandenburg University of Technology,
Cottbus, Germany

Article Outline

Glossary

Order parameter(s) (Field) variable(s) characterizing the spatio-temporal state of a system. Other state variables such as velocity, temperature, density, etc. can be computed if the order parameter(s) are known.

Control parameter(s) Parameters which are fixed and can be tuned from outside of the system under consideration.

Critical point, threshold, onset The points in control parameter space where new and qualitatively different solutions bifurcate from (usually simpler) ones.

Slaving principle Stated by H. Haken in 1975, the slaving principle allows for a huge reduction of degrees of freedom close to a critical point. It states that a very large number of linearly damped modes are slaved to and therefore completely determined by the few modes that grow in the vicinity of the critical point. The amplitudes of the growing modes are also called order parameters.

Natural patterns Spatial patterns showing a certain periodic (near) order, but also defects, grain boundaries etc.

Turing patterns Natural patterns that have a certain typical length scale and that show relaxation to a stationary state in the long term. Typical ingredients of Turing patterns are stripes, hexagons and squares.

Swift–Hohenberg equation Derived by Swift and Hohenberg in 1977 and nowadays established as the standard form for a scalar, real-valued order parameter equation showing Turing patterns at onset.

Coarsening The monotonic increase of the typical length scale of a structure in time. Often connected to spinodal decomposition, for example, of a binary mixture of non-mixing components such as water and oil. Small oil droplets in the beginning merge and finally form a large oil drop on the water surface. Coarsening slows down if the length scale increases.

Definition of the Subject

The state of a fluid is described by its velocity, density, pressure, and temperature. All these variables depend in general on space and time. Pattern formation refers to the situation where one or more of these variables are organized within a certain spatial and/or temporal order. This order has macroscopic length and time scales, that is, characteristic lengths and times are much larger than those of the atoms or molecules which constitute the fluid. Therefore a continuous description is appropriate.

Macroscopic fluid patterns may be encountered in nature as well as in technological applications for a large variety of different systems. Far from being complete, we mention some examples:

- Water waves caused by wind or by sea quakes and land slides (Tsunamis).
- Localized excitations of the surface of a fluid (solitons), such as that seen on shallow water channels.
- Shear instabilities in clouds or in multi-layer systems such as the Kelvin–Helmholtz instability or the Rayleigh–Taylor-instability.
- Surface deflections in the form of holes or drops of thin fluid films in coating or wetting processes.
- Convection instabilities in laboratory experiments, but also in the atmosphere, in the earth's interior or in stars.
- Creation and controlled growth of ordered structures in (nano-) technological applications.
- Biological applications: Behavior of liquid films on leaves or of the tear film on the cornea of the eye. Dynamics of thin blood layers, blood clotting.
- Films on the walls of combustion cells.
- Lubrication films in mechanical machines.

Fluid patterns may occur due to several mechanisms. One can distinguish between two main cases: Patterns excited and organized by some external forces or disturbances (such as Tsunamis) and those formed by instabilities. The latter may show the aspects of self-organization and will be the focus of the present contribution.

Introduction

Since the first observations of Michael Faraday almost 180 years ago [42] (Fig. 1), pattern formation in liquids or gases (*fluids*) has been subject to innumerable experimental [19,80,90], theoretical [32,34,43] and, later on, numerical work [5,26,73]. After the famous experiments by Henri Bénard around 1901 [17], convection in a single or later in multi-component fluids came into the focus of interest. The first theoretical studies were made by Lord Rayleigh [78]. Theoretical computations up to the early 1960s were restricted on the linearized basic equations and could explain the existence of critical points in parameter space as well as the observed length scales of the structures found experimentally [18,72]. In the meantime, Alan Turing [89] showed in his famous paper of 1952 that similar patterns could emerge out of equilibrium in reaction-diffusion systems. It took almost 40 years for an experimental confirmation using the so-called CIMA reaction [31,71].

With the appearance and rapid development of computers, the field gained further momentum from the new discipline of nonlinear dynamics and nonlinear system
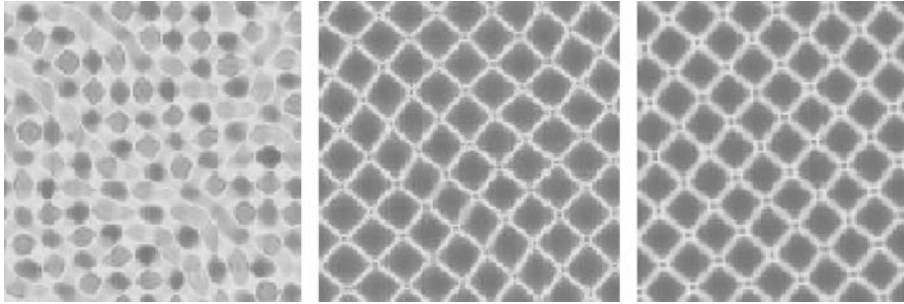
Fluid Dynamics, Pattern Formation, Figure 1
Michael Faraday observed surface patterns on a liquid horizontal layer if the whole layer vibrated vertically with a certain amplitude and frequency. Very often, regular squares are found, as shown in the time series as a numerical result of the shallow water equations (see Sect. "Surface Waves")

theory [4,45,48,50]. Early computations in 1963 by Edward Lorenz of a system of three coupled ordinary differential equations derived by a crudely truncated mode expansion of the Navier–Stokes equations revealed the first chaotic attractor of a dissipative system [60]. Though the chaotic behavior seen in the Lorenz equations does not originate from hydrodynamic equations and has nothing to do with irregular fluid behavior, the Lorenz model now stands on its own as a paradigm for a relatively simple system showing low dimensional chaos [87].

Patterns that emerge from an instability roughly pass through two phases. As long as amplitudes (or order parameters) are small, the behavior is often determined by the linear parts of the system and exponential growth of a certain part of the mode spectrum is found. In the second phase, nonlinearities come into play and may lead to saturation and selection of certain mode configurations, seen then as regular structures in configuration space (Fig. 2). The full mathematical description of hydrodynamic systems has been well known for a long time. Fluid motion is described by the Euler or Navier–Stokes equations, temperature fields by the heat equation and chemical concentrations by some nonlinear reaction-diffusion equations. The location and spatio-temporal evolution of surfaces or interfaces can be computed by the kinematic boundary conditions if the velocity of the fluid near the interface is known. All these equations can be coupled and provided with suitable boundary and initial conditions, resulting in rather complicated systems of nonlinear partial differential equations. Even today in the age of supercomputers, their further treatment, especially in three spatial dimensions, remains a challenge.

On the other hand, directly solving the basic equations, can be considered merely as another experiment. For these reasons and to get a deeper insight into the physics behind pattern formation, other methods have been devised. Very often one of the three spatial dimension is distinguished, either for physical reasons or simply due to the geometry of the system. A good example is surface waves on a water layer. Here, the behavior of the solutions in the vertical direction (z) is very different from those in the horizontal ones. For shallow water waves (wave length long compared to the layer depth) the velocities are more or less independent on z, where in the other limit of deep water waves, fluid motion takes only place along a small layer under the surface and decreases exponentially with depth. In both cases one may reduce the dimension of the basic problem by an expansion with respect to simple functions for the vertical dependence of the variables [33]. An analogue method can be applied describing thin film surface patterns [70]. Also for convection cells, the vertical dimension plays a special role and the solution can be projected onto a few modes near the critical point [25,73].

Another concept that reduces the number of dependent variables and equations is that of order parameters. The notion of the "order parameter" goes back to Landau [58] and refers originally to a variable that measures the order of a certain system. Rather a variable than a parameter, the order parameter normally depends on time and, in theories describing the formation of natural patterns, also on space [66]. Thus, the order parameter equation (abbreviated: OPE) is a partial differential equation with certain nonlinear terms that become important for pattern selection and saturation.

Theoretical methods developed by the Haken school [47,48,50] starting in the 1970s allow for a systematic derivation of the OPEs (sometimes also called "generalized Ginzburg–Landau equations") for a great variety of nonequilibrium and open systems from physics, chemistry and biology. The key idea is to find a reduced

N=2   N=3   N=5
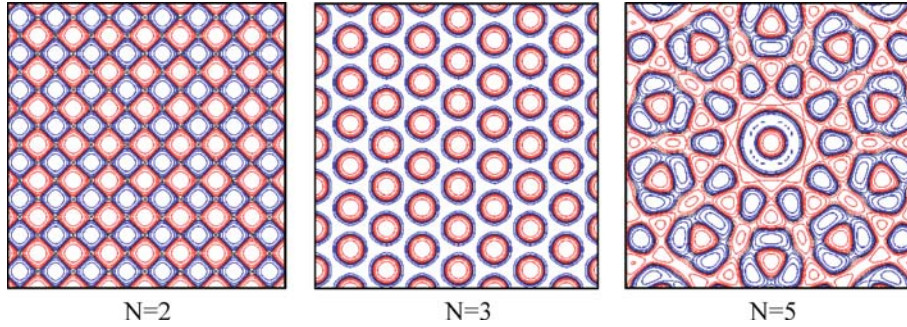
Fluid Dynamics, Pattern Formation, Figure 2
The composition of plane waves with the same wave number but different orientation in 2D space results in regular patterns. For two modes ($N = 2$) one sees squares, for $N = 3$ hexagons and for $N > 3$ quasi periodic structures in space or Penrose tilings [74] are found

description in terms of relevant or active modes close to a certain bifurcation point. The amplitudes of these active modes, the order parameters, now generalized to a nonequilibrium, pattern forming system, obey unified and simplified equations, namely the OPEs. It turns out that the structure of these equations depends not so much on the particular system under consideration as on the type of bifurcation. To each type of bifurcation a special "normal form" of OPE is related [35]. In deriving the OPEs, the slaving principle [50] allows us to eliminate a huge number of slaved variables and express them by the active ones.

This contribution is concerned mainly with structures in fluids that originate from self-organized processes. It tries to bring together direct numerical solutions of hydrodynamic equations with the modern concepts of pattern formation. After introducing the basic equations (Sect. "The Basic Equations of Fluid Dynamics") of fluid dynamics, it presents a short section on waves and descriptions reduced by geometrical reasons. Several types of instabilities are discussed in Sect. "Instabilities", together with computer solutions for the different cases. Section "Order Parameter Equations" presents different types of two-dimensional order parameter equations. Finally, Sect. "Conserved Order Parameter Fields" is devoted to the special case of conserved order parameters.

The Basic Equations of Fluid Dynamics

Let the state of a fluid be described by its velocity, its density, its pressure, and its temperature field

$$\vec{v}(\vec{r}, t), \quad \rho(\vec{r}, t), \quad p(\vec{r}, t), \quad T(\vec{r}, t). \tag{1}$$

In this section we wish to specify the basic hydrodynamic equations that rule the spatio-temporal behavior

of these seven variables. They have to be completed by suitable boundary conditions (abbreviated: b.c.) which we shall present later with the particular systems under consideration.

Continuity Equation

The conservation of mass yields the continuity equation

$$\partial_t \rho + \text{div}(\rho \vec{v}) = \partial_t \rho + (\vec{v} \cdot \nabla)\rho + \rho \ \text{div} \ \vec{v} = 0. \tag{2}$$

In most cases, liquids are difficult to compress. One can usually assume that a volume element does not change its density while it moves with the fluid (Lagrangian description)

$$\partial_t \rho + (\vec{v} \cdot \nabla)\rho = 0.$$

From (2) one finds the condition of incompressibility

$$\text{div} \ \vec{v}(\vec{r}, t) = 0, \tag{3}$$

or, in other words, the velocity field is free of sources and sinks. Equation (3) can be satisfied by the ansatz

$$\vec{v}(\vec{r}, t) = \text{curl} \ \vec{A}(\vec{r}, t) \tag{4}$$

where $\vec{A}$ plays the role of a vector potential. In (4) one can use the particular decomposition [5,32]

$$\vec{v}(\vec{r}, t) = \text{curl}(\Phi \acute{e}_z) + \text{curl curl}(\Psi \acute{e}_z)$$
$$= \begin{pmatrix} \partial_y \Phi + \partial_z \partial_x \Psi \\ -\partial_x \Phi + \partial_z \partial_y \Psi \\ -\Delta_2 \Psi \end{pmatrix} \tag{5}$$

with the two independent scalar functions $\Phi(\vec{r}, t)$ and $\Psi(\vec{r}, t)$ and $\Delta_2 = \partial_{xx}^2 + \partial_{yy}^2$ as the 2D-Laplacian.

If the velocity field is irrotational, that is without vortices (curl $\vec{v} = 0$), it can be derived from a scalar potential

$$\vec{v} = \operatorname{grad} \phi . \tag{6}$$

For incompressible and irrotational flows, hydrodynamics is reduced to a boundary value problem, since the potential must fulfill the Laplace equation

$$\operatorname{div} \vec{v} = \Delta \phi = 0 \tag{7}$$

and the velocity field is solely determined by its boundary conditions.

Euler Equations

For a perfect fluid, a fluid with no viscosity, one derives the Euler equations from the law of conservation of momentum [56]. They read

$$\rho(\vec{r}, t) \left[\partial_t \vec{v}(\vec{r}, t) + (\vec{v}(\vec{r}, t) \cdot \nabla) \, \vec{v}(\vec{r}, t) \right] = \\ - \operatorname{grad} \, p(\vec{r}, t) + \vec{f}(\vec{r}, t) , \tag{8}$$

where $\vec{f}$ denotes external volume forces. Together with a state equation of the form

$$p = p(\rho, T) , \tag{9}$$

the continuity equation (2) and the temperature equations (to be shown below) (Subsect. "Transport Equations") constitute the basic set for the seven state variables (1).

Incompressible Fluids   For an incompressible fluid, a state equation of the form (9) makes no sense since pressure will not change with density. So p can be eliminated by forming the curl of (8)

$$\partial_t \vec{\Omega} = \operatorname{curl}(\vec{v} \times \vec{\Omega}) + \frac{1}{\rho} \operatorname{curl} \vec{f} \tag{10}$$

where

$$\vec{\Omega} = \operatorname{curl} \vec{v} \tag{11}$$

denotes the vorticity. If p must be known, it can be computed from the divergence of (8) which yields

$$\nabla^2 p = \rho \{ - \operatorname{Tr} \left[(\nabla \circ \vec{v})(\nabla \circ \vec{v}) \right] + \operatorname{div} \vec{f} \} , \tag{12}$$

where $\circ$ is the dyadic product and $\operatorname{Tr}[\dots]$ the trace.

Incompressible Irrotational Fluids   If, in addition, the flow is free of vortices, one may integrate the Euler equations and find the theorem of Bernoulli

$$\partial_t \phi = -\frac{1}{\rho}(p + U) - \frac{1}{2}(\nabla \phi)^2 \tag{13}$$

where U is the potential to $\vec{f}$ ($\vec{f}$ must be irrotational, too). For stationary solutions, the velocity potential ϕ is found from (7) and (13) can be used to determine the pressure.

Navier–Stokes Equations

Compressible Fluids   In real fluids, shear stresses are a result of friction. They must be added to the balance of momentum and yield the Navier–Stokes equations. For a compressible Newtonian fluid they read

$$\rho \left[\partial_t \vec{v} + (\vec{v} \cdot \nabla) \, \vec{v} \right] \\ = - \operatorname{grad} p + \vec{f} + \eta \, \Delta \vec{v} + \left(\zeta + \frac{\eta}{3} \right) \operatorname{grad} \operatorname{div} \vec{v} \tag{14}$$

where η denotes the first and ζ the second viscosity [59].

Incompressible Fluids   The Navier–Stokes equations for incompressible fluids are simpler:

$$\rho \left[\partial_t \vec{v} + (\vec{v} \cdot \nabla) \, \vec{v} \right] = -\nabla p + \vec{f} + \eta \, \Delta \vec{v} . \tag{15}$$

Again, pressure can be eliminated by forming the curl. Taking the ansatz (5), the z-components of the curl and of the curl of the curl of (15), we have

$$\{ \nu \, \Delta - \partial_t \} \Delta_2 \Phi(\vec{r}, t) \\ = \left[\operatorname{curl}((\vec{v} \cdot \nabla)\vec{v}) \right]_z + \frac{1}{\rho} \left(\partial_x f_y - \partial_y f_x \right) \tag{16a}$$

$$\{ \nu \, \Delta - \partial_t \} \Delta \Delta_2 \Psi(\vec{r}, t) = \left[\operatorname{curl} \operatorname{curl}((\vec{v} \cdot \nabla)\vec{v}) \right]_z \\ + \frac{1}{\rho} \left(\Delta_2 f_z - \partial_x \partial_z f_x - \partial_y \partial_z f_y \right) . \tag{16b}$$

Here we have introduced the kinematic viscosity $\nu = \eta/\rho$. We note that this decomposition is of particular interest if $f_x = f_y = 0$ as is the case in convection problems of a plane layer.

Incompressible Fluids with a Small Reynolds Number
For some applications it is convenient to use the Navier–Stokes equations in dimensionless form. With scaling with respect to a characteristic length L and velocity V_0:

$$\vec{r} = L \cdot \vec{r}' , \quad t = (L/V_0) \cdot t' , \quad \vec{v} = V_0 \cdot \vec{v}' , \quad p' = \frac{L}{\eta V_0} \cdot p ,$$

(15) turns into

$$R_e \left[\partial_{t'} \vec{v}' + (\vec{v}' \cdot \nabla') \, \vec{v}' \right] = -\nabla' p' + \Delta' \vec{v}' \,. \tag{17}$$

(We assumed a potential for $\vec{f}$ which can be confined into p.) The dimensionless quantity

$$R_e = \frac{LV_0}{\nu} \tag{18}$$

is the Reynolds number. If $R_e \ll 1$, the left hand side of (17) can be neglected and the Navier–Stokes equations become linear (primes omitted):

$$\Delta \vec{v} = \nabla p \,. \tag{19}$$

This is the Stokes equation, in which no time derivative of $\vec{v}$ occurs. Thus, as known from over-damped motion, the velocity field directly follows the pressure gradients.

Transport Equations

Scalar fields such as temperature or concentration of a mixture that may diffuse into and be transported with the fluid are ruled by the transport equation. Let $S(\vec{r}, t)$ be the scalar field, then the transport equation reads

$$\partial_t S + (\vec{v} \cdot \nabla) \, S = D_s \Delta S \,, \tag{20}$$

where D_s is the appropriate diffusion coefficient.

Surface Waves

The only elastic forces in fluids are those coming from volume changes and may exist, therefore, only in compressible fluids. They give rise to longitudinal compression waves which usually have small amplitudes and behave linearly in a good approximation. A linear wave equation can be derived with a (space dependent) sound speed [56].

A transversal wave which is also possible in incompressible fluids can be formed along a deformable interface. Gravity and, for small wavelengths, surface tension provide the stabilizing mechanism of a flat surface, around which oscillations (gravity waves) may occur. If their amplitudes are big enough, nonlinearities may play an essential role for surface waves, as is clearly seen by solitons [40,64] and wave breaking [39]. For this reason we shall discuss only surface waves in this section.

Gravity Waves

If one assumes an irrotational flow of a perfect and incompressible fluid on a flat substrate and with a free, deformable surface (Fig. 3), then the velocity is determined by the Laplace equation (7) which must be accomplished by boundary conditions at $z = 0$

$$v_z \big|_{z=0} = \partial_z \phi \big|_{z=0} = 0 \tag{21}$$

and at $z = h(x, y, t)$

$$\partial_t \phi \big|_{z=h} = -gh - p(h) \big/ \rho - \frac{1}{2} (\nabla \phi)^2 \tag{22}$$

where g denotes the gravitational acceleration. Equation (22) is nothing other than the Bernoulli equation (13) evaluated at the surface. The surface itself is determined by the so-called kinematic boundary condition that reads (see Fig. 3, right frame).

$$\partial_t h = v_z \big|_{z=h} - v_x \big|_{z=h} \partial_x h - v_y \big|_{z=h} \partial_y h$$
$$= \partial_z \phi \big|_{z=h} - (\partial_x h)(\partial_x \phi)_{z=h} - (\partial_y h)(\partial_y \phi)_{z=h} \,. \tag{23}$$

Shallow Water Equations

For shallow water waves, one can introduce the small parameter

$$\delta = d/\ell \tag{24}$$

which is the ratio of the water depth d and a typical horizontal scale (such as a wavelength) ℓ. Then (7) can be
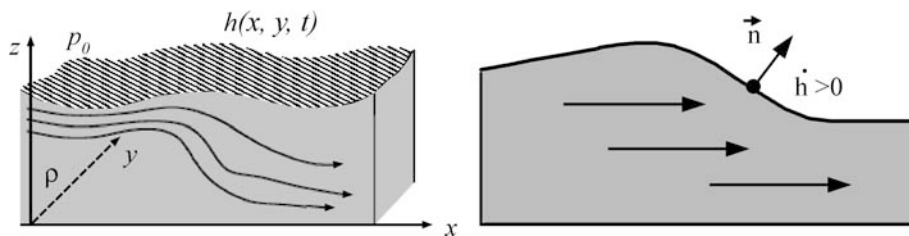
Fluid Dynamics, Pattern Formation, Figure 3
Left: an (incompressible) fluid with a free and deformable surface located at $z = h(x, y, t)$, on which a constant external pressure p_0 is applied. *Right*: The location of a certain point of the surface changes if the fluid is in motion

solved iteratively; the result is a power series in δ^2 [33,39]:

$$\phi(\vec{r}, t) = \Phi(x, y, t)$$
$$+ \delta^2 \left[-\frac{z^2}{2} \Delta_2 \Phi(x, y, t) + \varphi^{(1)}(x, y, t) \right] + O(\delta^4)$$

$$(25)$$

with an arbitrary function φ_1. Inserting (25) into (23) and (22) yields up to the lowest order in δ the *shallow water equations*

$$\partial_t h = -h \Delta_2 \Phi - (\partial_x h)(\partial_x \Phi) - (\partial_y h)(\partial_y \Phi) \quad (26a)$$

$$\partial_t \Phi = -gh - p(h)/\rho - \frac{1}{2}(\partial_x \Phi)^2 - \frac{1}{2}(\partial_y \Phi)^2 . \quad (26b)$$

This is the first example of how to derive a two-dimensional system starting from three-dimensional fluid motion. Equations (26) constitute a closed system of partial differential equations for the evolution of the two functions $h(x, y, t)$ and $\Phi(x, y, t)$. Using (25), one immediately finds from the latter the velocity field (up to the order δ^2).

Numerical Solutions

Figure 4 shows numerical solutions of the shallow water equations (left frame in one dimension, right frame in two dimensions). In one dimension, one sees clearly traveling surface waves which may run around due to the periodic boundary conditions in x. On the other hand, one can recognize a second wave with a smaller amplitude going to the left hand side. Both waves seem to penetrate each other without further interaction. The reason seems to be the smallness of the amplitude which results in a more or less linear behavior. In the two-dimensional frame, a snapshot of the temporal evolution of the surface is presented. The initial condition was chosen randomly. For numerical stability reasons, an additional damping of the form $\tilde{\nu} \Delta_2 \Phi$ was added to the right hand side of (26b) which filters out the short wave lengths. This could be justified phenomenologically by friction and leads in the long term to a fluid at rest, if only gravity acts.

Instabilities

Laplace Pressure and Disjoining Pressure   To discuss Eqs. (26) further, we must elaborate a little on the dependence of the surface pressure on the depth $h(x, y, t)$ and on its curvature $-\Delta h$.

The length scale of the surface structures is proportional to the depth of the fluid layer. If the films are very thin, we expect to have scales in the range or even well below the capillary length $a = \sqrt{\Gamma/g\rho}$ where Γ denotes the surface tension. Then one has to take into account the additional pressure which originates from the curvature of the surface, the so-called *Laplace pressure* [59] $-\Gamma \Delta_2 h$. Thus we substitute in (26b)

$$p(h) = p_1(h) - \Gamma \Delta_2 h , \quad (27)$$

with a function p_1 (the disjoining pressure) that will be specified later [38,91].

Linear Stability Analysis of the Flat Surface   To see if the flat film $h = h_0$ is stable against small perturbations, one may perform a linear stability analysis. Inserting

$$(h - h_0, \Phi) = (a, b) \exp(\lambda t + ikx)$$
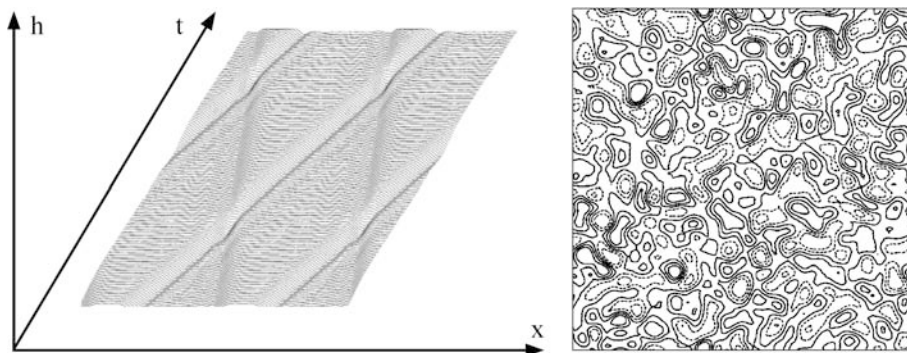
Fluid Dynamics, Pattern Formation, Figure 4
Numerical solutions of the shallow water equations, *left frame* shows a temporal evolution in one dimension, *right frame* a snapshot in two dimensions. *Dashed contour lines* mark troughs, *solid ones* correspond to peaks of the sea

into (26) yields, after linearization with respect to a, b, a linear eigenvalue problem with the solvability condition

$$\lambda_{12}(k) = -\frac{\tilde{v}k^2}{2}$$
$$\pm i|k|\sqrt{h_0\big(g + p_1' + \Gamma k^2\big)/\rho - \tilde{v}^2 k^2/4}, \quad (28)$$

where

$$p_1' = \left.\frac{\mathrm{d}p_1}{\mathrm{d}h}\right|_{h=h_0} .$$

We assume that the artificial viscosity is small, $\tilde{v}^2 \ll \Gamma/\rho$. An instability occurs first at $k = 0$ if the expression under the integral can be negative, that is, for $p_1' + \rho g < 0$. This corresponds to the region of initial thickness h_0 where the generalized pressure

$$p_1(h) + \rho g h \qquad (29)$$

has a negative slope. For that case, the real part of λ_1 starts at $k = 0$ at zero with positive slope, has a maximum at $k = k_c$ and decreases again to the value $-\nu k^2/2$. We shall revisit this instability in the next section and call it there, in a more systematic classification, a type II instability. How can (29) have a negative slope for a certain range of h_0? It is obvious that one has to assume that the pressure p_1 depends in some nonlinear non-monotonic fashion on the value of h (Fig. 5). As we shall see later, this can be the case for very thin films where van der Waals forces between the solid support and the free surface come into play [38,52,91]. But also, in thicker films, this should be possible in non-isothermal situations, where the surface temperature, and therefore the surface tension, changes with the vertical coordinate (Marangoni effect, see Sect. "Instabilities", Fig. 8). If we take (for instance) as a model the polynomial

$$p_1 = c \cdot h \cdot (h - h_1) \cdot (h - h_2), \quad c > 0, \qquad (30)$$

then the flat surface is unstable for h between the two spinodals

$$h_- < h < h_+$$

with $h_\pm$ being the roots of

$$3h^2 - 2h(h_1 + h_2) + h_1 h_2 + \rho g/c = 0 .$$

Figure 6 shows a numerically determined time series of a random dot initial condition. The mean thickness h_0 was chosen in the unstable region. The formation shows traveling waves in the linear phase, followed by coarsening to a large scale structure, in this case one big region of depression, or a hole. This hole becomes more and more symmetric while the velocity decays due to the friction term. Finally, a steady state of a big circular hole remains.
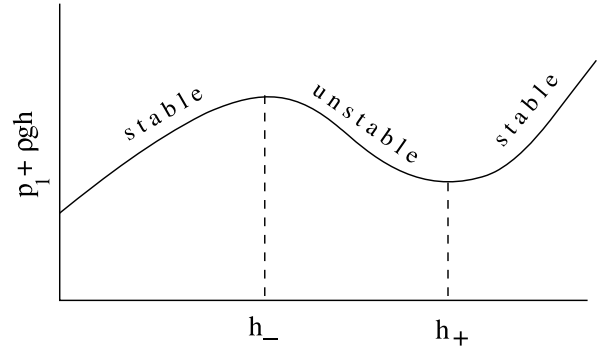
Fluid Dynamics, Pattern Formation, Figure 5
If the pressure depends on h and has a certain region with a negative slope, the flat film is unstable in this region and pattern formation sets in

Parametric Excitation of a Thin Bistable Fluid Layer
One way to replace the energy lost by damping (to "open" the system) is to accelerate the whole layer periodically in the vertical direction. This was done first in an experiment by Michael Faraday in 1831 [42]. He obtained regular surface patterns normally in the form of squares, see Fig. 1.

Faraday patterns can be seen as a solution of the shallow water equations if the gravity constant g is modulated harmonically [7]

$$g(t) = g_0 + g_1 \cos \omega t . \qquad (31)$$

A linear stability analysis leads to a Mathieu equation [1]. The flat film is unstable if frequency and amplitude fall into certain domains, the so-called *Arnold tongues*. There, one usually finds squares for not-too-supercritical values.

We conclude this section by showing a numerical solution of (26) with parameters as in Fig. 6, but now with an additional periodic excitation (Fig. 7). Coarsening is still present, but now oscillating drops emerge in the form of stars. No time stable structure is found in the long time limit.

Instabilities

Mechanisms of Instability in Fluids

We start with the specific case of a plane layer of a viscous fluid with a vertically applied, constant temperature gradient β (Fig. 11), where

$$\beta = (T_1 - T_0)/d \qquad (32)$$

and T_0, T_1 are the temperatures at the lower, upper side
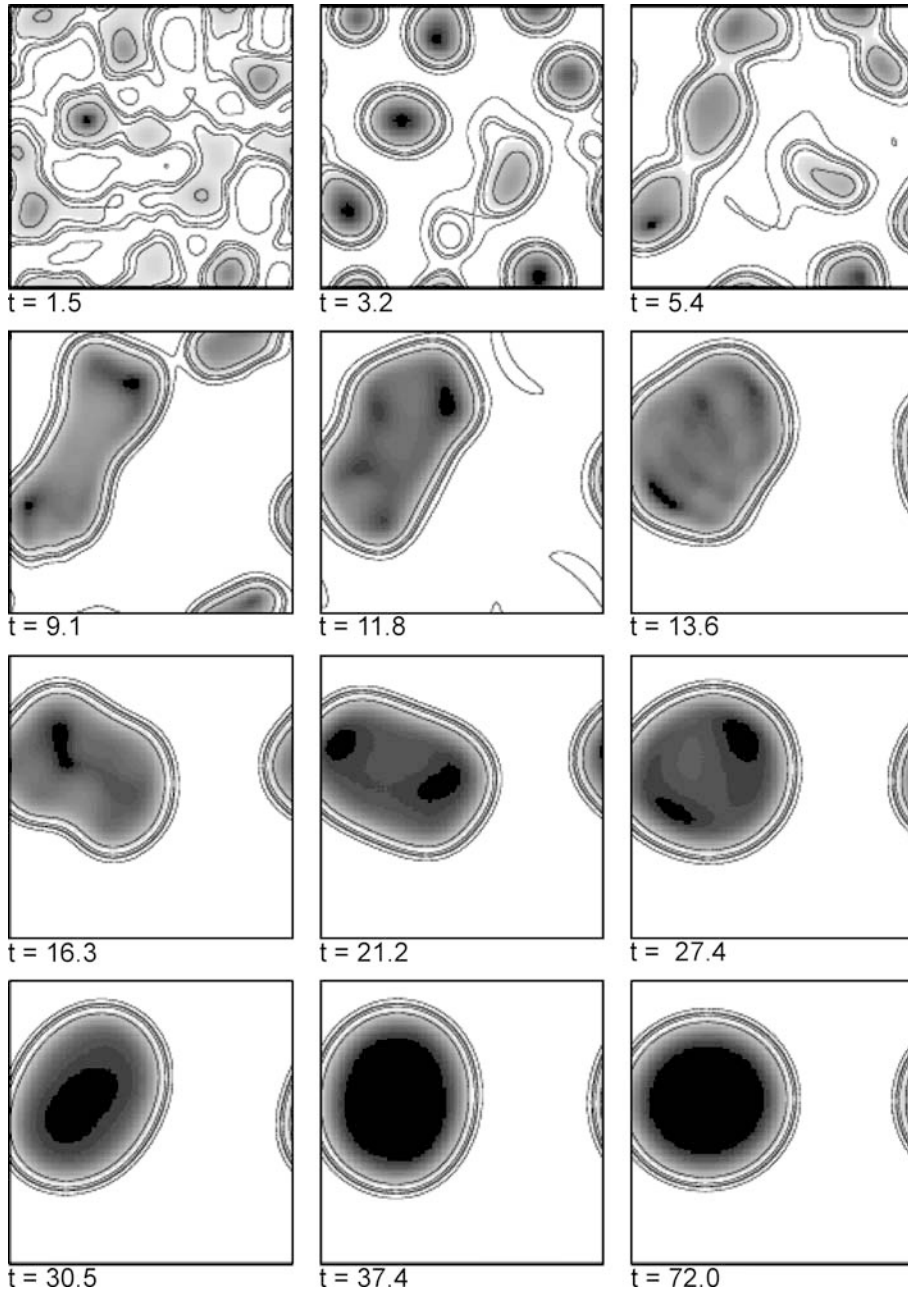
Fluid Dynamics, Pattern Formation, Figure 6
Time series from a numerical solution of (26) with artificial damping and bistable pressure according to (30). Coarsening dominates the nonlinear evolution and eventually a stationary circular region of surface depression (a hole) remains. Periodic boundary conditions in both horizontal directions have been used

of the layer. We assume that a motionless stationary state exists as a (stable or unstable) solution of (15) and of an equation such as (20) for the temperature. The temperature and pressure distribution of that state can then be computed from (15, 20) by putting $\vec{v}$ and all time derivatives to zero:

$$\nabla p^0 = \vec{f} \tag{33a}$$

$$\Delta T^0 = 0 . \tag{33b}$$
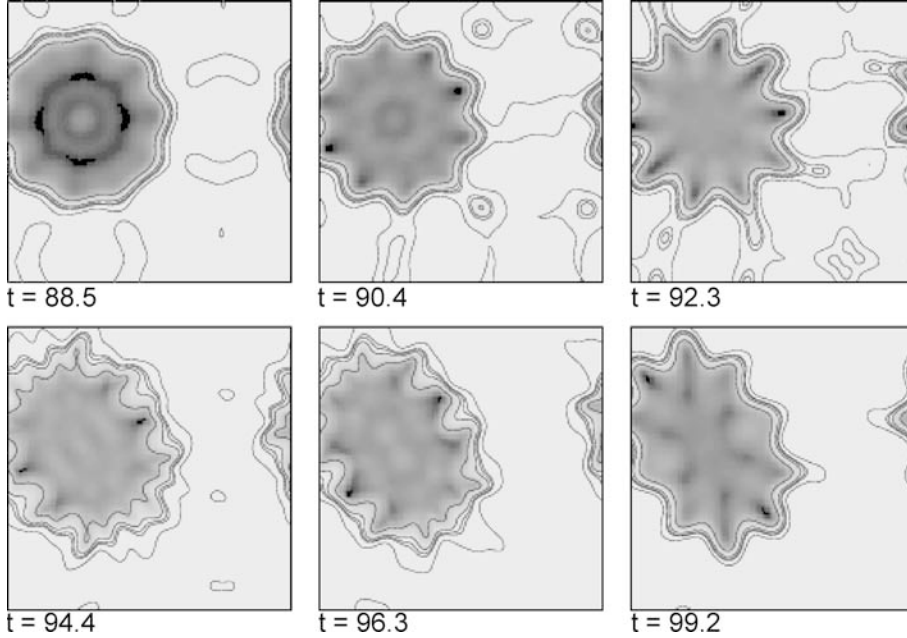
Fluid Dynamics, Pattern Formation, Figure 7
Continuation of the series of Fig. 6, but with additional parametric excitation according to (30) switched on at $t = 72$. Instead of stationary patterns pulsating stars are found

If an external force is provided by buoyancy, we may align the z-axis of the coordinate system along $\vec{f}$ which yields

$$\vec{f} = -g\,\rho(\vec{r})\,\hat{e}_z \,,$$

where g is the gravitational acceleration. Equation (33a) can be solved only if ρ does not depend on x and y. If one assumes that the density depends on temperature

$$\rho = \rho(T^0) \tag{34}$$

then T^0 can also depend only on z. Thus one finds from (33b)

$$T^0(z) = a + bz \,. \tag{35}$$

Taking a linear relation for (34)

$$\rho(T) = \rho_0[1 - \alpha(T - T_0)] \tag{36}$$

with the heat expansion coefficient $\alpha \equiv -\rho_0^{-1}\,d\rho/dT$ and ρ_0 as the density at the reference temperature T_0, one may integrate (33a) and find for the pressure of the motionless state (hydrostatic pressure)

$$p^0(z) = -g\int \rho dz = -g\rho_0\left(z - \frac{1}{2}\alpha\beta z^2\right) \tag{37}$$

where we put $a = T_0$ and $b = \beta$, in agreement with (32).

A linear stability analysis [32] shows that the motionless, nonequilibrium state (35) can become unstable if the temperature gradient β exceeds a certain critical value, depending on the fluid properties and the geometry of the layer. There are two different mechanisms, if the fluid layer is heated from below:

(1) *Buoyancy*: Hot fluid particles (volume elements) near the bottom are lighter than colder ones and want to rise. Colder particles near the top want to sink. If the stabilizing forces of thermal conduction and friction in the fluid are exceeded by the externally applied temperature gradient, patterned fluid motion sets in.

(2) *Surface tension*: If the upper surface of the fluid is free, that is, in contact with the ambient air, tangential surface tension normally increases with decreasing surface temperature (Fig. 8a–c). If a fluid particle near the surface moves by fluctuations, say, to the right, then warmer fluid is pulled up from the bottom, increasing the surface temperature locally. Due to laterally increasing surface tension with respect to the neighbored points, even more hot fluid is pumped up from the bottom and the fluid starts to move. This is called the Marangoni effect and works even without gravity, that is, in space experiments.

In both cases, the typical length of the structures which bifurcate from the motionless state is of the order of the layer depth. These instabilities are sometimes called *small scale instabilities*.
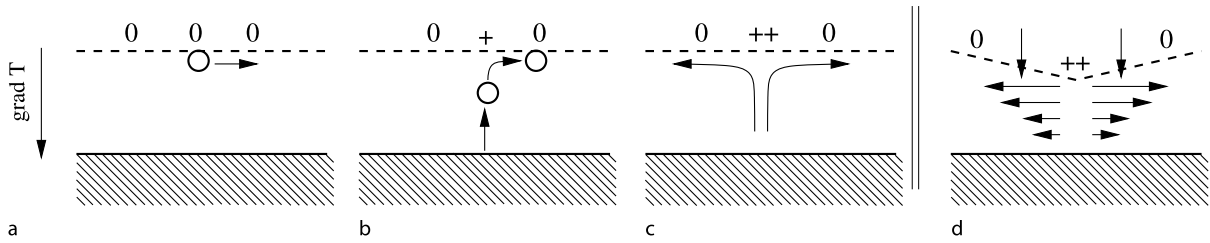
Fluid Dynamics, Pattern Formation, Figure 8
a–c The Marangoni effect may destabilize a fluid layer at rest and may generate a (regular) fluid motion. The surface remains flat (to a good approximation). If the surface is deformable **d**, a large scale instability may occur as a consequence of the Marangoni effect and mass conservation. For both instabilities, it is sufficient to assume the surface tension as a linear function of temperature. (+/-/0) denote relative temperatures
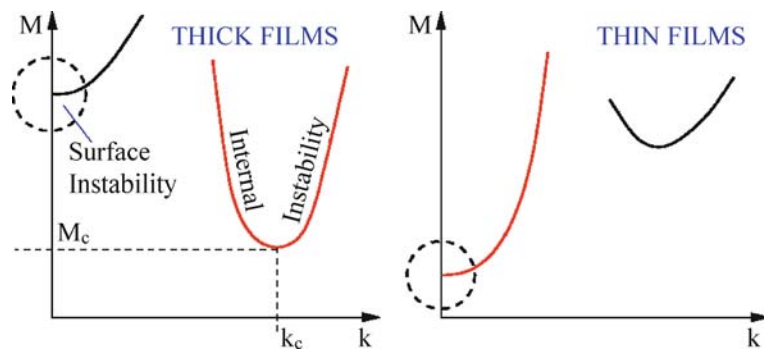
Fluid Dynamics, Pattern Formation, Figure 9
The two cases "thick films" and "thin films" are defined by the instability that comes first when the temperature gradient is increased. The two instabilities differ in the wavelength Λ (wave number $k = 2\pi/\Lambda$) of the growing structures

In the situation described above, the surface can be assumed to be flat and undeformable. Of course this is only an approximation, but valid for not-too-thin fluid layers and parameters not too far from threshold. If, on the other hand, the thickness of the fluid layer is less than a certain value which is on the order of 10^{-4} m for common silicone oils, another mechanism comes to the foreground. This mechanism is based on

Surface Deformation. If the surface is locally depressed by an arbitrary fluctuation, the depressed part is heated up due to the vertical temperature gradient. A lateral surface tension gradient is formed which pulls the liquid outside the depressed region (Fig. 8d). Since the continuity equation must hold, the surface becomes even more depressed and an instability occurs. The same mechanism leads to the growth of elevated parts of the surface, under which fluid is pumped in from adjacent regions [44,65,70]. The deformation mode belongs to the so-called *large-scale instability*. This means that the fastest growing modes have a wavelength that is very large compared to the layer depth. It is the depth of the layer which

distinguishes which instability occurs first if the temperature gradient is increased from the sub-critical region (Fig. 9).

In ultra-thin films (depth of few 100 nm or less), other mechanisms are possible. Van der Waals forces between the free surface and the solid substrate then become important. They have a potential and can be expressed in the pressure by an extra term, *disjoining pressure*, as already shown in Sect. "Surface Waves". If that pressure increases with decreasing layer depth, fluid is pressed out of depressed regions and pumped into elevated regions and an instability occurs, even for isothermal cases (Fig. 10).

Pattern Formation – Examples

What happens if the critical value for the temperature gradient is exceeded? Since the famous experiments of Henri Bénard [17] in the beginning of the 20th century, one knows that the fluid starts to move in form of hexagons if the surface is free and the layer is "thick" (Fig. 11). These kinds of experiments were repeated many times under ex-
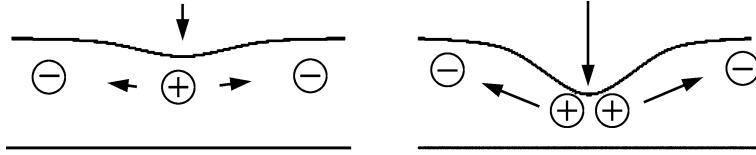
Fluid Dynamics, Pattern Formation, Figure 10
In ultra-thin films [79,85,92], van der Waals forces between free surface and solid substrate may destabilize a plane fluid layer even without an external temperature gradient (+/— denote relative values of the disjoining pressure)
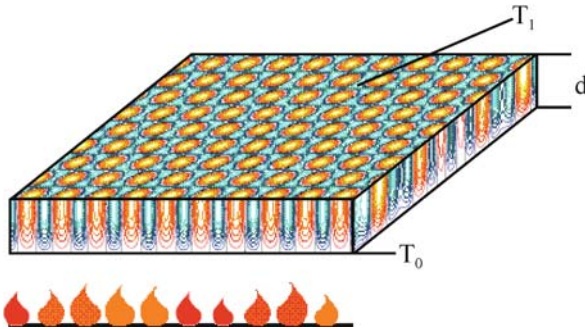
Fluid Dynamics, Pattern Formation, Figure 11
Hexagonal motion of a fluid heated from below, found by computer solution. Shown are contour lines of the temperature field (after [5])
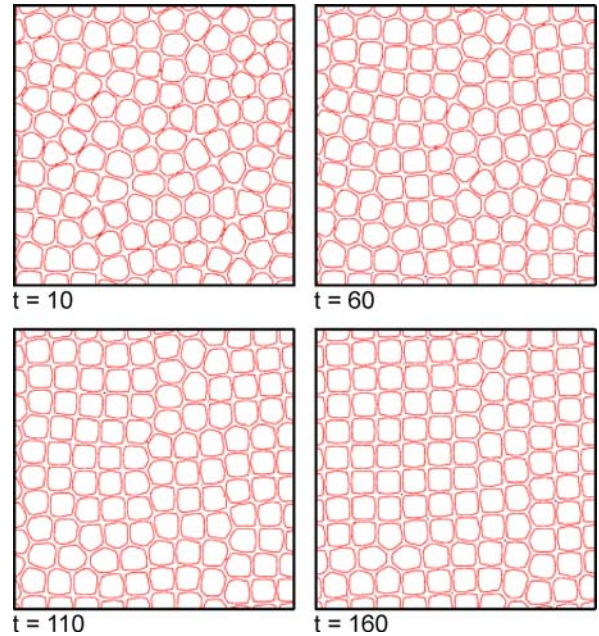
Fluid Dynamics, Pattern Formation, Figure 12
Regular squares as a secondary instability of hexagons. Numerical solution of the basic equations [6]

cellent conditions, for free and closed surfaces, with different fluids, even under micro gravity conditions [55,83].

Surprisingly, a secondary instability takes place for a larger external temperature gradient, which was not known before 1995, almost 100 years after Bénard. This instability shows the occurrence of rather regular squares and was discovered by Eckert and Thess in Dresden, Germany [6,41,67] and, in the meantime but independently, by Schatz and Swinney in Austin, Texas [80,81] (Fig. 12).

If the fluid is covered by a good thermal conductor (a sapphire plate, for instance), hexagons are not the typically found structure at onset, but rather stripes or rolls are encountered [90]. This can be understood in the frame of reduced order parameter equations by simple symmetry arguments. We shall discuss this in more detail in Sect. "Order Parameter Equations". For small Prandtl numbers (the ratio between viscosity and thermal diffusivity of the fluid) more complicated and time dependent patterns are found in the form of spirals (Fig. 13) [15,61,73].

The initial growth of patterns, with a certain horizontal length scale of the order of the depth of the fluid layer, is typical for pattern formation in thick films. In the long term, these structures can be stationary or time dependent, depending on several control and fluid parameters (temperature gradient, material properties, etc.). On the other

hand, the spatio-temporal behavior is completely different for thin and ultra-thin films. Here one finds, after a rather short initial phase, the formation of larger and larger structures, known as *coarsening*. Eventually, the dynamics converge to a stationary state that consists of a single elevation (drop) or suppression (hole) on the surface (Fig. 14 left panel). This development can be interrupted by rupture of the film. Rupture is obtained if the surface touches the substrate and the thickness reaches zero in certain domains. Rupture can be avoided by introducing a repelling disjoining pressure acting for a very small depth. In this situation, a completely dry region cannot exist but the substrate is, rather, covered by a so-called (ultra thin) precursor film [38,70], already proposed by Hardy in 1919 [51].

If, in addition, horizontal forces are applied, that is, by inclining the fluid layer, interesting studies of falling
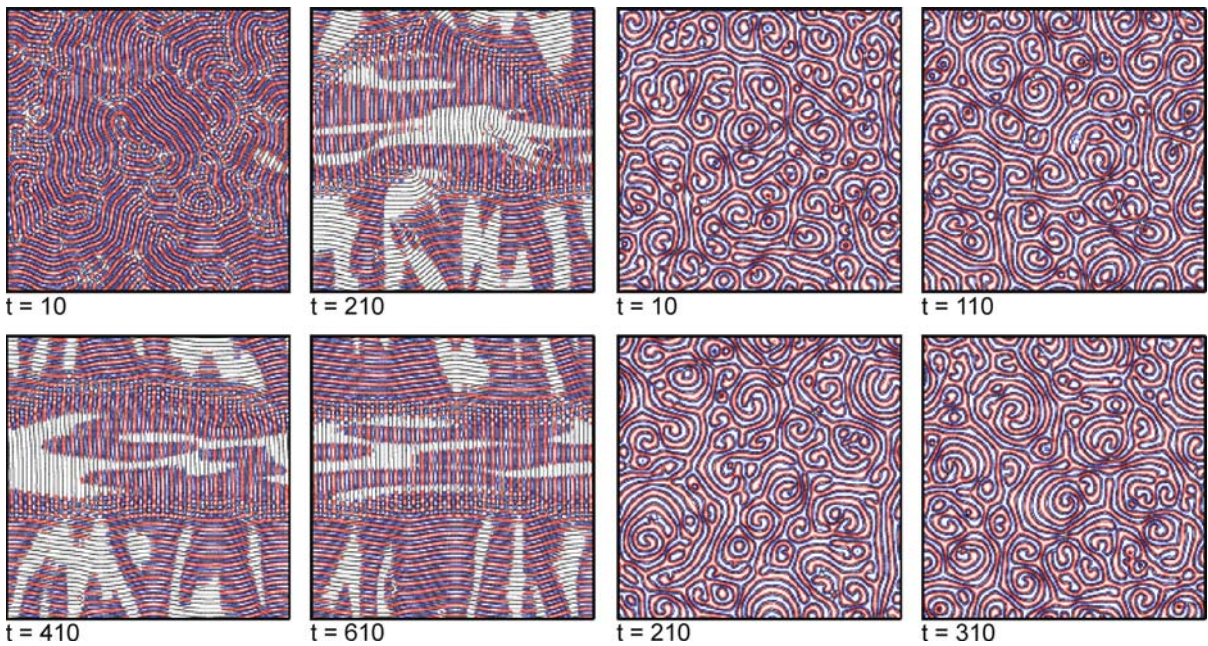
Fluid Dynamics, Pattern Formation, Figure 13
Left: **Rolls for high Prandtl number (Pr) fluids;** *Right*: **spirals for low Pr are found if the surface is covered by a good thermal conductor**
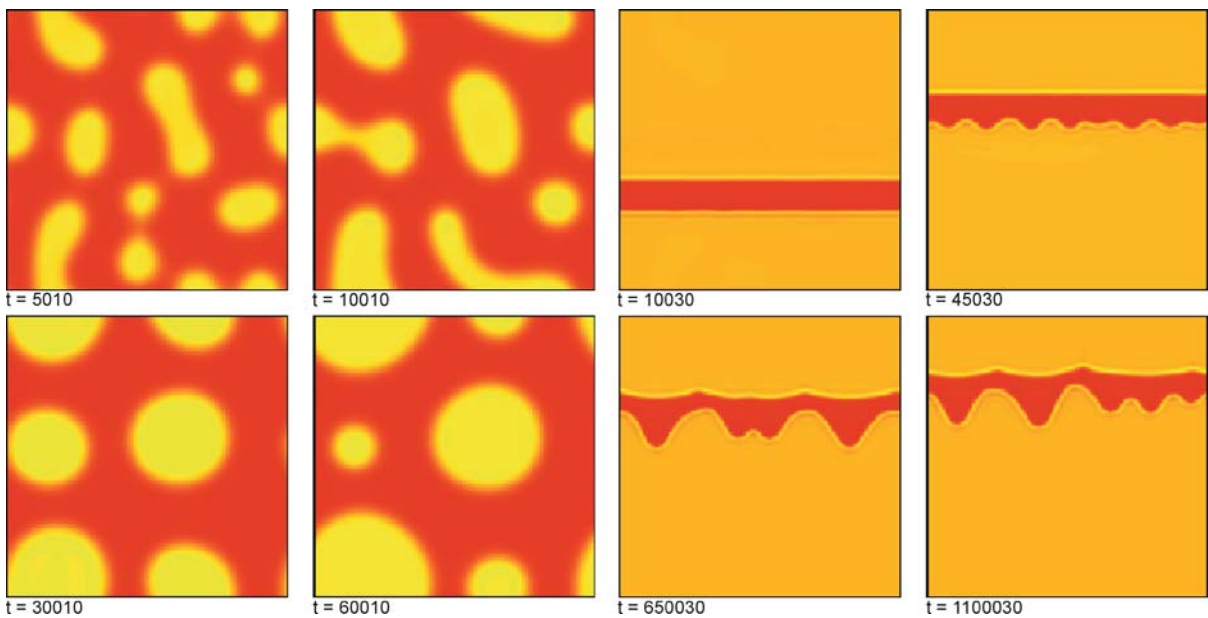
Fluid Dynamics, Pattern Formation, Figure 14
Numerical solution of the thin film equation (see Sect. "Conserved Order Parameter Fields"), *red*: **elevation, yellow: suppression.** *Left*:
Coarsening is the typical spatial behavior for a thin film. Finally, a stationary solution consisting of one single hole would survive.
Right: **If the layer is inclined, the motion of fronts and the development of front instabilities can be examined. From [16]**

films and front instabilities can be made in the frame of the thin film equation [13,82]. A typical example is shown in Fig. 14, right panel.

Types of Instabilities

Different types of instabilities can be classified according to their linear behavior at onset. Consider a mode having the complex eigenvalue

$$\lambda(k^2) = i\omega(k^2) + \sigma(k^2) \tag{38}$$

with real valued frequency ω and real valued growth rate σ. Due to rotation symmetry with respect to the horizontal coordinates, all values depend only on the modulus of the wave vector of the unstable mode (assumed as a plane wave in horizontal direction).

According to [36], we use the following notions:

Type III_s "s" denotes stationary or monotonic and refers to the temporal behavior of the unstable mode close to onset. The type number specifies the spatial behavior of the modes. Type III means slowly varying or even constant in space ($k \approx 0$). The spatial structure beyond instability is then mainly dominated by the geometry and boundary conditions of the system under consideration. For (38) this means

$$\omega = 0 \quad \text{and} \quad \left.\frac{d\sigma}{dk}\right|_{k=0} = 0 \,,$$

See Fig. 15. A typical example for a type III_s instability is the real Ginzburg–Landau equation. A computer solution clearly showing the spatially (and temporally) slowly varying behavior can be seen in Fig. 16.

Type III_o "o" stands for oscillatory and denotes a non-vanishing imaginary part of (38) at threshold. This type includes Hopf-instabilities which have the same slow spatial behavior as III_s. In (38) we have

$$\omega \neq 0 \quad \text{and} \quad \left.\frac{d\sigma}{dk}\right|_{k=0} = 0 \,.$$

For this kind of instability one needs at least two coupled diffusion equations. It is often encountered in reaction diffusion systems, as for instance the "Brusselator" [76,77].

Type I_s The short scale pattern forming instabilities shown in Figs. 11–13 with periodicity in space, $k_c \neq 0$ are of this type, see Fig. 15 middle frame. Again one needs at least two coupled diffusion equations to obtain such an instability. For the eigenvalue,

$$\omega = 0 \quad \text{and} \quad \left.\frac{d\sigma}{dk}\right|_{k=k_c} = 0 \quad \text{with} \ k_c \neq 0$$

holds. Sometimes these kinds of patterns are called *Turing structures* or *Turing instabilities*, after the seminal work of Alan Turing, who predicted this patterns in skin, scales, or hair coating of certain animals [89] (Fig. 17). For more details and pattern formation in biology see [63].

Type I_o Denotes oscillating Turing structures, sometimes also called *wave instabilities*. The eigenvalue λ then has the form

$$\omega \neq 0 \quad \text{and} \quad \left.\frac{d\sigma}{dk}\right|_{k=k_c} = 0 \quad \text{with} \ k_c \neq 0 \,.$$

For this instability, the system must be described by at least three coupled diffusion equations. In fluid mechanics, this kind of instability can be encountered in binary mixtures and give rise to a very complicated, in general chaotic, spatio-temporal behavior just at onset [9].

Type II_s This type is realized in the surface patterns of thin films, Fig. 14. Here, λ depends on k as shown in
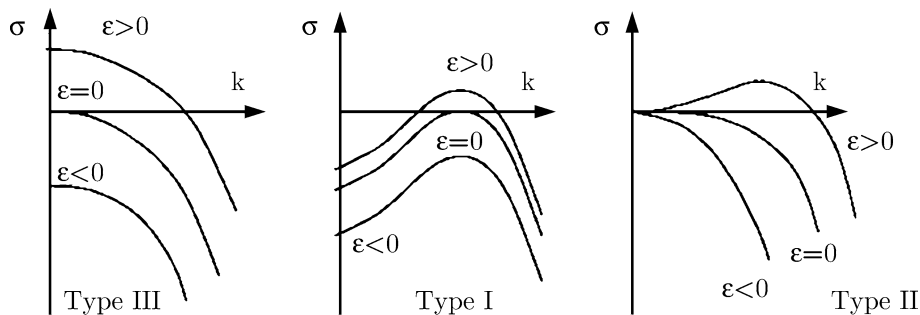
Fluid Dynamics, Pattern Formation, Figure 15
Schematic drawing of the real part of the eigenvalue (38) as function of the wave vector for the three types of instabilities
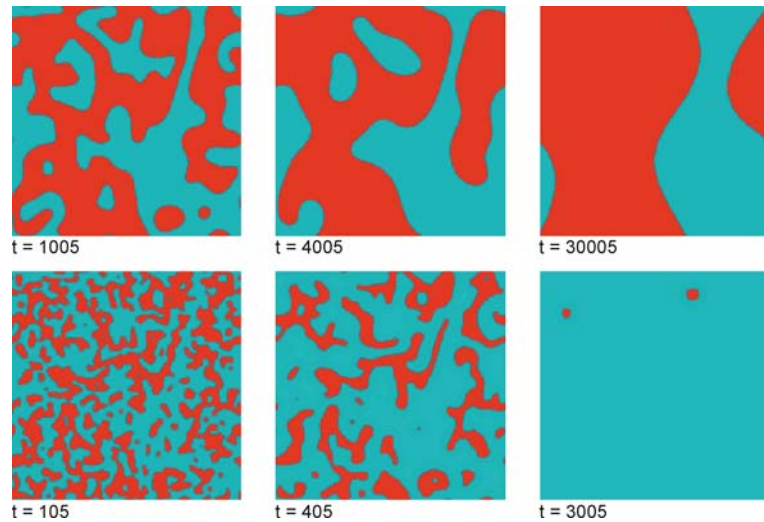
Fluid Dynamics, Pattern Formation, Figure 16
Numerical solution (time series) of the real Ginzburg–Landau equation. (39) which shows a III$_s$ instability. The Ginzburg–Landau equation can be considered as a simple model for the magnetization of a ferro magnet. Then the two rows show the spatio-temporal evolution of the magnetization, *Top*: without external field, *Bottom*: with external field

Fluid Dynamics, Pattern Formation, Figure 17
After a theory by A. Turing the painting on skin, scales, or coats of animals is organized by a nonequilibrium chemical reaction during the embryonal phase. *Left*: regular spots arranged in a hexagonal manner on the panther fish, *Right*: stripes with defects on the lion fish (pictures taken by the author in the Berlin Zoo)

the right frame of Fig. 15. One has

$$\omega = 0 \quad \text{and} \quad \frac{d\sigma}{dk}\bigg|_{k=k_c} = 0 \quad \text{with } k_c \neq 0$$

and in addition

$$\sigma(k = 0) = 0 \,.$$

From the last condition one sees that modes with $k = 0$, that is, those which are homogeneous in space, are marginally stable, meaning neither stable nor un-

stable. One may then add a constant to the order parameter (the mode amplitude)

$$\xi' = \xi + \text{const} \,,$$

where ξ' is still a solution of the linear part of the order parameter equation. This property usually has its origin in a symmetry of the basic problem. We shall discuss this instability type in Sect. "Conserved Order Parameter Fields" on thin films. There, the symmetry corresponds to a global shift of the surface in vertical direction.

Type II$_0$ The same as II$_s$ but with an additional imaginary part $\omega \neq 0$. We mention this type only for the sake of completeness; there will be no further examples in this contribution.

Order Parameter Equations

Order Parameters

In this section, we wish to describe pattern formation in the weakly nonlinear regime. We shall mainly restrict ourselves to the case of monotonic (non-oscillatory) instabilities. For further references see [10,12,36,47,50]. Close to a bifurcation point to a new state, it is natural to expand nonlinearities with respect to small deviations from the old, unstable state. These deviations can be written as a composition of certain Galerkin functions or modes; the amplitudes of these modes are called *order parameters*. If the order parameters are functions only of time, the dynamics given by the order parameter equations (ordinary differential equations, abbreviated: ODE) are perfect patterns, for instance parallel stripes, squares (two order parameters) or hexagons (three order parameters). Natural patterns having defects and grain boundaries, as for instance the structures shown in Figs. 11, 12 and 13 can also be described in this frame. One then has to make the additional assumption that the order parameters also vary (slowly) in space and are ruled by partial differential equations (abbreviated: PDE).

The Ginzburg–Landau Equation

A prominent (and historically the first) example of such a PDE order parameter equation is the Ginzburg–Landau equation [3,58]. In one spatial dimension it has the normal form

$$\partial_t \xi(x,t) = \varepsilon\, \xi(x,t) + q_0^2\, \partial_{xx}^2 \xi(x,t) - c_3 |\xi(x,t)|^2 \xi(x,t) \tag{39}$$

and describes the spatio-temporal evolution of the complex order parameter field ξ. If ξ is the mode amplitude of a roll structure with a certain wave number, for example, the critical one, then stripes with defects are obtained if ξ varies (slowly) in space. If c_3 and q_0 are real valued, (39) is called the real Ginzburg–Landau equation. For complex values of the coefficients, an incomparably richer and much more complicated spatio-temporal behavior of the order parameter is encountered, for details we refer to [3].

In the theory of nonequilibrium pattern formation, writing down an equation such as (39) is far from being purely phenomenological. It can be derived rather systematically from the basic hydrodynamic equations [48,66].

To give an idea of that, we do it briefly for the (two-dimensional) case of convection (the reader who is not interested in technical details can skip the rest of this section).

Starting point are the Eqs. (16) and (20) where in the latter, S stands for temperature T.

Scaling of independent $(\vec{r}, t)$ and dependent $(\vec{v}, T)$ variables allows the reduction of the numbers of parameters:

$$\vec{r} = \vec{\tilde{r}} \cdot d\,, \quad t = \tilde{t} \cdot (d^2/\kappa)\,, \quad \vec{v} = \vec{\tilde{v}} \cdot (\kappa/d)\,, \quad T = \tilde{T} \cdot \beta \cdot d\,, \tag{40}$$

with the constant depth d and the externally applied temperature gradient (32). Note that if the liquid is heated from below, $\beta < 0$. Introducing the deviation Θ from the thermally conducting state

$$T(\vec{r}, t) = T^0(z) + \Theta(\vec{r}, t) = T_0 + \beta z + \Theta(\vec{r}, t) \tag{41}$$

transforms (20) into

$$\{\Delta - \partial_t\} \Theta(\vec{r}, t) = -\Delta_2 \Psi(\vec{r}, t) + (\vec{v} \cdot \nabla)\Theta(\vec{r}, t) \tag{42}$$

and (16) into

$$\left\{\Delta - \frac{1}{\mathrm{Pr}}\partial_t\right\} \Delta_2 \Phi(\vec{r}, t) = -\frac{1}{\mathrm{Pr}} \left[\mathrm{curl}((\vec{v} \cdot \nabla)\vec{v})\right]_z \tag{43a}$$

$$\left\{\Delta - \frac{1}{\mathrm{Pr}}\partial_t\right\} \Delta \Delta_2 \Psi(\vec{r}, t)$$
$$= -R\, \Delta_2 \Theta(\vec{r}, t) - \frac{1}{\mathrm{Pr}} \left[\mathrm{curl}\,\mathrm{curl}((\vec{v} \cdot \nabla)\vec{v})\right]_z\,. \tag{43b}$$

Two dimensionless numbers occurred. One is the material dependent *Prandtl number*

$$\mathrm{Pr} = \frac{\nu}{\kappa}\,, \tag{44}$$

which measures the ratio of the diffusion times of heat and momentum. The other one is called the *Rayleigh number* and turns out to be

$$R = -\frac{\beta g \alpha d^4}{\nu \kappa} \tag{45}$$

with α defined in (36). The system (42), (43) constitutes the basic equations for the three scalar fields Φ, Ψ, and Θ which describe convective motion and temperature of a plane fluid layer with a flat and undeformable surface onto a plane substrate. This can be further simplified by taking the large Prandtl number limes $1/\mathrm{Pr} = 0$ (good for fluids with high viscosity, oils, etc.). Then Φ vanishes ev-

erywhere and only two equations are left:

$$\Delta^2 \Psi(\vec{r}, t) = -R\Theta(\vec{r}, t) \tag{46a}$$

$$\{\Delta - \partial_t\} \Theta(\vec{r}, t) = -\Delta_2 \Psi(\vec{r}, t) + (\vec{v} \cdot \nabla)\Theta(\vec{r}, t). \tag{46b}$$

A general nonlinear (2D) solution of Eqs. (46) may be expressed by

$$\begin{bmatrix} \Psi(x, z, t) \\ \Theta(x, z, t) \end{bmatrix}$$
$$= \sum_\ell \int_{-\infty}^{\infty} dk \xi_\ell(k, t) \begin{bmatrix} f_\ell(k^2, z) \\ g_\ell(k^2, z) \end{bmatrix} e^{-ikx} \tag{47}$$

and

$$\xi_\ell(k, t) = \xi_\ell^*(-k, t)$$

where f and g are eigenfunctions of the ODE eigenvalue problem

$$(d_z^2 - k^2)^2 f_\ell + R \, g_\ell = 0$$
$$(d_z^2 - k^2 - \lambda_\ell(k^2)) \, g_\ell - k^2 f_\ell = 0. \tag{48}$$

Here, ℓ labels the different eigenfunctions. Equation (48) is obtained by inserting (47) with $\xi_\ell \sim \exp(\lambda t)$ into (46) and keeping only linear terms. The functions f_ℓ and g_ℓ can be calculated numerically by a finite difference method in vertical direction where suitable boundary conditions must be implemented.

Inserting (47) into (46) yields, after multiplication with the adjoint function $g_\ell^+ \exp(ikx)$ and integration over the spatial coordinates, the system:

$$\partial_t \xi_\ell(k, t) = \lambda_\ell(k^2) \xi_\ell(k, t)$$
$$- \sum_{\ell' \ell''} \int_{-\infty}^{\infty} dk' dk'' c_{\ell\ell'\ell''}(kk'k'') \xi_{\ell'}(k', t) \xi_{\ell''}(k'', t)$$
$$\delta(k - k' - k''), \tag{49}$$

where the coefficients c are matrix elements that can be computed directly from the basic equations for any given set of control parameters:

$$c_{\ell\ell'\ell''}(kk'k'')$$
$$\equiv k'^2 \int_0^1 dz g_\ell^+(k^2, z) f_{\ell'}(k'^2, z) \partial_z g_{\ell''}(k''^2, z)$$
$$- k'k'' \int_0^1 dz g_\ell^+(k^2, z) g_{\ell'}(k'^2, z) \partial_z f_{\ell''}(k''^2, z). \tag{50}$$

Here we are still at the same level of complexity; the infinitely many degrees of freedom intrinsic in the basic

partial differential equations are expressed by an infinite number of mode amplitudes $\xi_\ell(k, t)$. To eliminate the fast damped modes by the linearly growing ones, we divide the eigenmodes into two groups:

$$\lambda_l \longrightarrow \begin{cases} \lambda_u(k^2) \approx 0 \Longrightarrow \xi_u(k, t), \\ \qquad\qquad |k| \approx k_c, u = \ell = 1 \\ \lambda_s(k^2) \ll 0 \Longrightarrow \xi_s(k, t), \\ \quad s = \ell > 1 \text{ or } s = \ell = 1 \text{ but } |k| \neq k_c. \end{cases} \tag{51}$$

In the following we may therefore substitute the index ℓ by u (unstable) or s (stable), depending on the values of ℓ and $|k|$. Now we express the amplitudes of the enslaved modes invoking an adiabatic elimination (k_c denotes the wave vector that maximizes λ_u). In this case, the dynamics of the enslaved modes are neglected, they follow instantaneously to the order parameters. This is a special case of the slaving principle of *synergetics*, which can be used in many other disciplines beyond hydrodynamics [47].

The remaining equations for the order parameters ξ_u, the amplitude equations, read (here and in the following we suppress the index u at ξ and λ):

$$\partial_t \xi(k, t) = \lambda(k^2) \xi(k, t)$$
$$+ \int dk' dk'' dk''' B(k, k', k'', k''') \xi(k', t) \xi(k'', t)$$
$$\xi(k''', t) \delta(k - k' - k'' - k''') \tag{52}$$

where $|k|, |k'|, |k''|, |k'''| \approx |k_c|$. Note that there are no quadratic expressions in ξ. This is because $k - k' - k''$ cannot vanish if all wave numbers have the same (nonzero) absolute value. In three spatial dimensions this is different. Three k-vectors can then form a resonant triangle, which is the reason why stable hexagons may occur.

The Landau coefficient B is directly related to the matrix elements (50):

$$B(k, k', k'', k''') = \sum_s \frac{1}{\lambda_s((k'' + k''')^2)}$$
$$c_{suu}(k'' + k''', k', k'') \big[c_{uus}(k, k', k'' + k''')$$
$$+ c_{usu}(k, k'' + k''', k') \big]$$

where the indices u and s are defined in (51).

To arrive at the Ginzburg–Landau equation, one must transform back to real space. If we express the δ-function in (52) as

$$\delta(k - k' - k'' - k''') = \frac{1}{2\pi} \int dx \, e^{i(k - k' - k'' - k''')x}$$

and assume, that the coefficient B does not depend much on k (it can be evaluated at $k = \pm k_c$), the cubic part of (52) takes the form

$$\frac{\bar{B}}{2\pi} \int \mathrm{d}x e^{ikx} \int \mathrm{d}k' \xi(k', t) \, e^{-ik'x} \int \mathrm{d}k'' \xi(k'', t)$$
$$e^{-ik''x} \int \mathrm{d}k''' \xi(k''', t) \, e^{-ik'''x}$$
$$= \frac{\bar{B}}{2\pi} \int \mathrm{d}x \, e^{ikx} \Psi^3(x, t) \quad (53)$$

where we have introduced the Fourier transform

$$\Psi(x, t) = \int \mathrm{d}k \, \xi(k, t) \, e^{-ikx} \,. \quad (54)$$

Inserting (53) into (52), multiplying with $e^{-ik\tilde{x}}$ and integrating over k yields the order parameter equation in real space

$$\partial_t \Psi(\tilde{x}, t) = \int \mathrm{d}k \lambda(k) \xi(k, t) e^{-ik\tilde{x}} + \bar{B} \, \Psi^3(\tilde{x}, t) \,. \quad (55)$$

If we replace the k^2-dependence of λ under the integral by $-\partial_{\tilde{x}\tilde{x}}^2$, we may pull λ out of the integral and write (55) in the form

$$\partial_t \Psi(x, t) = \lambda(-\partial_{xx}^2) \, \Psi(x, t) + \bar{B} \, \Psi^3(x, t) \,. \quad (56)$$

The function $\Psi(x, t)$ can also be called an "order parameter", though it is not slowly varying in space compared to the small scale structure of the rolls, an idea which we shall work out in the following section. One big advantage can already be seen: the reduction of the number of space dimensions by one. We started with the hydrodynamic equations in two dimensions and get an order parameter equation in only one spatial dimension.

To find the form of the Ginzburg–Landau equation, we must introduce a slowly varying order parameter. This is done by recalling that the Fourier transform of Ψ is mainly excited around $k = \pm k_c$. Then it is natural to make a "rotating wave approximation" with respect to x of the form

$$\Psi(x, t) = \xi(x, t) \, e^{ik_c x} + \xi^*(x, t) \, e^{-ik_c x} \,. \quad (57)$$

Inserting this into (56), multiplying by $e^{-ik_c x}$ and integrating with respect to x over one period $2\pi/k_c$ yields with the assumption of constant (slowly varying) ξ in this period

$$\partial_t \xi(x, t) = \lambda(-(\partial_x + ik_c)^2) \, \xi(x, t) + 3\bar{B} \, |\xi(x, t)|^2 \xi(x, t).$$

The last approximation is concerned with the evaluation of the eigenvalue in form of a differential operator.

Close to k_c, it has the form of a parabola, see Fig. 15 middle frame. Thus we may approximate

$$\lambda(k^2) = \varepsilon - q^2 (k^2 - k_c^2)^2 \quad (58)$$

and also

$$\lambda(-(\partial_x + ik_c)^2) = \varepsilon - q^2((\partial_x + ik_c)^2 + k_c^2)^2$$
$$= \varepsilon - q^2(\partial_{xx}^2 + 2ik_c\partial_x)^2$$
$$\approx \varepsilon + 4q^2 k_c^2 \partial_{xx}^2 \,. \quad (59)$$

For the last conversion, we neglect higher derivatives, which is justified due to the slowly varying spatial dependence of ξ. After scaling of ξ and the additional assumption $\bar{B} < 0$ we finally have derived the Ginzburg–Landau equation (39).

The Swift–Hohenberg Equation

In two spatial dimensions, the drawback of the Ginzburg–Landau equation is its lack of rotational symmetry. Therefore, it is better to pass on the rotating wave approximation (57) and to consider instead the fully space-dependent function Ψ as an order parameter, but now in two spatial dimensions. The resulting evolution equation in its lowest nonlinear approximation is the Swift–Hohenberg equation [88]

$$\dot{\Psi}(\vec{x}, t) = \left[\varepsilon - (1 + \Delta_2)^2\right] \Psi(\vec{x}, t) - \Psi^3(\vec{x}, t) \,, \quad (60)$$

which we shall derive now.

Non-local Order Parameter Equations To this end we go back to (52) and write it down in two dimensions, now including the quadratic terms ($\vec{k} = (k_x, k_y)$):

$$\partial_t \xi(\vec{k}, t) = \lambda(k^2) \xi(\vec{k}, t) + \int \mathrm{d}^2\vec{k}' \mathrm{d}^2\vec{k}'' A(\vec{k}, \vec{k}', \vec{k}'')$$
$$\cdot \xi(\vec{k}', t) \xi(\vec{k}'', t) \delta(\vec{k} - \vec{k}' - \vec{k}'') + \int \mathrm{d}^2\vec{k}' \mathrm{d}^2\vec{k}'' \mathrm{d}^2\vec{k}'''$$
$$\cdot B(\vec{k}, \vec{k}', \vec{k}'', \vec{k}''') \xi(\vec{k}', t) \xi(\vec{k}'', t) \xi(\vec{k}''', t)$$
$$\cdot \delta(\vec{k} - \vec{k}' - \vec{k}'' - \vec{k}''') \,. \quad (61)$$

Introducing the (2D) Fourier transform ($\vec{x} = (x, y)$),

$$\Psi(\vec{x}, t) = \int \mathrm{d}^2\vec{k} \, \xi(\vec{k}, t) \, e^{-i\vec{k}\vec{x}} \,. \quad (62)$$

and transforming (61) to real space yields the integro-differential equation

$$\partial_t \Psi(\vec{x}, t) = \lambda(\Delta)\Psi(\vec{x}, t)$$
$$+ \iint d^2\vec{x}' \, d^2\vec{x}'' \, G^{(2)}(\vec{x} - \vec{x}', \vec{x} - \vec{x}'')\Psi(\vec{x}', t)\Psi(\vec{x}'', t)$$
$$+ \iiint d^2\vec{x}' d^2\vec{x}'' d^2\vec{x}''' \, G^{(3)}(\vec{x} - \vec{x}', \vec{x} - \vec{x}'', \vec{x} - \vec{x}''')$$
$$\Psi(\vec{x}', t)\Psi(\vec{x}'', t)\Psi(\vec{x}''', t)$$
$$(63)$$

where the kernels are computed by the Fourier transforms:

$$G^{(2)}(\vec{x}, \vec{x}')$$
$$= \frac{1}{16\pi^4} \int d^2\vec{k}d^2\vec{k}' A(\vec{k} + \vec{k}', \vec{k}, \vec{k}')e^{-i\vec{k}\vec{x}} e^{-i\vec{k}'\vec{x}'},$$
$$G^{(3)}(\vec{x}, \vec{x}', \vec{x}''')$$
$$= \frac{1}{64\pi^6} \int d^2\vec{k}d^2\vec{k}'d^2\vec{k}'' B\left(\vec{k} + \vec{k}' + \vec{k}'', \vec{k}, \vec{k}', \vec{k}''\right)$$
$$\cdot e^{-i\vec{k}\vec{x}} e^{-i\vec{k}'\vec{x}'} e^{-i\vec{k}''\vec{x}''}.$$
$$(64)$$

Gradient Expansion Although Eq. (63) has a rather general form, its further numerical treatment is not practicable, at least not in two dimensions. Each integral must be approximated somehow as a sum over mesh points. The cubic coefficients would result in a 6-fold sum with, if N is the number of mesh points, N^6 summands, which is, if N is around the size of 100, rather hopeless.

On the other hand, the excitation of ξ mainly close to k_c, in two dimensions on a (narrow) ring in Fourier space with radius k_c, makes it natural to expand Ψ under the integrals around $\vec{x}$. This works well if the kernels (64) have a finite (small) range with significant contribution only for $|\vec{x} - \vec{x}'| < \Lambda$ with $\Lambda = 2\pi/k_c$.

To save space we demonstrate the method only for the quadratic term of (63) and in one spatial dimension. A Taylor expansion of Ψ leads to

$$\iint dx' \, dx'' G^{(2)}(x - x', x - x'')$$
$$\sum_{m,n=0}^{\infty} \frac{1}{m!n!} \frac{\partial^m \Psi}{\partial x^m} \frac{\partial^n \Psi}{\partial x^n} (x - x')^m (x - x'')^n,$$

where the derivatives must be evaluated at x. They can be written in front of the integrals, yielding

$$\sum_{m,n=0}^{\infty} g_{mn}^{(2)} \frac{\partial^m \Psi}{\partial x^m} \frac{\partial^n \Psi}{\partial x^n} \qquad (65)$$

with the moments

$$g_{mn}^{(2)} = \frac{1}{m!n!} \iint dx_1 \, dx_2 \, G^{(2)}(x_1, x_2) \, x_1^m x_2^n \,.$$

A similar expression can be found for the cubic coefficient. A series of the form (65) is called *gradient expansion*. In this way, a local order parameter equation results, but which now has infinitely many nonlinear terms. It reads

$$\partial_t \Psi = \lambda(\Delta)\Psi + \sum_{m,n=0}^{\infty} g_{mn}^{(2)} \frac{\partial^m \Psi}{\partial x^m} \frac{\partial^n \Psi}{\partial x^n}$$
$$+ \sum_{\ell,m,n=0}^{\infty} g_{\ell mn}^{(3)} \frac{\partial^\ell \Psi}{\partial x^\ell} \frac{\partial^m \Psi}{\partial x^m} \frac{\partial^n \Psi}{\partial x^n} \quad (66)$$

with

$$g_{\ell mn}^{(3)}$$
$$= \frac{1}{\ell!m!n!} \iiint dx_1 \, dx_2 \, dx_3 \, G^{(3)}(x_1, x_2, x_3) x_1^\ell x_2^m x_3^n.$$

For more details see [10].

Swift–Hohenberg–Haken Equation The series in (66) will converge rapidly if the kernels have a short range. Here we consider only the extreme case of δ-shaped kernels, now in two dimensions:

$$G^{(2)}(\vec{x}_1, \vec{x}_2) = A \cdot \delta(\vec{x}_1)\delta(\vec{x}_2),$$
$$G^{(3)}(\vec{x}_1, \vec{x}_2, \vec{x}_3) = B \cdot \delta(\vec{x}_1)\delta(\vec{x}_2)\delta(\vec{x}_3).$$

All coefficients vanish, except $g_{00}^{(2)}$ and $g_{000}^{(3)}$. Then (66) simplifies to

$$\partial_t \Psi(\vec{x}, t) = \lambda(\Delta)\Psi(\vec{x}, t) + A \, \Psi^2(\vec{x}, t) + B \, \Psi^3(\vec{x}, t). \quad (67)$$

For the linear part we again use the expansion (58) and replace k^2 by $-\Delta$. After rescaling of length, time and Ψ, (67) turns into the canonical form

$$\dot{\Psi}(\vec{x}, t) = \left[\varepsilon - (1 + \Delta_2)^2\right] \Psi(\vec{x}, t) + a \, \Psi^2(\vec{x}, t) - \Psi^3(\vec{x}, t) \quad (68)$$

with

$$a = \frac{A}{\sqrt{-B}}.$$

Equation (68) is the Swift–Hohenberg–Haken equation derived first using the theoretical methods of synergetics by Haken [11,49].

Numerical solutions of (68) with $a = 0$ are shown in Fig. 18. Stripes as known from convection, but also from Turing instabilities, can be clearly seen. If $|a|$ exceeds a certain value which depends on $\sqrt{\varepsilon}$, hexagonal structures are found which agree qualitatively with those obtained in Bénard–Marangoni convection (Fig. 19). It can be shown that the symmetry break $z \to -z$ caused by the different vertical boundary conditions on top and bottom of the fluid gives rise to a (positive) quadratic coefficient. In the Swift–Hohenberg equation, this violates the symmetry $\Psi \to -\Psi$ and may stabilize two different sorts of hexagons, namely the already mentioned ℓ- and g-hexagons. The first ones are found for large enough positive a, the latter for negative a.
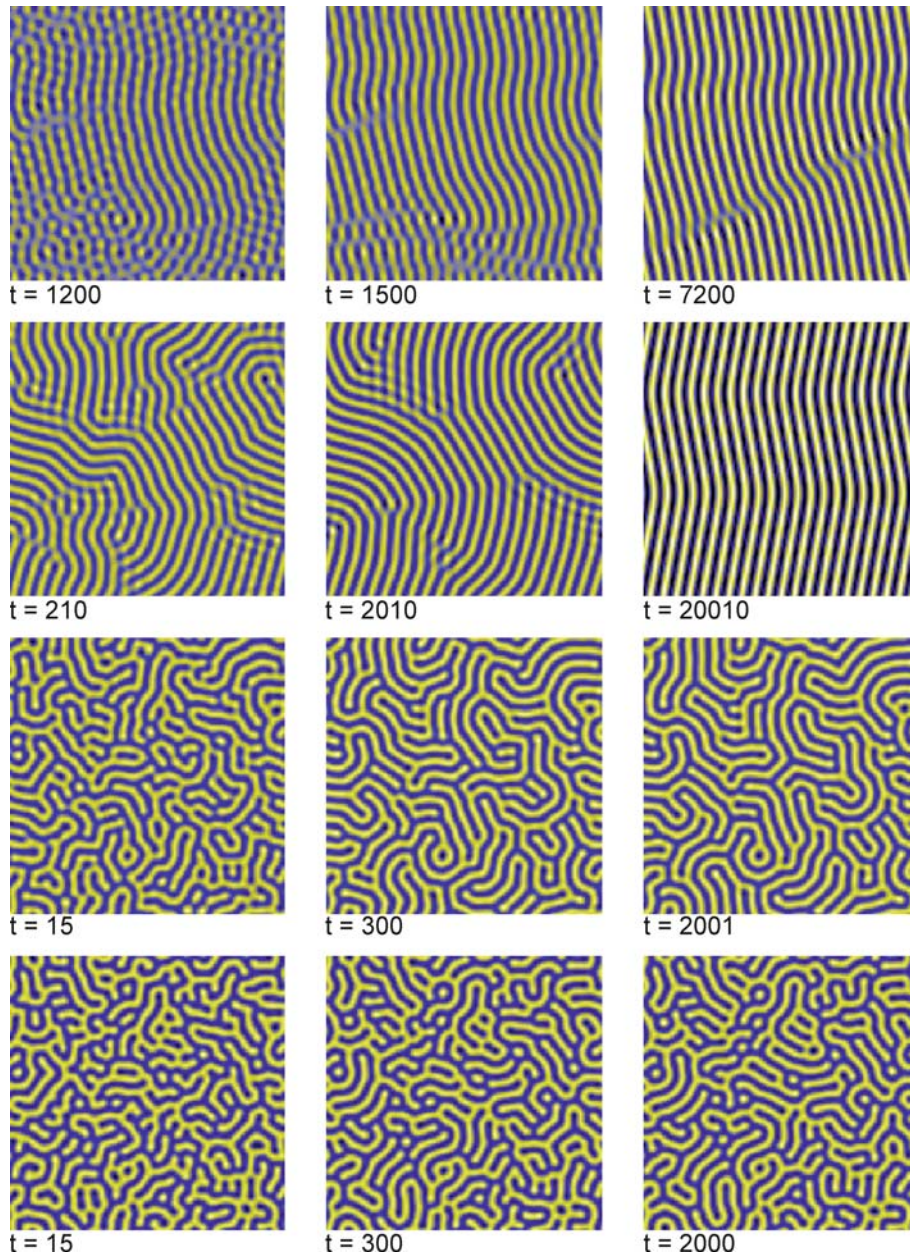
Fluid Dynamics, Pattern Formation, Figure 18
Computer solutions of the Swift–Hohenberg Eq. (60) for several $\varepsilon = 0.01, 0.1, 1.0, 2.0$ (*top to bottom*). The evolution time scales with $1/\varepsilon$, the number of defects increases with ε
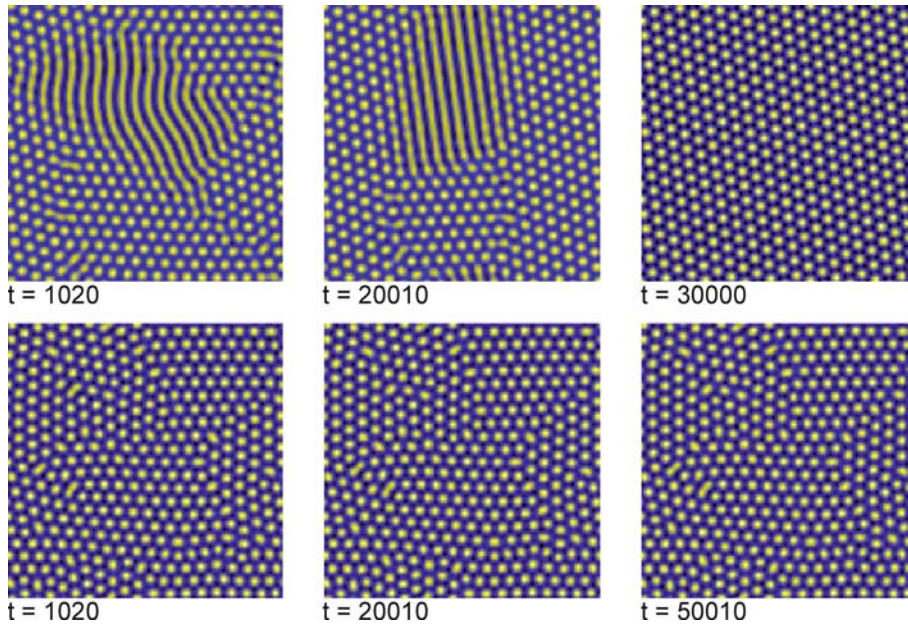
Fluid Dynamics, Pattern Formation, Figure 19
Evolution of a random dot initial condition from (68) with $\varepsilon = 0.1$, $a = 0.26$ (*top*) and $a = 1.3$ (*bottom*). For a in the bistable region, top row, stripes and hexagons coexist for a long time until hexagons win. *Bottom*: for rather large a hexagons are formed soon showing many defects and grain boundaries. The defects survive for quite a long time

The Swift–Hohenberg equation can be considered as normal form of type I_s instabilities. The bifurcation scenario is general (Fig. 20): hexagons are the generic form at onset if symmetry breaking (quadratic) terms occur, which is normal. Even very small symmetry breaking effects lead to hexagons, although their stability region will decrease and finally shrink to the critical point $\varepsilon = 0$ if $a \rightarrow 0$. Well above threshold, stripes are expected – or squares.

Squares A linear stability analysis of the Swift–Hohenberg Eq. (68) shows that squares are always unstable in favor of rolls (or hexagons). Therefore there exists no stable square pattern as a solution. This can be changed including higher order terms in the gradient expansion (66), for details see [14]. In this spirit, the equation

$$\partial_t \Psi = \varepsilon\, \Psi - (\Delta + 1)^2 \Psi - b\, \Psi^3 - c\, \Psi\, \Delta^2(\Psi^2) \quad (69)$$

has a stable square solution for $-32c/9 < b < 0$. In Fig. 21 we present numerical solutions of (69) for two different values of the parameter b.

Regular squares are found in convection experiments with two poorly conducting top and bottom plates [28] or in binary mixtures with a certain mean concentra-
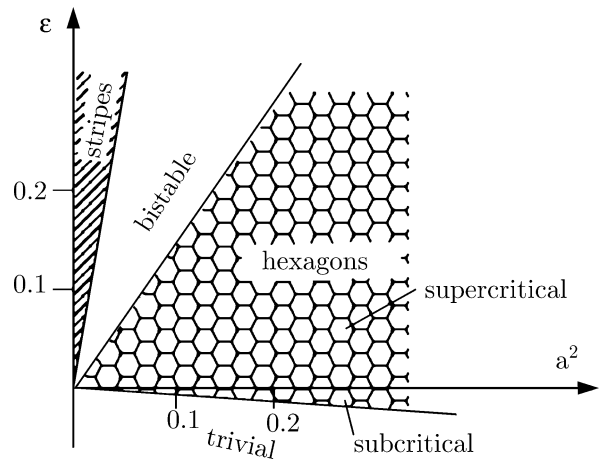
Fluid Dynamics, Pattern Formation, Figure 20
Stability regions in the parameter plane of Eq. (68). Hexagons bifurcate subcritically from the trivial solution $\Psi = 0$. As a secondary instability, stripes emerge. The transition hexagons-stripes as well as trivial sol-hexagons both show hysteresis

tion [62]. If the fluid viscosity is strongly temperature dependent (non-boussinesq effects), squares are also preferred, as shown in [27]. For all these cases, an equation of the form (69) can be approximately derived close to onset.
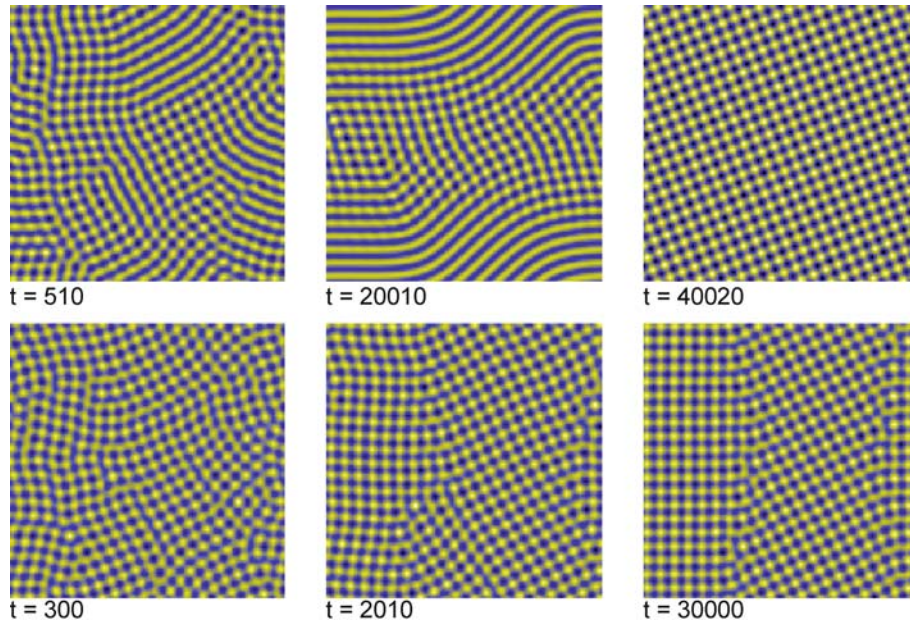
Fluid Dynamics, Pattern Formation, Figure 21
Numerical solutions of (69) for $\varepsilon = 0.1$, $c = 1/16$ and $b = 0$ *(top)*, $b = -0.1$ *(bottom)*. For $b = 0$ both squares and stripes are stable. After a longer time squares win the competition. *Bottom*: clearly in the square region of parameter space. Squares are formed soon having many defects and grain boundaries. Finally, a rather regular square pattern evolves

Conserved Order Parameter Fields

In the previous section, the OPE had the general form

$$\partial_t \xi(\vec{r}, t) = F(\xi, \nabla \xi, \Delta \xi) \tag{70}$$

with no further restrictions (except of boundary conditions) for the order parameter field ξ. However, there are many cases where the physical meaning of the order parameter is that of a density belonging to a conserved quantity such as total mass, volume or charge. Let ξ be such a density; then the mean value

$$M = \langle F \rangle \equiv \frac{1}{V} \int_V d^3\vec{r}\, F(\xi, \nabla \xi, \Delta \xi) \tag{71}$$

should vanish, if $\langle \xi \rangle$ is a conserved quantity in the constant volume V. Then F can be written as

$$F(\xi, \nabla \xi, \Delta \xi) = -\operatorname{div} \vec{j}(\vec{r}, t) \tag{72}$$

if the total flow of the current density $\vec{j}$ through the surface A of V vanishes

$$\oint_{A(V)} d^2\vec{f} \cdot \vec{j}(\vec{r}, t) = 0 . \tag{73}$$

With (72), Eq. (70) takes the form of a continuity equation. In this section we wish to consider OPEs that fulfil (72) and (73).

Thin Films

Consider a fluid with a free and deformable surface located at $z = h(x, y, t)$ as already shown in Fig. 3. If the fluid is incompressible and there is no flow through the sidewalls, the total volume of the fluid layer

$$A \cdot \langle h \rangle = \int_A dx\, dy\, h(x, y, t) \tag{74}$$

is a conserved quantity, where A is the base area of the layer. As a consequence, the evolution equation for h must have the form

$$\partial_t h = -\operatorname{div} \vec{j} = -\partial_x j_x - \partial_y j_y . \tag{75}$$

Comparing (75) with the kinematic boundary conditions (23) and taking $v_z|_{z=h}$ from the integral of the incompressibility condition (3)

$$v_z|_{z=h} = -\int_0^h dz\, (\partial_x v_x + \partial_y v_y) + v_z|_{z=0}$$

one finds with $v_z|_{z=0} = 0$

$$\vec{j} = \int_0^h dz\, \vec{v}_H , \tag{76}$$

where $\vec{v}_H$ denotes the two horizontal velocity components.

The Lubrication Approximation   To close the Eqs. (75), (76), it is necessary to compute $\vec{v}_H$ as a function of h. For thin films, the Reynolds number is small and the Stokes equation (19) determines the fluid velocity to a good approximation. Using scaling [70]

$$x = \tilde{x} \cdot \ell , \;\; y = \tilde{y} \cdot \ell , \;\; z = \tilde{z} \cdot \mathrm{d} , \;\; t = \tilde{t} \cdot \tau , \;\; h = \tilde{h} \cdot \mathrm{d},$$
(77)

(19) turns into

$$\left(\delta^2 \left(\partial_{\tilde{x}\tilde{x}}^2 + \partial_{\tilde{y}\tilde{y}}^2 \right) + \partial_{\tilde{z}\tilde{z}}^2 \right) \vec{v}_H = \tilde{\nabla}_2 \tilde{P}$$
(78a)

$$\delta^2 \left(\delta^2 \left(\partial_{\tilde{x}\tilde{x}}^2 + \partial_{\tilde{y}\tilde{y}}^2 \right) + \partial_{\tilde{z}\tilde{z}}^2 \right) \tilde{v}_z = \partial_{\tilde{z}} \tilde{P} .$$
(78b)

with the 2D-gradient $\nabla_2 = (\partial_x, \partial_y)$. In (78) we have introduced the dimensionless velocity and pressure

$$\vec{v}_H = \vec{\tilde{v}}_H \cdot \frac{\ell}{\tau} , \quad v_z = \tilde{v}_z \cdot \frac{d}{\tau} , \quad P = \tilde{P} \cdot \frac{\eta}{\delta^2 \tau}$$

and $\delta = d/\ell$ as a small parameter already defined in (24). In the limit $\delta \to 0$ it follows from (78b)

$$\partial_{\tilde{z}} \tilde{P} = 0 \quad \text{or} \quad \tilde{P} = \tilde{P}(\tilde{x}, \tilde{y}) .$$

Thus one can integrate (78a) twice over $\tilde{z}$ and finds with the no-slip condition $\vec{\tilde{v}}_H(0) = 0$

$$\vec{\tilde{v}}_H(\tilde{x}, \tilde{y}, \tilde{z}) = \vec{f}(\tilde{x}, \tilde{y}) \cdot \tilde{z} + \frac{1}{2} (\tilde{\nabla}_2 \tilde{P}(\tilde{x}, \tilde{y})) \cdot \tilde{z}^2$$
(79)

with a function $\vec{f}(\tilde{x}, \tilde{y})$ which can be determined by the boundary conditions. To this end we consider an inhomogeneous surface tension (caused, for example, by a temperature gradient) at the free surface, which yields the condition

$$\eta \; \partial_z \vec{v}_H|_{z=h} = \nabla_2 \Gamma|_{z=h} .$$

Inserting (79) there one finds

$$\vec{f} = \tilde{\nabla}_2 \tilde{\Gamma} - (\tilde{\nabla}_2 \tilde{P}) \cdot \tilde{h}$$

with the non-dimensional surface tension

$$\tilde{\Gamma} = \Gamma \; \frac{\tau d}{\eta \ell^2} .$$

Inserting everything into (76) and integrating by $\tilde{z}$ finally yields (all tildes omitted)

$$\partial_t h = -\nabla_2 \cdot \left[-\frac{h^3}{3} \nabla_2 P + \frac{h^2}{2} \nabla_2 \Gamma \right] .$$
(80)

This is the basic equation for the evolution of the surface of a thin film in the so-called lubrication approxima-

tion [68]. Equation (80) is sometimes denoted as the *thin film equation* [70,92].

The Disjoining Pressure for Ultra–thin Films   Gravitation and surface tension can be included into the pressure P as already outlined in Sect. "Surface Waves". They both stabilize the flat film. On the other hand, an instability mechanism is encountered in very thin (ultra-thin) films where the thickness is some 100 nm or even less [52,79,84]. Then, van der Waals forces between free surface and solid substrate can no longer be neglected [52]. For an attractive force between surface and substrate one has

$$d_h P < 0 .$$

But there can also exist a repelling van der Waals force with $d_h P > 0$ which stabilizes the flat surface. Attractive and repelling forces have different ranges. Usually, the repelling force is short range, the attractive one long range. Then, the initially "thick" film can be unstable due to attraction but rupture is avoided by repulsion. In this way completely dry regions cannot exist but the substrate always remains covered by an extremely thin film (some nm), called *precursor film*, Fig. 22 [51].

The complete expression for such an attractive/repulsive disjoining pressure including gravity and surface tension would be (Fig. 23)

$$P(h) = \frac{A_3}{h^3} - \frac{A_9}{h^9} + Gh - q \, \Delta_2 \, h$$
(81)

where A_3 and A_9 are material parameters, the Hamaker constants, and

$$G = \frac{d^3 g \tau}{\ell^2 \nu} , \quad q = \Gamma \; \frac{\tau d^3}{\ell^4 \eta}$$

denote the dimensionless gravitation number and the surface tension, respectively.

Spinodal Dewetting – Numerical Results   If a thin liquid film is exposed to a non- or partially wetting substrate, a small perturbation is sufficient to destabilize the flat surface. The fluid then bubbles and many small drops are formed. This phenomenon can be seen for instance if rain falls on a waxed cloth or on a well polished car roof. Such a process is called *spinodal dewetting* and refers to the unstable region of Fig. 23, [13,86]. As already explained in Sect. "Instabilities", an instability of the flat film occurs in the region where P has a negative slope. This instability is of type II, as is shown in Fig. 23, right frame, and has the growth rate (dispersion relation)

$$\lambda = \frac{1}{3} h_0^3 \left(-D(h_0) k^2 - k^4 \right)$$
(82)
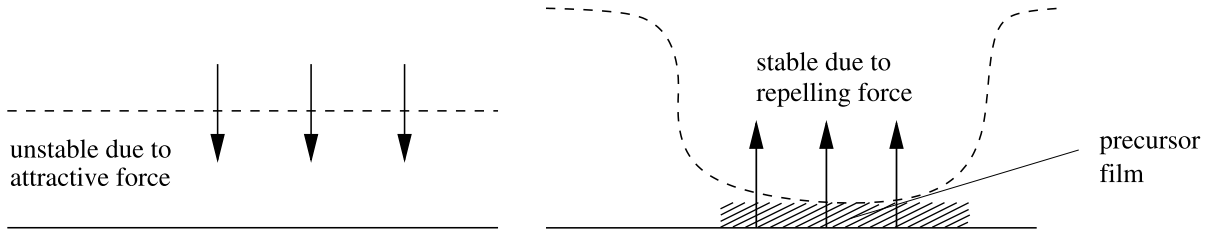
Fluid Dynamics, Pattern Formation, Figure 22
Left: Thin flat films are unstable due to an attractive, long range van der Waals force between the free surface and the solid substrate of the fluid. *Right*: If the film is extremely thin (some nm), a repelling short range force acts as a stabilizer and the precursor film remains intact instead of rupturing
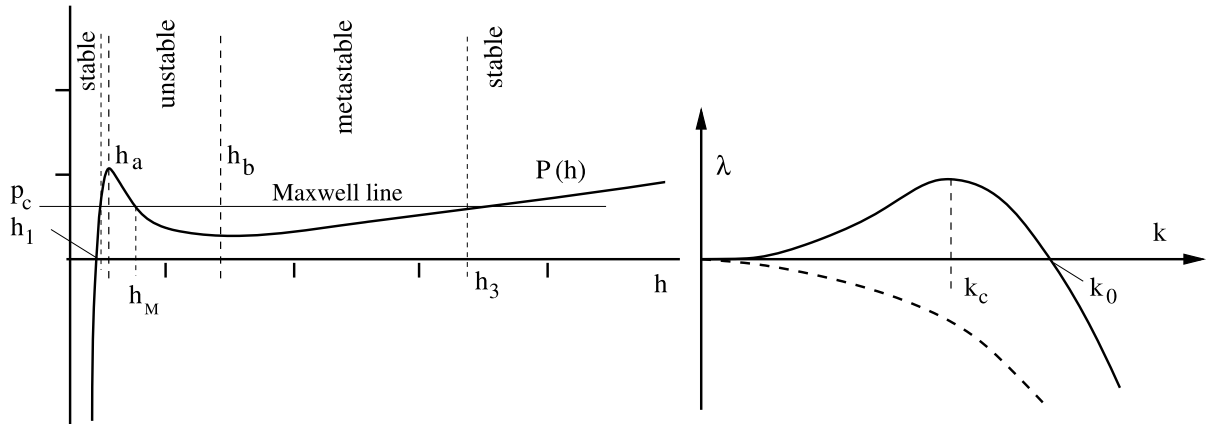
Fluid Dynamics, Pattern Formation, Figure 23
Left: The disjoining pressure for a film with uniform thickness h including gravitation, $A_3 = 3, A_9 = 1, G = 0.1$. The region of unstable films is bounded by h_a and h_b. The critical pressure (depth) P_c (h_M) where drops turn into holes is determined by a Maxwell construction. *Right*: Growth rates of periodic disturbances of the plane surface with wave number k. The solid line corresponds to a film with a mean thickness in the unstable regime. Waves having a wave number $0 < k < k_0$ grow exponentially, the mode with $k = k_c$ has the largest growth rate (most dangerous mode). The instability is of type II

with the "diffusion coefficient"

$$D(h) = d_h P .$$

(Here we restrict our further study to fluids with a uniform surface tension. For non-isothermal films with $\nabla_2 \Gamma \neq 0$ we refer to [16,69]). Next we wish to present numerical solutions of the fully nonlinear Eq. (80) with (81). To this end we used the parameters of Fig. 23 and several initial depths h_0. As initial condition a random distribution around the average depth h_0 was chosen.

In the early stage of the evolution (top line of Fig. 24), structures having a length scale of the critical wave length $\Lambda = 2\pi/k_c$, occur, were k_c is the wave number of the fastest growing mode

$$k_c = \sqrt{-\frac{D}{2}} .$$

This can be called "linear phase" since the amplitudes are still small and nonlinearities play no important role. The structure grows on the typical time scale

$$\tau = \lambda^{-1}(k_c) = \frac{12}{h_0^3 D^2} = \frac{12}{h_0^3} \left(P'(h_0) \right)^{-2} ,$$

which is inverse to the square of the slope of the disjoining pressure. This is the reason why pattern formation in thicker films takes much longer (right column in Fig. 24). As a consequence, the small scale (linear phase) structures are overlayed by holes created by certain seeds. After the linear phase, the position of h_0 with respect to the Maxwell point h_M (Fig. 23, left frame) is decisive. If $h_0 > h_M$, holes are formed, for $h_0 < h_M$, one finds drops. If $h_0 \approx h_M$, maze-like patterns are obtained in form of bent, rather irregular stripes (Fig. 24, middle column). In a last, strongly nonlinear phase, coarsening is observed. The final stationary structure (long term) is often a single entity in the form
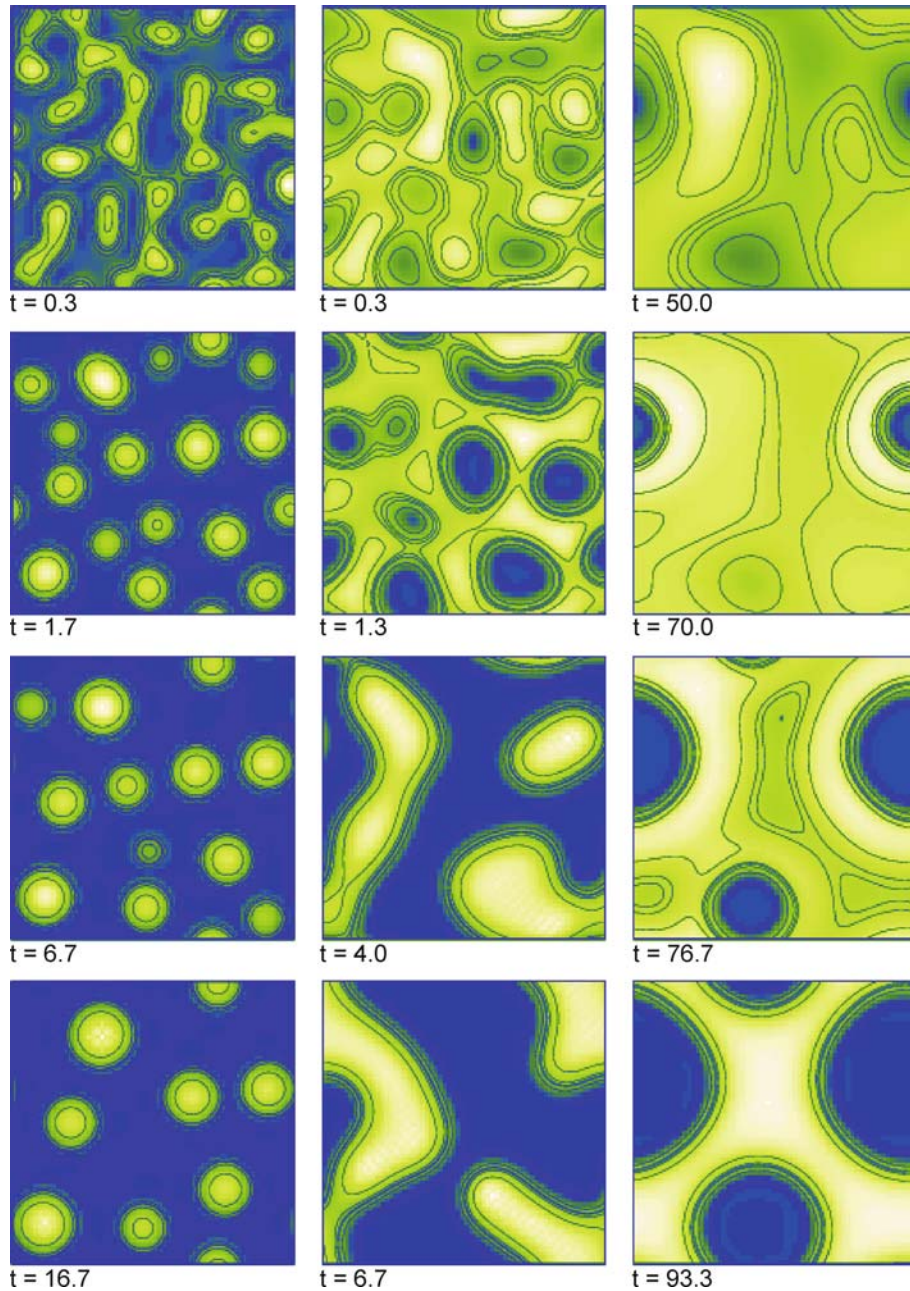
Fluid Dynamics, Pattern Formation, Figure 24
Time series found by numerical integration of (80) for $h_0 = 1.2$ *(left column)*, 1.862 *(middle)*, and 2.8 *(right)*. Light areas correspond to elevated regions of the surface (from [8])

of one big drop or hole. The whole spatio-temporal evolution is transient and can be formulated as a gradient dynamics. The potential plays the role of a generalized free energy reaching its minimum in the steady end state [16].

The flat film is unstable with respect to infinitesimal disturbances if h_0 is in the region between h_a and h_b. On the other hand, two meta-stable domains exist, where the flat film is stable, although the free energy could be lowered by pattern formation. Then, a finite disturbance is necessary, which can be caused by seeds coming, for instance, from impurities. Such a process is called *nucleation* and can be seen in the right column of Fig. 24. There, the seeds

were provided by the random dot initial conditions and two holes are formed. Both processes (nucleation and wetting) converge in this region and it is a question of time scales which one emerges first. In experiments, the formation of holes by nucleation is seen quite often. The reason is that for a Lennard–Jones like disjoining pressure as (81), the meta-stable hole region is much larger compared to that of drops (Fig. 23, left frame).

Phase Field Models

In solidification processes, phase fields are introduced as additional variables to describe the state, here liquid or solid, of the system [57]. Phase fields depend on space and time and governing equations for the phase field variables must be stated or derived. If the phase field obeys an equation of the form of (70), it is called Model A, according to a classification given by Hohenberg and Halperin [46].

Model B Here, we are more interested in phase field equations belonging to Model B. The phase field (we call it Φ) is conserved and a continuity equation

$$\partial_t \Phi = -\operatorname{div}\vec{j} \qquad (83)$$

must hold. As in nonequilibrium thermodynamics [30] one assumes that the current density $\vec{j}$ is proportional to a generalized force $\vec{f}$

$$\vec{j} = Q(\Phi) \cdot \vec{f} \qquad (84)$$

where Q stands for a non-negative mobility, which is normally a function of the phase field itself, but may also explicitly depend on space coordinates. If the force can be derived from a potential P (pressure)

$$\vec{f} = -\nabla P(\Phi) \qquad (85)$$

which in turn can be written as functional derivative of another (thermodynamic) potential (free energy) F

$$P = \frac{\delta F}{\delta \Phi}, \qquad (86)$$

we finally obtain a closed equation for (83) of the form

$$\partial_t \Phi = \operatorname{div}\left[Q(\Phi)\nabla \frac{\delta F}{\delta \Phi} \right]. \qquad (87)$$

With (87) it is easy to show that $\mathrm{d}_t F \leq 0$.

The Cahn–Hilliard Equation As known from the Ginzburg–Landau equation, one may expand the free energy with respect to powers of the phase field. The surface

term $(\nabla \Phi)^2$ penalizes phase field variations with respect to space by an increase of F:

$$F[\Phi] = \int_V d^3\vec{r}\Big[\frac{D}{2}(\nabla\Phi)^2 + a_0\Phi + \frac{a_1}{2}\Phi^2$$
$$+ \frac{a_2}{3}\Phi^3 + \frac{a_3}{4}\Phi^4 + \dots \Big]. \qquad (88)$$

Substituting this into (87) yields

$$\partial_t \Phi = \operatorname{div}\big[Q(\Phi)\nabla\big(-D\Delta\Phi + a_0 + a_1\Phi$$
$$+ a_2\Phi^2 + a_3\Phi^3 \big)\big]. \qquad (89)$$

We further assume $a_2 = 0$ (this can be always obtained by a simple shift of Φ) and $a_1 < 0,\ a_3 > 0$. If we restrict us to the case of a constant mobility, we arrive from (89) after a suitable scaling at the Cahn–Hilliard Eq. [29]

$$\partial_t \Phi = -\Delta\Phi - \Delta^2\Phi + \Delta(\Phi^3). \qquad (90)$$

Equation (90) can be considered as a simple model for a conserved order parameter. A family of stationary solutions of (90) is given by $\Phi = \Phi_0 = $ constant. A linear stability analysis shows that these solutions are type II unstable if $\Phi_0^2 < \frac{1}{3}$ holds. Since (90) belongs to Model B, an infinitesimal disturbance can grow only in a way that keeps the mean value of $\Phi = \Phi_0$ constant. Therefore, spatially structured solutions are expected (Fig. 25).

The density of the free energy of a homogeneous solution reads

$$f(\Phi) = -\frac{\Phi^2}{2} + \frac{\Phi^4}{4} \qquad (91)$$

and has its minima at $\Phi_m = \pm 1$. From Fig. 25 it becomes clear that the stationary pattern forming solutions are located between these two minima independently from the mean Φ_0. If the mean value is increased, the regions with $\Phi \approx 1$ grow at the cost of the regions with $\Phi \approx -1$ and vice versa. Taking (90) as a simple model for the phase transition from liquid to gas, the phase field defines the state of aggregation. The density can then be found from the linear relation

$$\rho(\vec{r}, t) = \frac{1}{2}(\rho_f - \rho_g)\,\Phi(\vec{r}, t) + \frac{1}{2}(\rho_f + \rho_g) \qquad (92)$$

with ρ_g (ρ_f) as the density of the gaseous (liquid) state. Regions where $\Phi \approx -1$ are gaseous, those with $\Phi \approx +1$ liquid.

Equation (90) has no free parameters. On the other hand, the mean $\Phi_0 = \langle\Phi\rangle$ is a conserved quantity which influences the dynamics of pattern formation qualitatively
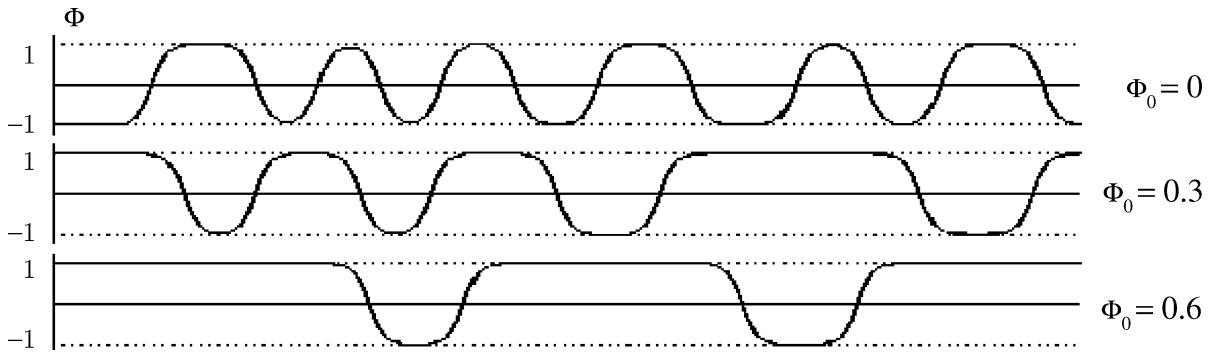
Fluid Dynamics, Pattern Formation, Figure 25
Stationary solutions of the 1D Cahn–Hilliard equation for several mean values Φ_0

and which can be considered as a control parameter. Integrating (92) over the volume, it turns out that Φ_0 is linked to the total mass M of the gas/liquid system

$$M = \frac{1}{2}(\rho_f - \rho_g)\,\Phi_0 \cdot V + \frac{1}{2}(\rho_f + \rho_g) \cdot V \,.$$

The stable homogeneous solutions $\Phi_0^2 > 1/3$ correspond to a pure gas phase ($\Phi_0 < 0$, small total mass), or to a pure liquid phase ($\Phi_0 > 0$, large total mass). In the unstable regime $\Phi_0^2 < 1/3$ the (homogeneous) system has a medium density; this corresponds either to an oversaturated vapor atmosphere ($\Phi_0 < 0$) or to a liquid with a temperature above its boiling point. In both cases, an infinitesimally small disturbance is sufficient to trigger pattern formation in the form of phase separation. In the first case, one observes drops in the gas atmosphere, in the latter, bubbles in the liquid. Figure 26 shows a numerical simulation of (90) in three dimensions.
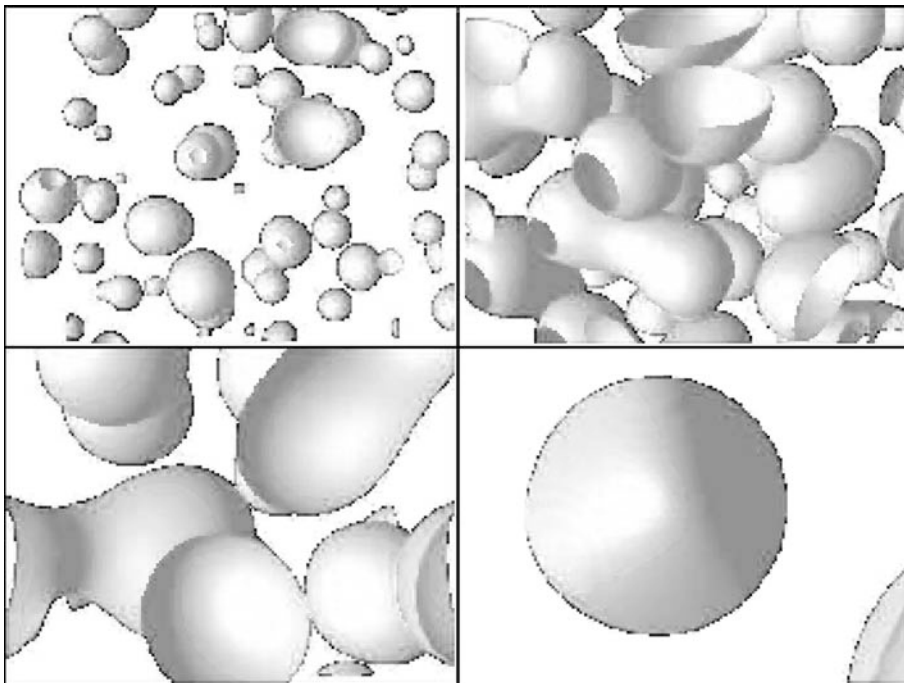
Fluid Dynamics, Pattern Formation, Figure 26
Numerical solution of the Cahn–Hilliard Eq (90) in three room dimensions. The time series *(top left to bottom right)* shows how liquid drops are formed in an oversaturated gas atmosphere. Finally they merge to one big drop by coarsening, a typical dynamic for a type II instability (from [7])

The Fluid Density as Phase Field

Writing down an equation such as (87) and an expansion such as (88) seems to be rather ad hoc. However, for pure fluids it is evident to use the density itself as the phase field, if one is interested in the liquid/gas phase transition. Then, the continuity equation (2) may serve as a phase field equation in lieu of (87). Consequently, the fluid can no longer be considered incompressible.

The Model The Navier–Stokes equations for a compressible fluid (14) must be extended by a force term coming from spatial variations of the phase field (density). They read [21,53]

$$\rho\left[\partial_t \vec{v} + (\vec{v} \cdot \nabla)\,\vec{v}\right] = -\operatorname{grad} p + \vec{f} + \eta\,\Delta\vec{v}$$
$$+ \left(\zeta + \frac{\eta}{3}\right)\operatorname{grad}\operatorname{div}\vec{v} + \mathcal{K}\rho\,\operatorname{grad}\Delta\rho\,. \quad (93)$$

The extra term at the end of (93) was first used by Korteweg in 1901 and is sometimes called Korteweg stress [54]. For (93) we assumed constant material parameters η, ζ, and $\mathcal{K}$. Using the methods of thermodynamics,

the pressure is related to the free energy density f [2,37]

$$p(\rho) = \rho\,\frac{\partial f(\rho)}{\partial\rho} - f(\rho) \quad (94)$$

and the free energy as a functional of ρ reads

$$F[\rho] = \int_V \mathrm{d}^3\vec{r}\left[\frac{\mathcal{K}}{2}(\nabla\rho)^2 + f(\rho)\right]\,, \quad (95)$$

according to (88). Equations (93) with (94) and (2) form a closed system for the variables $\vec{v}$ and ρ. Wetting properties and contact angles at the walls depend on the boundary conditions $\rho = \rho_w$ along the wall [75]. The choice $\rho_w = \rho_f$ corresponds to a completely wetting (hydrophilic) material, $\rho_w = \rho_g$ to a non-wetting (hydrophobic) boundary. The boundary condition for $\vec{v}$ can either be no-slip (along a wall), no flux, or periodic. It is straightforward to include evaporation and condensation effects into the model, which is studied in [20].

Note that now the free energy (95) is not needed for determining the evolution of the phase field by (*ad-hoc*) gradient dynamics. However, it can be shown that the free energy decreases monotonically in time.
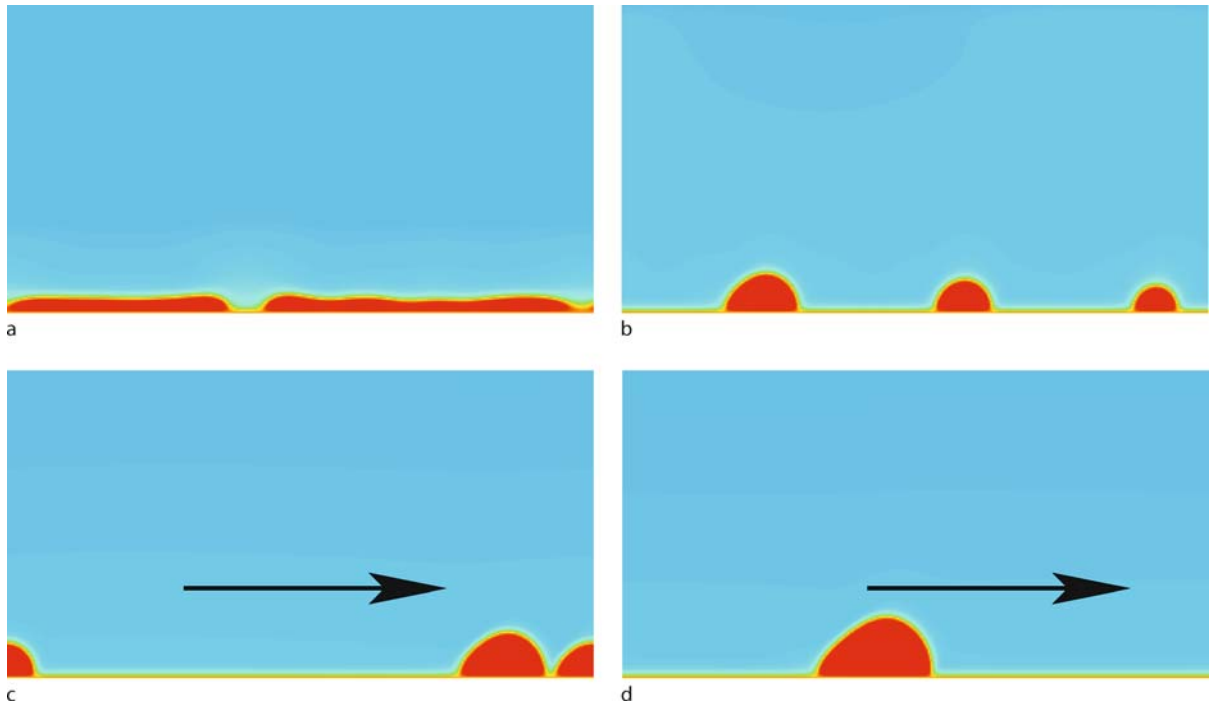
Fluid Dynamics, Pattern Formation, Figure 27
Transition from a flat unstable liquid layer to a drop running down on an inclined substrate (arrows) under gravity effects. Numerical simulation [22] of (93) with (96) and the material parameters for water/vapor from [24] and $\rho_w = 0.5\rho_f$
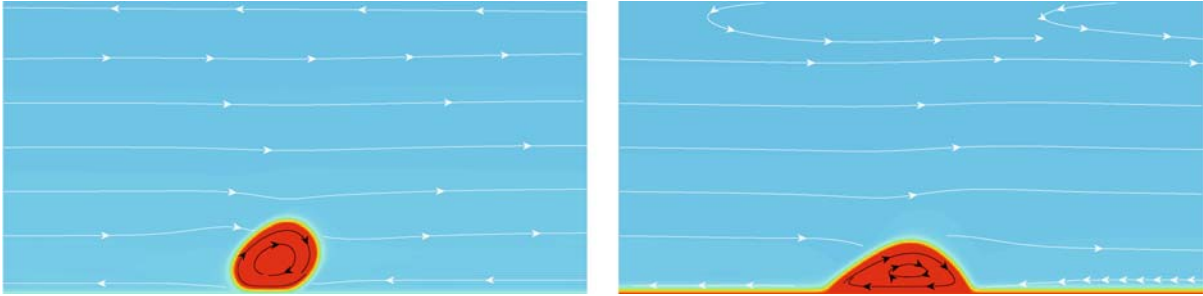
Fluid Dynamics, Pattern Formation, Figure 28
Two final states showing a drop sliding down the inclined substrate with $\rho_w = 0.1\rho_f$ (left, almost hydrophobic) and $\rho_w = 0.8\rho_f$ (*right*, almost hydrophilic). The flow in the gas and in the liquid is indicated by *small arrows* [23]

Results   Again, we wish to consider the formation of one state of aggregation on the background of the other. To account for the two stable states, liquid and gaseous, we take (for sake of simplicity we assume $\rho_g \approx 0$)

$$f(\rho) = \gamma\rho^2(\rho - \rho_f)^2 \, . \tag{96}$$

where γ is a positive material constant. In Fig. 27 we show results of the breakup of a flat liquid film aligned along a rigid bottom plate. The layer is inclined by an angle φ and under vertical gravitation. Thus, an external force density of the form

$$\vec{f} = \rho g \begin{pmatrix} \sin\varphi \\ -\cos\varphi \end{pmatrix}$$

occurs in (93). The bottom material is assumed to be partially wetting ($\rho_w = 0.5\rho_f$) and the initial film is unstable under these conditions. Periodic disturbances grow along the fluid's surface. After rupture, bubbles separate and travel from left to right due to downhill force. Figure 28 shows final states of a sliding drop for two boundary values ρ_w. Clearly, the contact angles are different [22].

The phase field description goes far beyond the one based on the thin film equation of Sect. "Thin Films", since there the treatment was restricted to small contact angles and rupture was excluded from the beginning.

Future Directions

There is a huge number of applications in science, industry, and technology where the methods and models outlined in the present article can be used and developed further. In the field of patterns not formed by self-organized processes, but rather by external events such as tidal waves, storm surges or Tsunamis, a reduced and simplified description as discussed in Sects. "Surface Waves", "Order Parameter Equations" should allow for a better under-

standing of the underlying mechanisms and their effects. Highly involved problems, as for instance the flow, temperature, and concentration fields inside a combustion cell, could be tackled by such models, extended in a suitable way.

Self-organized fluid patterns (Sect. "Instabilities") are the focus of attention in many actual fields of quite different disciplines and scales. The conditions that lead to the creation and stabilization of hurricanes are not yet completely known. The rather high probability of the occurrence of freak waves in the open sea still waits for an explanation. On a planetary scale, convection problems are encountered in the interior of planets and stars and may give rise to the spontaneous formation of a magnetic field. Another problem of great interest for the geophysicist is that of a fluid (such as oil) in a porous medium. The equations for that case differ only a little from that discussed in Sect. "The Basic Equations of Fluid Dynamics" and could therefore be treated in the same spirit.

Understanding the mechanisms responsible for pattern formation can also help to control systems to avoid the occurrence of spatial patterns. In this way, the quality of products obtained from industrial processes, such as coating or solidification (crystal growth), might be improved.

On the micro-scale, fluid problems in general ruled by the Stokes equations discussed in Sect. "Conserved Order Parameter Fields" form a major issue, founding the new discipline of micro-fluidics. But even on the nanoscale, there are new applications in view. The self-organized growth of structures could be a promising tool in the conception and construction of nano-circuits.

An extension of the treatment to complex fluids such as mixtures and emulsions, or to non-Newtonian fluids using the phase field approach (Sect. "Conserved Order Parameter Fields"), is desirable. These fluids are important for biological applications.

Bibliography

Primary Literature

1. Abramowitz M, Stegun IA (1965) Handbook of mathematical functions. Dover, New York
2. Anderson DM, Mc Fadden GB (1997) A diffuse-interface description of internal waves in a near-critical fluid. Phys Fluids 9:1870–1879
3. Aranson IS, Kramer L (2002) The world of the complex Ginzburg–Landau equation. Rev Mod Phys 74:99–143
4. Argyris JH, Faust G, Haase M (1994) An exploration of chaos: An introduction for natural scientists and engineers. North-Holland, Amsterdam
5. Bestehorn M (1993) Phase and amplitude instabilities for Bénard–Marangoni convection in fluid layers with large aspect ratio. Phys Rev 48:3622–3634
6. Bestehorn M (1996) Square patterns in Bénard–Marangoni convection. Phys Rev Lett 76:46–49
7. Bestehorn M (2006) Hydrodynamik und Strukturbildung. Springer, Berlin
8. Bestehorn M (2007) Convection in thick and thin fluid layers with a free interface. Eur Phys J Spec Top 146:391–405
9. Bestehorn M, Colinet P (2000) Bénard–Marangoni–convection of a binary mixture as an example of an oscillatory bifurcation under strong symmetry-breaking effects. Phys D 145:84
10. Bestehorn M, Friedrich R (1999) Rotationally invariant order parameter equations for natural patterns in nonequilibrium systems. Phys Rev E 59:2642–2652
11. Bestehorn M, Haken H (1983) A calculation of transient solutions describing roll and hexagon formation in the convection instability. Phys Lett A 99:265–268
12. Bestehorn M, Haken H (1990) Traveling waves and pulses in a two-dimensional large-aspect-ratio system. Phys Rev A 42:7195–7203
13. Bestehorn M, Neuffer K (2001) Surface patterns of laterally extended thin liquid films in three dimensions. Phys Rev Lett 87:046101
14. Bestehorn M, Pérez-García C (1992) Study of a model of thermal convection in cylindrical containers. Phys D 61:67–76
15. Bestehorn M, Neufeld M, Friedrich R, Haken H (1994) Comment on spiral-pattern formation in Rayleigh–Bénard convection. Phys Rev E 50:625
16. Bestehorn M, Pototsky A, Thiele U (2003) 3D Large scale Marangoni convection in liquid films. Eur Phys J B 33:457
17. Bénard H (1901) Les tourbillons cellulaires dans une nappe liquide. Ann Chim Phys 23:62–143
18. Block MJ (1956) Surface tension as the cause of Bénard cells. Nat Lond 176:650–651
19. Bodenschatz E, Pesch W, Ahlers G (2000) Recent developments in Rayleigh–Bénard convection. Ann Rev Fluid Mech 32:709–778
20. Borcia R, Bestehorn M (2005) Phase-filed simulations for evaporation with convection in liquid-vapor systems. Eur Phys J B 44:101–108
21. Borcia R, Bestehorn M (2007) Phase-field simulations for drops and bubbles. Phys Rev E 75:056309
22. Borcia R, Borcia I, Bestehorn M (2008) Phys Rev E (submitted)
23. Borcia R, Borcia I, Bestehorn M (2008) Static and dynamic contact angles. Eur Phys J Spec Top (in print)
24. Burelbach JP, Bankoff SG, Davis SH (1988) Nonlinear stability of evaporating/condensing liquid films. J Fluid Mech 195:463
25. Busse FH (1967) The stability of finite amplitude cellular convection. J Fluid Mech 30:625–649
26. Busse FH (1989) In: Peltier WR (ed) Fundamentals of thermal convection. Taylor, pp 23–95
27. Busse FH, Frick H (1985) Square-pattern convection in fluids with strongly temperature-dependent viscosity. J Fluid Mech 150:451–465
28. Busse FH, Riahi N (1980) Nonlinear convection in a layer with nearly insulating boundaries. J Fluid Mech 96:243–256
29. Cahn JW, Hilliard JE (1958) Free energy of a nonuniform system. J Chem Phys 28:258–267
30. Callen HB (1985) Thermodynamics and an introduction to thermostatistics. Wiley, New York
31. Castets V, Dulos E, Boissonade J, De Kepper P (1990) Experimental evidence of a sustained standing Turing-type nonequilibrium chemical pattern. Phys Rev Lett 64:2953–2956
32. Chandrasekhar S (1961) Hydrodynamic and hydromagnetic stability. Dover, New York
33. Cohen IM, Kundu PK (2004) Fluid Mechanics. Academic, Amsterdam
34. Colinet P, Legros JC, Velarde MG (2001) Nonlinear dynamics of surface tension driven instabilities. Wiley, Berlin
35. Cross MC (1988) Theoretical methods in pattern formation in physics, chemistry and biology. In: Garrido L (ed) Far from equilibrium phase transitions. Springer, Berlin
36. Cross MC, Hohenberg PC (1993) Pattern formation outside of equilibrium. Rev Mod Phys 65:851–1112
37. Davis HT, Scriven LE (1982) Stress and structures in fluid interfaces. Adv Chem Phys 49:357
38. de Gennes PG (1985) Wetting: statics and dynamics. Rev Mod Phys 57:827–863
39. Dean RG, Dalrymple RA (2000) Water wave mechanics for engineers and scientists. World Science, Singapore
40. Drazin PG, Johnson RS, Crighton DG (1989) Solitons: An Introduction. Cambridge University Press, Cambridge
41. Eckert K, Bestehorn M, Thess A (1998) Square cells in surface tension driven Bénard convection: experiment and theory. J Fluid Mech 356:155–197
42. Faraday M (1831) On a peculiar class of acoustical figures and on certain forms assumed by groups of particles upon vibrating elastic surfaces. Philos Trans R Soc Lond 121:299
43. Getling AV (1998) Rayleigh–benard convection: structures and dynamics. World Scientific
44. Golovin AA, Nepomnyashchy AA, Pismen LM (1997) Nonlinear evolution and secondary instabilities of Marangoni convection. J Fluid Mech 341:317–341
45. Guckenheimer J, Holmes P (2002) Nonlinear oscillations, dynamical systems, and bifurcations of vector fields. Springer, Berlin
46. Hohenberg PC, Halperin BI (1977) Theory of dynamic critical phenomena. Rev Mod Phys 49:435–47
47. See the up to now 69 volumes of the Springer Series of Synergetics, especially the monographs by Haken H (1983) Synergetics. An Introduction.; Advanced Synergetics.; Information and Self-Organization (2006).
48. Haken H (1975) Cooperative phenomena in systems far from thermal equilibrium and in nonphysical systems. Rev Mod Phys 47:67–121
49. Haken H (1983) Advanced Synergetics: Instability hierarchies of self-organizing systems and devices. Springer, Berlin

50. Haken H (2004) Synergetics introduction and advanced topics. Springer, Berlin

51. Hardy W (1919) The spreading of fluids on glass. Philos Mag 38:49–55

52. Israelachvili JN (1992) Intermolecular and surface forces. Academic Press, London

53. Jasnow D, Vinals J (1996) Coarse-grained description of thermo-capillary flow. Phys Fluids 8:660–669

54. Korteweg DJ (1901) Sur la forme que prennent les équation du mouvements des fluides. Arch Néerl Sci Exact Nat Ser II 6:1

55. Koschmieder EL (1993) Bénard cells and Taylor vortices. Cambridge University Press, Cambridge

56. Lai WM, Rubin D, Krempl E (1993) Introduction to continuum mechanics. Pergamon, Oxford

57. Langer JS (1980) Instabilities and pattern formation in crystal growth. Rev Mod Phys 52:1–28

58. Landau LD, Lifshitz EM (1996) Statistical physics. In: Course of theoretical physics, vol 5. Heinemann, Butterworth

59. Landau LD, Lifshitz EM (2004) Fluid dynamics. In: Course of theoretical physics, vol 6. Heinemann, Butterworth

60. Lorenz EN (1963) Deterministic nonperiodic flow. J Atmos Sci 20:130–141

61. Morris SW, Bodenschatz E, Cannell DS, Ahlers G (1993) Spiral defect chaos in large aspect ratio Rayleigh–Bénard convection. Phys Rev Lett 71:2026–2029

62. Moses E, Steinberg V (1986) Competing patterns in a convective binary mixture. Phys Rev Lett 57:2018–2021

63. Murray JD (1993) Mathematical biology. Springer, Berlin

64. Nekorkin VI, Velarde MG (2002) Synergetic phenomena in active lattices. Patterns, waves, solitons, chaos. Springer, Berlin

65. Nepomnyashchy AA, Velarde MG, Colinet P (2002) Interfacial phenomena and convection. Chapman & Hall, Boca Raton

66. Newell AC, Whitehead JA (1969) Finite bandwidth, finite amplitude convection. J Fluid Mech 38:279–303

67. Nitschke-Eckert K, Thess A (1995) Secondary instability in surface tension driven Bénard convection. Phys Rev E 52:5772–5775

68. Ockendon H, Ockendon JR (1995) Viscous flow. Cambridge University Press, Cambridge

69. Oron A (2000) Nonlinear dynamics of three-dimensional long-wave Marangoni instability in thin liquid films. Phys Fluids 12:1633–1645

70. Oron A, Davis SH, Bankoff SG (1997) Long-scale evolution of thin liquid films. Rev Mod Phys 69:931–980

71. Ouyang Q, Swinney HL (1991) Transition from a uniform state to hexagonal and striped Turing patterns. Nature 352:610–612

72. Palm E (1960) On the tendency towards hexagonal cells in steady convection. J Fluid Mech 19:183–192

73. Pesch W (1996) Complex spatiotemporal convection patterns. Chaos 6:348–357

74. Penrose R (1974) Role of aesthetics in pure and applied research. Bull Inst Math Appl 10:266

75. Pismen LM, Pomeau Y (2000) Disjoining potential and spreading of thin liquid layers in the Diffuse-interface model coupled to hydrodynamics. Phys Rev E 62:2480–2492

76. Prigogine I, Levever R (1968) Symmetry breaking instabilities in dissipative systems II. J Chem Phys 48:1695–1700

77. Prigogine I, Nicolis G (1967) On symmetry-breaking instabilities in dissipative systems. J Chem Phys 46:3542–3550

78. Rayleigh Lord (1915) On convection currents in a horizontal layer of fluid. Phil Mag 32:462–468

79. Reiter G, Sharma A, Casoli A, David M-O, Khanna R, Auroy P (1999) Thin film instability induced by long-range forces. Langmuir 15:2551–2558

80. Schatz MF, Neitzel GP (2001) Experiments on thermocapillary instabilities. Ann Rev Fluid Mech 33:93–127

81. Schatz MF, Van Hook SJ, Mc Cormick WD, Swift JB, Swinney HL (1999) Time-independent square patterns in surface-tension-driven Bénard convection. Phys Fluids 11:2577–2582

82. Scheid B, Oron A, Colinet P, Thiele U, Legros JC (2002) Nonlinear evolution of nonuniformly heated falling liquid films. Phys Fluids 14:4130–4151

83. Schwabe D (2006) Marangoni instabilities in small circular containers under microgravity. Exp Fluids 40:942–950

84. Sharma A, Khanna R (1998) Pattern formation in unstable thin liquid films. Phys Rev Lett 81:3463–3466

85. Sharma A, Khanna R (1999) Pattern formation in unstable thin liquid films under the influence of antagonistic short and long-range forces. J Chem Phys 110:4929–4936

86. Seemann R, Herminghaus S, Jacobs K (2001) Dewetting patterns and molecular forces: A reconciliation. Phys Rev Lett 86:5534–553

87. Sparrow C (1982) The Lorenz equations. Springer, Berlin

88. Swift JB, Hohenberg PC (1977) Hydrodynamic fluctuations at the convective instability. Phys Rev A 15:319; Chem Phys 48:1695–1700

89. Turing AM (1952) The chemical basis of morphogenesis. Phil Trans R Soc Lond B237:37

90. Van Dyke M (1982) An album of fluid Motion. Parabolic, Stanford

91. Van Oss CJ, Chaudhury MK, Good RJ (1988) Interfacial Lifshitz–van der Waals and polar interactions in macroscopic systems. Chem Rev 88:927–941

92. Vrij A (1966) Possible mechanism for the spontaneous rupture of thin, free liquid films. Disc Faraday Soc 42:23–33

Books and Reviews

Emmerich H (2003) The diffusive interface approach in material science. Springer, Berlin

Fletcher CAJ (1988) Computational techniques for fluid dynamics, vol 1,2. Springer, Berlin

Manneville P (1990) Dissipative structures and weak turbulence. Academic Press, London

Platten JK, Legros JC (1984) Convection in liquids. Springer, Berlin

Pismen LM (2006) Patterns and interfaces in dissipative dynamics. Springer, Berlin

Simanovskii IB, Nepomnyashchy AA (1993) Convective instabilities in systems with interface. Gordon Preach

Schlichting H, Gersten K (2000) Boundary-layer theory. Springer, Berlin

Fluid Dynamics, Turbulence

RUDOLF FRIEDRICH[1], JOACHIM PEINKE[2]

[1] Institute for Theoretical Physics, University of Münster, Münster, Germany

[2] Institute of Physics, Carl-von-Ossietzky University Oldenburg, Oldenburg, Germany

Article Outline

Glossary

Basic equation of fluid dynamics Fluid motion is mathematically treated on the basis of a continuum theory. The fundamental evolution equations are the Euler equation for ideal fluids and the Navier–Stokes equation for Newtonian fluids.

Vortex motions Numerical calculations of turbulent flow fields show that the flows are dominated by coherent structures in form of vortex sheets or tube-like vortices. The question, why vorticity tends to be condensed in localized objects, is one of the central issues of fluid dynamics. Regarding two-dimensional flows there are attempts to approximate fluid motion by a collection of point vortices. This allows one to investigate properties of flows on the basis of a finite dimensional (Hamiltonian) dynamical system.

Turbulence modeling and large eddy simulations The evolution equation for the average velocity field of turbulent flows contains the Reynolds stresses, whose origin are the turbulent pulsations. Turbulence modeling consists of relating the Reynolds stresses to averaged quantities of the fluid motion. This allows one to perform numerical computations of large-scale flows without resolving the turbulent fine structure.

Phenomenological theories of the fine structure of turbulence Phenomenological theories play an important role in physics, and are quite often formulated before a microscopic understanding of the physical problem has been achieved. Phenomenological theories have been developed for the fine structure of turbulence. Of great importance is the theory of Kolmogorov, which he formulated in the year 1941 and refined in 1962.

The so-called K41 and K62 theories focus on the self-similar behavior of statistical properties of velocity increments, i. e. the velocity difference between two points with a spatial distance r. Recently, phenomenological theories have been developed that consider the joint probabilities of velocity increments at different scales. It is expected that multiple scale analysis of turbulence will provide new insights into the spatio-temporal complexity of turbulence.

Turbulent cascades Fluid motions are dissipative systems. Stationary flows can only be maintained by a constant energy input in the form of shear flows or body forces. Usually, the length and time scales related to the energy input are widely separated from the ones on which energy is dissipated. A consequence is the establishment of an energy transport across scales. It is believed that this energy transport is local in scale leading to the so-called energy cascades. These cascades are related to the emergence of scaling behavior. There is a direct energy cascade in three dimensions from large to small scales and an inverse cascade of energy from small scales to large scales in two-dimensional flows.

Analytical theories of turbulence Analytical theories of turbulence try to assess the experimental results on turbulent flows directly from a statistical treatment of the basic fluid dynamical equations. Analytical theories rely on renormalized perturbation expansions and use methods from quantum field theory and renormalization group methods. However, no generally accepted theory has emerged so far.

Definition of the Subject

Fluid flows are open systems far from equilibrium. Fluid motion is sustained by energy injected at a certain scale, the so-called integral scale and is dissipated by viscosity mainly in small-scale structures. If the integral scale and the dissipative scale are widely separated and the motions on the integral scale are sufficiently strong, the fluid develops a range of spatio-temporal structures. In three-dimensional flows these structures steadily decay into smaller structures and are generated by the instability of larger structures. This leads to a cascading process which transports energy across scales. Turbulence appears if the fluid motion is driven far away from equilibrium. It develops through sequences of instabilities and processes of selforganization. From this respect, turbulence is a highly ordered phenomenon, whose spatio-temporal complexity, however, has still to be explored.

Introduction

Turbulence is one of the outstanding problems in the field of nonlinear dynamics and complex systems. Although the basic equations of ideal fluid dynamics were formulated by L. Euler 250 years ago and the equations for viscous flows, the so-called Navier–Stokes equation, were established about 150 years ago [13], only a few analytical solutions have been found so far, because of the inherent nonlinear character of fluid flows. Furthermore, fluid motions are systems far from equilibrium. Their maintenance requires a constant input of energy, which is transformed by the viscous flow into heat. A measure for the distance from equilibrium, which corresponds to vanishing fluid velocity, is the Reynolds number,

$$\text{Re} = \frac{UL}{\nu} \tag{1}$$

where U is a characteristic velocity, L a characteristic length scale, and ν is the kinematic viscosity. Flows with Reynolds numbers larger than $\text{Re} = 1000$ usually are turbulent. A turbulent field generated in a free jet experiment is exhibited in Fig. 1. By increasing the Reynolds number one observes the occurrence of various types of instabilities resulting in time-dependent and chaotic patterns making these systems paradigms of self-organization. Whereas the flows generated by the first few instabilities can be treated satisfactorily the transitions and properties of flows at higher Reynolds numbers are by far less understood. This lack of understanding hinders the scientific development in various fields, ranging from astrophysics, engineering to the life sciences.

Basic research on turbulence has always stimulated and contributed to the formulation of new scientific concepts like self-organization and pattern formation, chaos, and the theory of fractals. As a classical nonlinear field theory the description of fluid motion has advanced the mathematical understanding of infinite dimensional nonlinear dynamical systems and the development of efficient computational tools. It is expected that combined experimen-

tal and theoretical efforts will lead to a satisfactory understanding of high Reynolds number flows in the near future.

The Basic Hydrodynamic Equations

Fluid motions are described in terms of a continuum theory. The basic ingredients of continuum theories are balance equations for a density $h(\mathbf{x}, t)$ of a physical quantity like mass or momentum defined at location $\mathbf{x}$ and time t:

$$\frac{\partial}{\partial t} h + \nabla \cdot [\mathbf{u} h + \mathbf{j}_h] = q. \tag{2}$$

Here, $\mathbf{u}$ denotes the fluid velocity, $\mathbf{j}_h$ the current corresponding to the density h and q denotes a source term [11,47].

The balance equation for the mass density ρ reads

$$\frac{\partial}{\partial t} \rho + \nabla \cdot \rho \mathbf{u} = 0. \tag{3}$$

Since mass is conserved, the source term vanishes identically, $q = 0$.

Incompressible fluid motions are characterized by the condition

$$\nabla \cdot \mathbf{u} = 0. \tag{4}$$

In the present review we shall mainly focus on incompressible fluids.

The balance equation for the density of momentum, $\rho \mathbf{u}(\mathbf{x}, t)$, takes the form

$$\frac{\partial}{\partial t} \rho u_i + \sum_j \frac{\partial}{\partial x_j} u_j \rho u_i = -\frac{\partial}{\partial x_i} p + \sum_j \frac{\partial}{\partial x_j} \sigma_{ij} + f_i, \tag{5}$$

where the momentum current $\mathbf{j}_h$ is expressed by pressure p and the viscous stress tensor σ_{ij}. External forces are summarized in f_i. A complete description requires the formulation of boundary conditions for the velocity field.

It is straightforward to derive the balance equation for the density of the kinetic energy, $\rho \mathbf{u}^2(\mathbf{x}, t)/2$, from the con-

Fluid Dynamics, Turbulence, Figure 1
Development of turbulent structures in a free jet experiment

servation law of momentum (5) for incompressible flows:

$$\frac{\partial}{\partial t} \frac{\rho}{2} \mathbf{u}^2 + \sum_j \frac{\partial}{\partial x_j} \left\{ u_j \left[\frac{\rho}{2} \mathbf{u}^2 + p \right] - \sum_i \sigma_{ji} u_i \right\}$$
$$= -\sum_{ij} \sigma_{ij} \frac{\partial u_i}{\partial x_j} + \sum_j u_j f_j . \quad (6)$$

This equation shows that energy is conserved provided the viscous stresses σ_{ij} vanish.

Ideal Fluids: Euler's Equation

For ideal fluid motions, the kinetic energy is conserved provided external forces are absent. The balance Eq. (6) shows that in this case the viscous stresses σ_{ij} have to vanish leading to the Euler equation for incompressible fluid motions:

$$\left[\frac{\partial}{\partial t} + \mathbf{u} \cdot \nabla \right] \mathbf{u} = -\frac{1}{\rho} \nabla p + \frac{1}{\rho} \mathbf{f} . \quad (7)$$

The dynamics of ideal fluid motion is restricted by Kelvin's theorem. The circulation $\oint \mathbf{u}(\mathbf{x}, t) \, d\mathbf{r}$ along closed curves going with the flow remains constant [11,47].

Newtonian Fluids: Navier–Stokes Equation

Newtonian fluids are characterized by the presence of viscous stresses. They are assumed to be proportional to the strain matrix S_{ij}

$$\sigma_{ij} = \nu \rho S_{ij} = \nu \rho \frac{1}{2} \left[\frac{\partial u_i}{\partial x_j} + \frac{\partial u_j}{\partial x_i} \right] . \quad (8)$$

Assuming isotropic material properties of the fluid as well as incompressibility one obtains the Navier–Stokes equation:

$$\left[\frac{\partial}{\partial t} + \mathbf{u} \cdot \nabla \right] \mathbf{u} = \nu \Delta \mathbf{u} - \nabla p + \mathbf{f} . \quad (9)$$

The kinematic viscosity ν characterizes different fluids. The local energy dissipation rate, denoted by ϵ is obtained from the balance equation of the density of kinetic energy, (6):

$$\epsilon = \frac{\nu}{2} \sum_{ij} \left(\frac{\partial u_i}{\partial x_j} + \frac{\partial u_i}{\partial x_j} \right)^2 . \quad (10)$$

This quantity plays a crucial role for the understanding of turbulent fluid motions.

At first glance, the Navier–Stokes equation seems to be underdetermined due to the appearance of the gradient pressure term. However, as a result of incompressibility, the pressure is uniquely defined by the Poisson equation in connection with suitable boundary conditions:

$$\Delta p = -\sum_{ij} \frac{\partial u_i}{\partial x_j} \frac{\partial u_j}{\partial x_i} . \quad (11)$$

The pressure can be determined with the help of Green's function G of the Laplacian,

$$\Delta G(\mathbf{x} - \mathbf{x}') = -\delta(\mathbf{x} - \mathbf{x}') , \quad (12)$$

and yields the pressure as a functional of the velocity field:

$$p = \int d\mathbf{x}' \, G(\mathbf{x} - \mathbf{x}') \sum_{ij} \frac{\partial u_i}{\partial x_j'} \frac{\partial u_j}{\partial x_i'} . \quad (13)$$

This clearly demonstrates that incompressible fluid motions are governed by nonlinear, nonlocal interactions. The gradient pressure term can be regarded as a Lagrange parameter which guaranties the incompressibility of fluid motion.

Vorticity Formulation of Incompressible Fluid Dynamics It is possible to formulate the basic fluid dynamic equations using the vorticity $\boldsymbol{\omega}(\mathbf{x}, t) = \nabla \times \mathbf{u}(\mathbf{x}, t)$. Provided the vorticity is known, one can obtain the velocity field by the analogue of Maxwell's equation of magnetostatics:

$$\nabla \times \mathbf{u} = \boldsymbol{\omega} , \quad \nabla \cdot \mathbf{u} = 0 . \quad (14)$$

The velocity field is determined by the analogy to Biot–Savart's law

$$\mathbf{u}(\mathbf{x}, t) = \int d\mathbf{x}' \, \boldsymbol{\omega}(\mathbf{x}', t) \times \mathbf{K}(\mathbf{x} - \mathbf{x}') + \nabla \Phi , \quad (15)$$

where $\mathbf{K}$ is related to Green's function $G(\mathbf{x})$ of the Laplacian $\mathbf{K}(\mathbf{x}) = \nabla G(\mathbf{x})$. The potential Φ has to fulfill $\Delta \Phi = 0$.

It is straightforward to derive an evolution equation for the vorticity:

$$\frac{\partial}{\partial t} \boldsymbol{\omega} + \mathbf{u} \cdot \nabla \boldsymbol{\omega} = \boldsymbol{\omega} \cdot \nabla \mathbf{u} + \nu \Delta \boldsymbol{\omega} + \mathbf{f}_\omega . \quad (16)$$

Here, an important difference between two- and three-dimensional fluid motion becomes evident. For two-dimensional flows the vorticity only has a component perpendicular to the motion and the so-called vortex stretching term $\boldsymbol{\omega} \cdot \nabla \mathbf{u}$ vanishes identically.

Lagrangian Formulation of Incompressible Fluid Dynamics   Up to now, we have treated fluid dynamics from the Eulerian point of view by considering fields defined at a fixed spatial location. There is an alternative approach to the description of fluid motion. This so-called Lagrangian treatment is based on the introduction of the Lagrangian velocity $\mathbf{U}(\mathbf{y}, t)$ and the Lagrangian path $\mathbf{X}(\mathbf{y}, t)$ of a point moving with the flow starting at time $t = 0$ at the location $\mathbf{y}$. For obvious reasons the quantity $\mathbf{X}(\mathbf{y}, t)$ is also denoted as *Lagrangian map*. The inverse map is denoted as $\mathbf{y}(\mathbf{x}, t)$. The basic fluid dynamical equations can also be formulated in this Lagrangian picture.

As an example we formulate the evolution equation for the Lagrangian vorticity for two-dimensional incompressible flows. To this end we introduce the Lagrangian vorticity $\Omega(\mathbf{y}, t) = \boldsymbol{\omega}(\mathbf{X}(\mathbf{y}, t), t)$. The first equation defines the Lagrangian path, the second the evolution of the Lagrangian vorticity.

$$\frac{\partial}{\partial t}\mathbf{X}(\mathbf{y}, t) = \int d\mathbf{y}' \, \Omega(\mathbf{y}', t)\{\mathbf{e}_z \times \mathbf{K}[\mathbf{X}(\mathbf{y}, t) - \mathbf{X}(\mathbf{y}', t)]\},$$

$$\frac{\partial}{\partial t}\Omega(\mathbf{y}, t)$$
$$= \nu\left[\sum_{ijkl}\frac{\partial Y_l}{\partial x_i}\frac{\partial Y_k}{\partial x_i}\frac{\partial^2}{\partial y_l \partial y_k} + \sum_{li}\frac{\partial^2 Y_l}{\partial x_i \partial x_i}\frac{\partial}{\partial y_l}\right]\Omega(\mathbf{y}, t).$$
(17)

For two-dimensional flows the gradient of Green's function of the Laplacian takes the form $\mathbf{K}(\mathbf{x}) = \frac{\mathbf{x}}{2\pi|\mathbf{x}|^2}$. It is immediately obvious that in the ideal fluid case, the two-dimensional vorticity is conserved along Lagrangian trajectories. A similar formulation exists for three-dimensional flows, where, however, the evolution equation for vorticity contains the vortex stretching term.

Recently, the Lagrangian formulation of fluid dynamics has become important due to the possibility to measure the path of passive tracer particles [46,63,69]. Its importance for the description of turbulence has already been emphasized by Taylor [82] and Richardson [77].

Existence and Smoothness Results

Although the Euler and the Navier–Stokes equations are of fundamental interest for various fields ranging from astrophysics to applied mechanics and engineering, their mathematical properties still remain puzzling. Especially for three-dimensional flows results on the existence (or nonexistence) and smoothness of solutions could not yet be obtained. This topic is one of the millennium problems formulated by the Clay Mathematics Institute. As an introduction to the subject we refer the reader to the webpage of the Clay institute with an outline of the mathematical problem due to Fefferman [90], as well as the two monographs [16,20].

Vortex Solutions of the Navier–Stokes Equation

Figure 2 shows a volume rendering of the absolute value of vorticity above a certain threshold obtained from a direct numerical solution of the vorticity equation. The field is characterized by the presence of elongated vortex structures [39]. Whereas fully developed turbulent flows tend to be dominated by vortex-like objects it seems that modest turbulent flows are characterized by the presence of sheet-like structures. There are several exact solutions of the Navier–Stokes equation, which seem to be related with the vortex structures observed in fully developed turbulence. They can be investigated using symmetry arguments and methods from group theory [31].

Axisymmetric Vortices: Lamb–Oseen Vortex

An axisymmetric vorticity distribution $\boldsymbol{\omega} = \Omega(r)\mathbf{e}_z$ generates a purely azimuthal velocity field. This fact has the consequence that in the vorticity equation both, the convective term $\mathbf{u} \cdot \nabla\boldsymbol{\omega}$ and the vortex stretching term $\boldsymbol{\omega} \cdot \nabla\mathbf{u}$, identically vanish, leading to an equation which is the radially symmetric heat equation

$$\frac{\partial}{\partial t}\Omega = \nu\frac{1}{r}\frac{\partial}{\partial r}r\frac{\partial}{\partial r}\Omega.$$
(18)
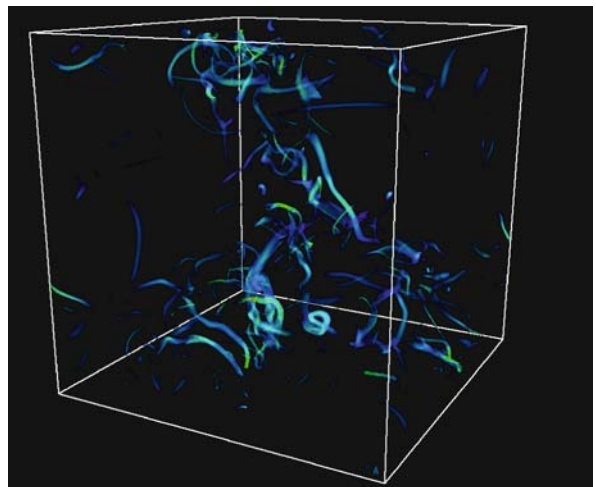
Fluid Dynamics, Turbulence, Figure 2
Direct numerical calculation of the vorticity field of a turbulent fluid motion: Absolute value of vorticity above a given threshold. After [88]
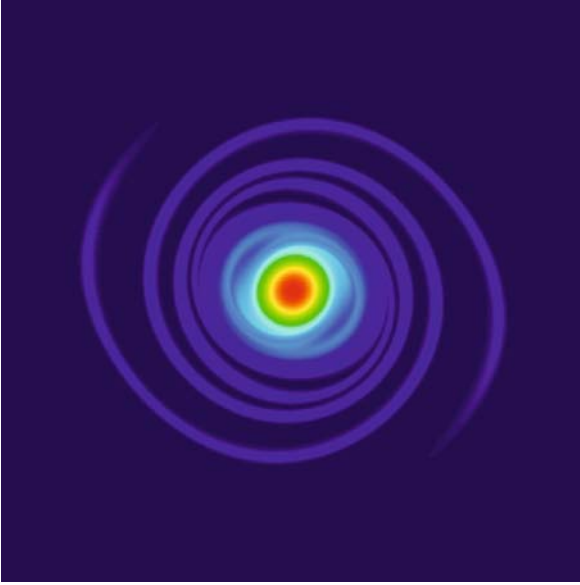
Fluid Dynamics, Turbulence, Figure 3
Vorticity field of a two-armed Lundgren spiral

A solution is the so-called Lamb–Oseen vortex,

$$\Omega = \frac{\Gamma}{\pi r_{\mathrm{B}}^2} e^{-r^2/r_{\mathrm{B}}^2}, \quad r_{\mathrm{B}}^2 = 4\nu t . \tag{19}$$

The corresponding velocity field is azimuthal and has the form

$$v_\varphi(r) = \frac{\Gamma}{2\pi r}\left[1 - e^{-r^2/r_{\mathrm{B}}^2}\right] . \tag{20}$$

It decays like $\Gamma/2\pi r$ for large distances. Because of viscosity the velocity field at the origin vanishes identically.

The Lundgren Spiral

Another vortex solution in the form of a spiral vorticity distribution has been considered by Lundgren [53]. It is generated by a localized central, strong vorticity distribution whose velocity field drags surrounding, weaker vorticity distributions into spiral arms. In this way it is also possible to generate multiple-armed spirals. This process is depicted in Fig. 3. Recently the Lundgren spiral has been generated experimentally by the group of Petitjeans [12].

Stretched Vortices

A highly interesting class of solutions has been found by Lundgren. He considered the velocity field

$$\mathbf{u}(\mathbf{x}, t) = \frac{a(t)}{2}[-x, -y, 2z]$$
$$+ [w_1(x, y, t), w_2(x, y, t), 0] \tag{21}$$

and showed that the two-dimensional velocity field $\mathbf{w}(x, y, t)$ can be obtained from the velocity field $\mathbf{W}(\xi_1, \xi_2, \tau)$ by the *Lundgren transformation*

$$\mathbf{w}(x, y, t) = A(t)\mathbf{W}[A(t)x, A(t)y, \tau(t)]$$
$$A(t) = e^{\int_0^t dt' a(t')}, \quad \tau(t) = \int_0^t dt' A(t')^2 . \tag{22}$$

Thereby, the field $\mathbf{W}(\xi_1, \xi_2, \tau)$ obeys the two-dimensional Navier–Stokes equation.

In the case of a time constant a, the decaying *Lamb–Oseen vortex*, Eq. (19) is changed into the *Burgers vortex*, where r_{B} becomes constant, $r_{\mathrm{B}}^2 = \frac{4\nu}{a}$.

The corresponding two-dimensional velocity field is an azimuthal field, which, for large values of r decays like $\frac{1}{r}$. In the limit $\nu \to 0$ the vorticity field approaches a delta-distribution.

Vorticity Alignment

Figure 2 exhibits the vorticity field obtained from a numerical simulation of the Navier–Stokes equation. Exhibited is a volume rendering of the absolute value of vorticity above a given value and it is quite evident that Burgers-like vortices play a major role in the spatio-temporal organization of turbulence. Consequently, the Burgers vortices have been denoted as the *sinews of turbulence* [60]. Although the emergence of vortex-like objects as organization centers of turbulence has not yet been fully clarified, it is clear that it is related to the phenomenon of vorticity alignment [28]. It has been emphasized that locally the vorticity vector is predominantly aligned to the eigenvector of the intermediate eigenvalue λ_2 of the strain matrix S, Eq. (8). This matrix is symmetric and has three real eigenvalues $\lambda_1 \le \lambda_2 \le \lambda_3$. Vorticity alignment is still investigated intensively [36].

Vorticity alignment has also played a major role in the discussion of the possibility of finite time singularities in Euler flows [30,32], a question which is fundamentally related to the question of the existence of solutions of the Euler and Navier–Stokes equations [16,20].

**Modeling Turbulent Fields
by Random Vortex Distributions**

There have been several attempts to model the fine structure of turbulent fields by statistically distributed vortex solutions of the Navier–Stokes equations. Townsend [84] used a random arrangement of Burgers vortices. As already mentioned Lundgren [53,54] considered the above-mentioned spiral structures in a strain field. He showed

that suitable space-time average over a decaying Lundgren spiral leads to an energy spectrum predicted by Kolmogorov's phenomenological theory of turbulence. A cascade interpretation of Lundgren's model has been given by Gilbert [29]. Kambe [37] discussed randomly arranged Burgers vortices in a strain field and was able to model intermittency effects of Eulerian velocity increments.

Patterns, Chaos, and Turbulence

Pattern Formation and Routes to Chaos in Fluid Dynamics

Experiments on fluid dynamics in confined geometries like the Rayleigh–Bénard system or the Taylor–Couette experiment exhibit a variety of instabilities leading from stationary patterns to time periodic structures and to chaotic motions. For an overview with further references on fluid instabilities we refer the reader to the monograph of Manneville [55]. These flows are characterized by temporal complexity, however the flow structures remain spatially coherent. The scale of energy injection and the scale of energy dissipation usually are not widely separated in these systems. This has the consequence that only few degrees of freedom are excited. Such types of flows can be successfully treated on the basis of the slaving principle of synergetics [34,35]. In mathematical terms, these types of flows are related to the existence of center or inertial manifolds [83] in phase space. This allows explanation of the various *routes to chaos* or *routes to turbulence* observed in fluid motions, especially in confined geometries, on the basis of the theory of low-dimensional dynamical systems.

Point Vortex Motion

The *Lagrangian map* $\mathbf{X}(\mathbf{y}, t)$ of a two-dimensional ideal fluid motion is determined by the solution of the integro-differential Eq. (17), where the Lagrangian vorticity $\Omega(\mathbf{y}, t)$ is temporally constant. This integrodifferential equation can be reduced to a finite set of ordinary differential equations by considering fields with strongly localized vorticity

$$\Omega(\mathbf{y}, t) = \sum_j \Gamma_j \delta(\mathbf{y} - \mathbf{y}_j). \tag{23}$$

Introducing the notation $\mathbf{X}(\mathbf{y}_j, t) = \mathbf{x}_j(t)$ we obtain the set of differential equations for the positions $\mathbf{x}_j(t)$ of the point vortices:

$$\dot{\mathbf{x}}_i = \sum_{j \neq i} \frac{\Gamma_j}{2\pi} \mathbf{e}_z \times \frac{\mathbf{x}_i - \mathbf{x}_j}{|\mathbf{x}_i - \mathbf{x}_j|^2}. \tag{24}$$

This set of equations for the vortex positions was already known by Kirchhoff [41]. Since then, there have been many studies of this problem [3,4,56,64]. We mention that the N-vortex problem of two-dimensional fluid motion shares many properties with the N-body problem of classical mechanics.

The Lagrangian motion of an arbitrary point $\mathbf{X}(\mathbf{y}, t)$ can be determined by the solution of the nonautonomous differential equation

$$\dot{\mathbf{X}}(\mathbf{y}, t) = \sum_{i=1}^{N} \frac{\Gamma_i}{2\pi} \mathbf{e}_z \times \frac{\mathbf{X}(\mathbf{y}, t) - \mathbf{x}_i(t)}{|\mathbf{X}(\mathbf{y}, t) - \mathbf{x}_i(t)|^2}, \tag{25}$$

where the point vortex positions $\mathbf{x}_i(t)$ are given by the evolution equations (24). The point vortex approximation, thus, allows one to study *mixing in two-dimensional flows* on the basis of sets of ordinary differential equations [3,70,81].

Since the point vortex system is obtained from the two-dimensional Euler equation it is evident that the kinetic energy is conserved. In fact, introducing the Hamilton function

$$\begin{aligned} H &= -\frac{1}{4\pi} \sum_{i \neq j} \Gamma_i \ln |\mathbf{x}_i - \mathbf{x}_j| \Gamma_j \\ &= \frac{1}{2} \sum_{i \neq j} \Gamma_i G(\mathbf{x}_i, \mathbf{x}_j) \Gamma_j \end{aligned} \tag{26}$$

we can rewrite the evolution equations in Hamiltonian form

$$\Gamma_i \dot{\mathbf{x}}_i = \mathbf{e}_z \times \nabla H. \tag{27}$$

There are further conserved quantities due to symmetry. The *center of vorticity*,

$$\mathbf{R} = \sum_i \Gamma_i \mathbf{x}_i, \tag{28}$$

as well as the quantity

$$I = \sum_i \Gamma_i \mathbf{x}_i^2 \tag{29}$$

are time constants.

The advantage of considering the motion of point vortices in two-dimensional ideal flows is the applicability of methods and notions from the theory of finite Hamiltonian systems. Explicitly, one can investigate the integrability of point vortex motions, i. e. the formation of time periodic as well as quasiperiodic vortex motions and the emergence of chaotic motions. Because of the existence of
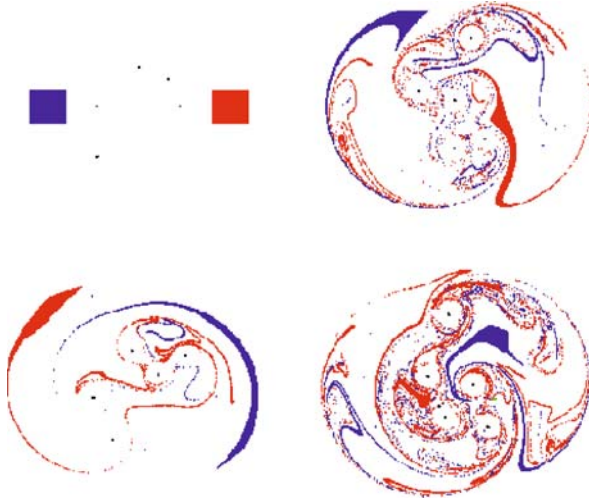
Fluid Dynamics, Turbulence, Figure 4
Mixing induced by the motion of four point vortices

the integrals of motions (26), (28), (29) the motion of two vortices as well as three point vortices is integrable leading to quasiperiodic motions in time. Furthermore, it has been shown that in the four vortex problem, in addition to quasiperiodic motions, chaotic motion is possible. We refer the reader to the review articles of Aref [3,4] and the monograph of Newton [64]. Also the investigations of mixing by point vortex motion has revealed interesting insights into the geometry and the complexity of the mixing process. An example of the missing process by four point vortices is exhibited in Fig. 4. Similar concepts are expected to be applicable for general two- and three- dimensional turbulent flows.

Onsager's Statistical Theory
of Two-Dimensional Turbulence

Onsager [68] recognized the importance of two-dimensional vortex models for turbulent flows. Since the dynamics is Hamiltonian, one can perform a statistical treatment along the lines of equilibrium statistical mechanics. One can determine the corresponding probability distributions. If we focus onto the canonical ensemble, i. e. a point vortex system in connection with a heat bath, the probability distribution reads

$$f(\mathbf{x}_i) = Z^{-1}(\beta) e^{-\beta \left[\frac{1}{2} \sum_{ij} \Gamma_i G(\mathbf{x}_i, \mathbf{x}_j) \Gamma_j \right]} = Z^{-1}(\beta) e^{-\beta H} . \tag{30}$$

Onsager recognized that this probability distribution is normalizable both for negative as well as for positive values of β, which in statistical mechanics is related to tempera-

ture. A consequence of the existence of these *negative temperature states* is the tendency of the vortices to form large-scale flow structures. This property can be seen quite easily by recognizing that the probability distribution is the stationary probability distribution of the set of Langevin equations

$$\frac{\mathrm{d}}{\mathrm{d}t}\mathbf{x}_i = -\beta \nabla_{\mathbf{x}_i} H + \eta_j . \tag{31}$$

Here, η_j represents Gaussian white noise. For negative temperatures the force between two point vortices is attractive for $\Gamma_i \Gamma_j > 0$ and repulsive in the other case. This leads to the formation of large-scale flows. We refer the reader to the recent review [18].

Extension to Three Dimensions

The extension of the philosophy of point vortex motion to three dimensions leads one to consider vortex filaments. The location of a single vortex filament is given by the local induction equation for the position $\mathbf{X}(s, t)$ of the filament as a function of arclength s and time t. This equation is again obtained from Biot–Savart's law assuming a filamentary vorticity distribution:

$$\frac{\partial}{\partial t}\mathbf{X}(s, t) = \frac{\partial}{\partial s}\mathbf{X}(s, t) \times \frac{\partial^2}{\partial s^2}\mathbf{X}(s, t) \tag{32}$$

It has been shown that the single filament equation is integrable. For a discussion we refer the reader to [64].

Turbulence: Determinism and Stochasticity

It is evident that an understanding of the characteristics of turbulent fluid motions has to be based on the deterministic dynamics generated by the basic fluid dynamical equations in combination with methods of statistical physics. This has led to the field of *Statistical Hydrodynamics*, a topic which has been treated extensively by the Russian school founded by Kolmogorov [61,62]. A good overview can be found in [26,50].

Statistical Averages

In *Statistical Hydrodynamics* the Eulerian velocity field and related fields are treated as *random fields* in a probabilistic sense. To this end one has to define suitable averages, where usually *ensemble averages* are chosen. They are defined by specifying the statistics of the initial flow field by probability distributions $f(\mathbf{u}_1, \mathbf{x}_1, t = 0)$, $f(\mathbf{u}_1, \mathbf{x}_1, t = 0; \ldots; \mathbf{u}_N, \mathbf{x}_N, t = 0)$ of the velocities $\mathbf{u}_i$ at positions $\mathbf{x}_i$ at initial time $t = 0$. The transition to a continuum of points requires one to consider a probability density functional $F[\mathbf{u}(\mathbf{x}), t = 0]$.

In practice, instead of ensemble averages, time averages are taken, provided the flow is stationary in a statistical sense.

The corresponding probability distributions at time t, which specify the temporal evolution of the considered statistical ensemble are given by

$$f(\mathbf{u}, \mathbf{x}, t) = \langle \delta(\mathbf{u} - \mathbf{u}(\mathbf{x}, t; \mathbf{u}_0)) \rangle . \tag{33}$$

Here, $\mathbf{u}(\mathbf{x}, t; \mathbf{u}_0)$ is the solution of the fluid dynamic equation with the initial value of the velocity field at point $\mathbf{x}$: $\mathbf{u}(\mathbf{x}, t = 0; \mathbf{u}_0) = \mathbf{u}_0$. The brackets denote an ensemble average. Joint probability distributions and the probability functional $F(\mathbf{u}(\mathbf{x}, t))$ are defined accordingly. Frequently, the characteristic functions, defined as the Fourier transform of the probability distributions are used. The characteristic functional

$$Z(\alpha) = \langle e^{i \int dt \int d^3\mathbf{x}\,\alpha(\mathbf{x},t) \cdot \mathbf{u}(\mathbf{x},t)} \rangle \tag{34}$$

is the functional Fourier transform of the probability functional. The ultimate goal of statistical hydrodynamics is the determination of this functional, since it contains all information on the various correlation functions of the Eulerian velocity fields of the statistical ensemble.

Similar probability distributions can be defined for the Lagrangian description of fluid dynamics.

Hierarchy of Moment Equations

The temporal evolution of moments of the velocity field are determined by the basic fluid dynamical equations. For instance, an equation for the moment $\langle u_i(\mathbf{x}, t)u_j(\mathbf{x}', t) \rangle$ can be obtained in a straightforward manner from the Navier–Stokes equation

$$
\begin{aligned}
&\frac{\partial}{\partial t}\langle u_i(\mathbf{x}, t)u_j(\mathbf{x}', t)\rangle + \sum_k \frac{\partial}{\partial x_k}\langle u_k(\mathbf{x}, t)u_i(\mathbf{x}, t)u_j(\mathbf{x}', t)\rangle \\
&+ \sum_k \frac{\partial}{\partial x'_k}\langle u_k(\mathbf{x}', t)u_i(\mathbf{x}, t)u_j(\mathbf{x}', t)\rangle \\
&= -\frac{\partial}{\partial x_i}\langle p(\mathbf{x}, t)u_j(\mathbf{x}', t)\rangle - \frac{\partial}{\partial x'_j}\langle p(\mathbf{x}', t)u_i(\mathbf{x}, t)\rangle \\
&+ \nu[\Delta_x + \Delta_{x'}]\langle u_i(\mathbf{x}, t)u_j(\mathbf{x}', t)\rangle .
\end{aligned}
\tag{35}
$$

Similar equations can be formulated for the higher-order moments $\langle u_i(\mathbf{x}_1, t_1)u_j(\mathbf{x}_2, t_2) \ldots u_j(\mathbf{x}_N, t_N) \rangle$ using the Navier–Stokes equation. The equation for the average flow field $\langle \mathbf{u}(\mathbf{x}, t) \rangle$ has been considered by O. Reynolds [76] (see the discussion below). The chain of evolution equations for the higher-order correlation functions are the so-called Friedmann–Keller equations.

The moment equations are not closed. The evolution equation containing the Nth order moment contains the $(N+1)$-th order moment. This is the famous closure problem of turbulence. It is an immediate consequence of the nonlinearity of the Navier–Stokes equation. A mathematical discussion of the closure problem is given in [27].

Evolution Equations for Probability Distributions

It is straightforward to derive evolution equations for the probability distributions from the Navier–Stokes equation. This has been emphasized by Lundgren [52] and Ulinich and Lyubimov [86]. The evolution equation for $f(\mathbf{u}, \mathbf{x}, t)$ reads

$$
\begin{aligned}
&\left[\frac{\partial}{\partial t} + \mathbf{u} \cdot \nabla_\mathbf{x}\right] f(\mathbf{u}, \mathbf{x}, t) \\
&= -\nabla_\mathbf{u} \cdot \int d\mathbf{u}'\, d\mathbf{x}'\, \mathbf{K}(\mathbf{x} - \mathbf{x}')(\mathbf{u}' \cdot \nabla_{\mathbf{x}'})^2 f(\mathbf{u}', \mathbf{x}', t; \mathbf{u}, \mathbf{x}, t) \\
&\quad - \nabla_\mathbf{u} \cdot \int d\mathbf{u}'\, d\mathbf{x}'\, \delta(\mathbf{x} - \mathbf{x}')\Delta_{\mathbf{x}'}\mathbf{u}'\, f(\mathbf{u}', \mathbf{x}', t; \mathbf{u}, \mathbf{x}, t) .
\end{aligned}
$$
$$\tag{36}$$

(Here, $\mathbf{K}(\mathbf{x})$ is the gradient of Green's function $G(\mathbf{x})$ of the Laplacian.) The dynamics of the single-point probability distribution $f(\mathbf{u}, \mathbf{x}, t)$ is coupled to the two-point probability distribution $f(\mathbf{u}', \mathbf{x}', t; \mathbf{u}, \mathbf{x}, t)$ by the nonlocal pressure term and the dissipative term. Similar equations can be obtained for higher-order probability distributions. Again, the hierarchy shows clearly the closure problem of turbulence theory. The hierarchy is of considerable interest for the so-called Lagrangian pdf (probability density function) model approach advocated by S.B. Pope [72,73]. In this approach, the terms involving the two-point pdf's are modeled leading to a description of the turbulent velocity in terms of a stochastic process.

For the case of Burgers equation, which is the Navier–Stokes equation without the pressure term, it has been possible to solve the corresponding Lundgren hierarchy without approximation for a certain external forcing [17].

Functional Equations

The Lundgren hierarchy of evolution equations arises due to the fact that the N-point probability distributions contain incomplete information on the evolution of the fluid continuum. A closed evolution equation is obtained for the characteristic functional (34). This equation is the famous Hopf functional equation [38], which forms a concise formulation of the statistical treatment of fluid motions. For a more detailed treatment we refer the reader to the monograph of Monin and Yaglom [62].

Path-Integral Formulation

As has been emphasized by Martin, Siggia and Rose [58], each classical field theory can be represented in terms of the path integral formalism. This is essentially true for a fluid motion driven by a fluctuating force, which is Gaussian and δ-correlated in time. The generating (MSR) functional has the following path integral representation

$$
\begin{aligned}
&Z(\alpha, \hat{\alpha}) \\
&= \int D\mathbf{u}D\hat{\mathbf{u}}\mathbf{e}^{S[\mathbf{u},\hat{\mathbf{u}}]+i\int d\mathbf{x}\int dt\boldsymbol{\alpha}(\mathbf{x},t)\cdot\mathbf{u}(\mathbf{x},t)+\hat{\boldsymbol{\alpha}}(\mathbf{x},t)\cdot\hat{\mathbf{u}}(\mathbf{x},t)} .
\end{aligned}
\tag{37}
$$

The functional $Z(\alpha, \hat{\alpha} = 0)$ is just the Hopf characteristic functional (34). The MSR-action is defined according to

$$
\begin{aligned}
S = {}& i \int d\mathbf{x} \int dt \hat{\mathbf{u}}(\mathbf{x}, t) \\
& \cdot \left\{ \dot{\mathbf{u}}(\mathbf{x}, t) + \mathbf{u}(\mathbf{x}, t) \cdot \nabla \mathbf{u}(\mathbf{x}, t) - \nu \Delta \mathbf{u}(\mathbf{x}, t) + \nabla p(\mathbf{x}, t) \right\} \\
& - \frac{1}{2} \int dt \int d\mathbf{x}' \int d\mathbf{x}\hat{\mathbf{u}}(\mathbf{x}, t)Q(\mathbf{x} - \mathbf{x}')\hat{\mathbf{u}}(\mathbf{x}', t) .
\end{aligned}
\tag{38}
$$

The MSR-formalism is a convenient starting point for an analytical determination of correlation functions of the velocity field. A naive perturbation expansion of this functional yields the diagrammatic representation of the series by Wyld [89]. A renormalized perturbation expansion leads to the so-called direct interaction approximation (DIA) of Kraichnan and related analytical approximations. For an overview we refer the reader to the monographs of Lesieur [49] and McComb [59]. Also the recent work by V. L'vov and I. Procaccia [45] is based on this approach.

Reynolds Equation and Turbulence Modeling

O. Reynolds [76] suggested to decompose a turbulent flow field $\mathbf{u}(\mathbf{x}, t)$ into a mean flow $\langle\mathbf{u}\rangle$ and turbulent pulsations $\mathbf{w}(\mathbf{x}, t)$:

$$
\mathbf{u}(\mathbf{x}, t) = \langle\mathbf{u}(\mathbf{x}, t)\rangle + \mathbf{w}(\mathbf{x}, t) .
\tag{39}
$$

By averaging the Navier–Stokes equation one ends up with the famous Reynolds's equation, which is the first equation of the hierarchy of moment equations (35):

$$
\begin{aligned}
&\left[\frac{\partial}{\partial t} + \langle\mathbf{u}\rangle \cdot \nabla \right] \langle\mathbf{u}\rangle \\
&= -\nabla p - \sum_{ij} \frac{\partial^2}{\partial x_i \partial x_j} \langle w_i w_j \rangle + \nu \Delta \langle\mathbf{u}\rangle + \langle\mathbf{f}\rangle .
\end{aligned}
\tag{40}
$$

This equation contains the so-called Reynolds stress tensor $\langle w_i w_j \rangle$, which cannot be neglected since the turbulent pulsations $\mathbf{w}$ can be larger than the averaged velocity $\langle\mathbf{u}\rangle$. The turbulent pulsations $\mathbf{w}$ are closely linked to the *fine structures of turbulence.*

If the Reynolds stress tensor is known as a functional of the mean velocity field, the Reynolds equation is a closed evolution equation determining average flow properties. This makes numerical computations of the average flow quantities rather efficient since the *fine structures* or the *small-scale flow* have not to be resolved by numerical schemes.

The Reynolds stress tensor has been the subject of various investigations. Since a general theory determining the Reynolds stress tensor is lacking, engineers have developed the area of *turbulence modeling* in order to overcome the closure problem. Famous turbulence models are eddy viscosity models, which replace the Reynolds stress tensor by an effective damping term modeling the energy flux from the averaged flow into the turbulent pulsations, or so-called $K - \epsilon$ models, which are based on the evolution equation for the local energy dissipation rate ϵ of the turbulent pulsations.

Turbulence modeling has to fulfill the *requirement of physical realizability* and should, for example, not lead to the development of negative kinetic energies. Furthermore, symmetry arguments should be taken into account [65].

Turbulence modeling has led to the development of the field of *Large Eddy Simulations* (LES), which has provided a variety of numerical, also commercially available schemes for calculating the large-scale flows of applied fluid dynamical problems. The LES-approach, however, is limited by the fact that the properties of the Reynolds stresses cannot yet be derived from a physical treatment of the small-scale turbulent pulsations of flows and the obtained numerical results have to be met with caution. For details we refer the reader to Piquet [71], Jovanovic [40], and the reviews of Métais and Leschziner in [50]. Of considerable interest are the two articles of Johansson and Oberlack in [66].

The Fine Structure of Turbulence

The fine structure of turbulence essentially influences the large-scale flows via the Reynolds stresses. Therefore, the investigation of the fine structure of turbulent flows is a central theme in turbulence research. It is commonly believed that the statistical characteristics of the turbulent fine structures are universal. A point emphasized by U. Frisch [26] is that in the fine structures symmetries of

the Euler equation of fluid dynamics are restored in a statistical sense. The symmetries of the Euler equations are translational symmetry, isotropy, and rescaling symmetry of space, time, and velocity. Although each of these symmetries are broken by the turbulent flows, the symmetries are restored for averaged quantities. If this hypothesis is true, then the turbulent fine structure is related with a universal state, which is called *fully developed stationary, homogeneous, and isotropic turbulence.*

Increments

The fine-scale structure of turbulence is evaluated by the introduction of the so-called velocity increment $v_x(r, t)$

$$v_x(r, t) = u(x + r, t) - u(x, t) . \tag{41}$$

Velocity increments are defined with respect to a reference point x. A mean flow is eliminated by the definition of increments. Furthermore, one may consider velocity increments with respect to a moving reference point $x = X(y, t)$. In the following we shall consider such moving increments.

The corresponding evolution equation for the incompressible velocity increment can easily be established using the definition (41) and the Navier–Stokes equation:

$$\left[\frac{\partial}{\partial t} + v(r, t) \cdot \nabla_r \right] v(r, t)$$
$$= - \nabla_r p(r, t) + \nu \Delta_r v(r, t)$$
$$- \left[- \nabla_r p(r, t) + \nu \Delta_r v(r, t) \right]_{r=0} ,$$
$$\nabla \cdot v_x(r, t) = 0 . \tag{42}$$

Length and Time Scales in Turbulent Flows

The Integral Scale   The integral scales are measures for a spatial distance L or a time interval T, across which the turbulent fluctuations become uncorrelated. The integral length scale is based on the velocity-velocity correlation function

$$\langle u(x + r, t)u(x, t) \rangle = \langle u(x, t)u(x, t) \rangle F\left(\frac{r}{L} \right) , \tag{43}$$

which decays to zero as a function of the distance r. The integral length scale L is defined by the integral

$$L = \int_0^\infty \mathrm{d}r F\left(\frac{r}{L} \right) . \tag{44}$$

Similarly, one can define temporal integral scales based on the decay of temporal correlations $\langle u(x, t + \tau)u(x, t) \rangle$

$= \langle u(x, t)u(x, t) \rangle F\left(\frac{\tau}{T} \right)$ leading to the definition of the integral time scale T. An integral velocity scale can be formed quite naturally by the ratio

$$U_{\text{int}} = L/T . \tag{45}$$

The velocity field $u(x, t)$ describes the turbulent pulsations. A mean flow has already been subtracted. The correlation functions (43) actually are tensors. For flows which are nonisotropic in the statistical sense, it might be necessary to introduce different integral length scales.

The Kolmogorov Scales   The Navier–Stokes equation involves the kinematic viscosity ν as well as a measure of the excitation of turbulence, the mean local energy dissipation ϵ. It is convenient to form length and time scales from these two quantities and obtain the so-called Kolmogorov scales

$$\eta = \left(\frac{\nu^3}{\epsilon} \right)^{1/4} , \quad \tau_\eta = \left(\frac{\nu}{\epsilon} \right)^{1/2} , \quad u_\eta = \frac{\eta}{\tau_\eta} = (\epsilon\nu)^{1/4} . \tag{46}$$

On the basis of these quantities one can define a Reynolds number which turns out to be unity, $\text{Re} = \frac{\eta u_\eta}{\nu} = 1$. Therefore, the Kolmogorov scales have to be related with small scale motions, which, due to dissipation, can be considered to be laminar on these scales.

Relation Between the Integral and the Kolmogorov Length Scale   One may find a relation between the integral length scale L and the Kolmogorov length scale η. This relation is based on the observation that the local energy dissipation rate ϵ can be dimensionally expressed in terms of the velocity at the integral scale, U_{int}, Eq. (45) via

$$\epsilon \approx \frac{U_{\text{int}}^3}{L} . \tag{47}$$

This relation allows one to determine the ratio of the intergal scale L and the Kolmogorov scale η leading to a Reynolds number dependence

$$\frac{L}{\eta} \approx \text{Re}^{3/4} = \left(\frac{U_{\text{int}} L}{\nu} \right)^{3/4} . \tag{48}$$

The ratio increases with the Reynolds number. Typical length scales estimated from experimental data of a grid flow are shown in Fig. 5. It is seen that the estimate (48) can be confirmed experimentally.

From a dynamical point of view the quantity $\left(\frac{L}{\eta} \right)^3 \approx \text{Re}^{9/4}$ is a measure of the number of active degrees of freedom of the fluid motion.
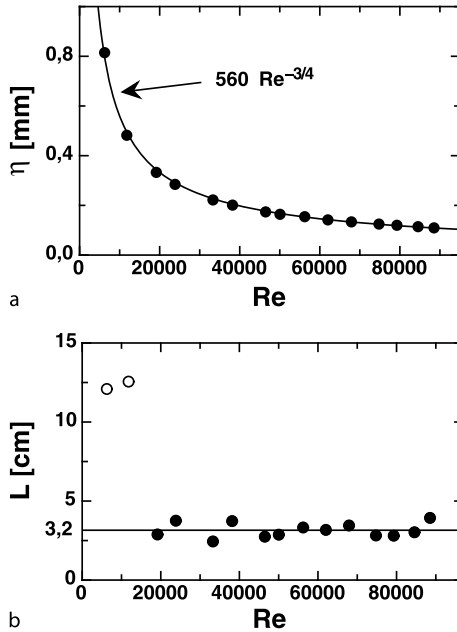
Fluid Dynamics, Turbulence, Figure 5
Reynolds number dependence of the Kolmogorov scale (**a**) and the integral length scale (**b**) estimated from data of a grid experiment, [51]

Fluid Dynamics, Turbulence, Figure 6
Reynolds number dependence of the Taylor scale (**a**) and the Taylor length-based Reynold number from data of a grid experiment, [51]

The Taylor Length and the Taylor-Based Reynolds Number   A third length scale has been used to characterize a turbulent flow field. This is the so-called Taylor length, which frequently is denoted also as Taylor microscale, λ.

It can be derived from the velocity u (here u denotes again the fluctuating part of a measured velocity signal) and the derivative $\partial u/\partial x$ as

$$\lambda^2 = \lim \frac{\langle u^2 \rangle}{\langle (\partial u/\partial x)^2 \rangle} . \tag{49}$$

Aronson and Löfdahl [6] suggested to estimate the Taylor length using the velocity increment $v(r) = u(x+r) - u(x)$:

$$\lambda^2 = \lim_{r \to 0} \frac{\langle u(x)^2 \rangle \, r^2}{\langle v(r)^2 \rangle} . \tag{50}$$

For the Taylor length it is expected that $\frac{L}{\lambda} \approx \mathrm{Re}^{1/2}$ holds, which is also found experimentally see Fig. 6a.

On the basis of this Taylor length it is also possible to define a new Reynolds number by the ratio of velocity times length scale and kinematic viscosity. The Taylor length-based Reynolds number R_λ should scale with the square of the usual Reynolds number, see Fig. 6b.

Statistics of Increments: Structure Functions

A great amount of work in turbulence research has been devoted to the so-called structure functions. In general, structure functions are equal time moments of the velocity increments:

$$\langle v_i(\mathbf{r}, t) v_j(\mathbf{r}, t) \rangle , \quad \langle v_i(\mathbf{r}, t) v_j(\mathbf{r}, t) v_k(\mathbf{r}, t) \rangle . \tag{51}$$

For stationary, homogeneous, and isotropic turbulence the tensorial quantities can be considerably reduced by symmetry arguments. For homogeneous and isotropic turbulence the second- and third-order moments can be related to the so-called longitudinal structure functions, i. e. the moments of the component of the velocity increment in the direction of $\mathbf{r}$, $\mathbf{e}_r \cdot \mathbf{v}(\mathbf{r}, t)$. They are defined according to

$$S^N(r) = \left\langle \left(\frac{\mathbf{r}}{r} \cdot \mathbf{v}(\mathbf{r}, t) \right)^N \right\rangle . \tag{52}$$

For small values of r, the structure functions have to behave like $S^N \approx r^N$, since for small r the velocity increment can be expanded in a Taylor series.

The structure functions of different orders are related through the Navier–Stokes dynamics. The evaluation of the structure functions from the resulting hierarchy of evolution equations is one of the major theoretical challenges in turbulence research. However, it seems that

so far only the lowest-order equations relating second- and third-order structure (the so-called Kolmogorov's -4/5 law) functions have been exploited rigorously.

Kolmogorov's -4/5 Law

Kolmogorov (c. f. [26,62]) showed that the mean kinetic energy in an *eddy of scale r*, $\langle \mathbf{v_x}(\mathbf{r}, t)^2 \rangle$, in a homogeneous, isotropic turbulent field is given by the equation

$$\frac{\partial}{\partial t}\left\langle \frac{\mathbf{v}_x(\mathbf{r},t)^2}{2} \right\rangle + \nabla_r \cdot \left\langle \mathbf{v}_x(\mathbf{r},t)\frac{\mathbf{v}_x(\mathbf{r},t)^2}{2} \right\rangle$$
$$= \nu \Delta \langle \mathbf{v_x}(\mathbf{r},t)^2 \rangle - \langle \epsilon \rangle , \quad (53)$$

where $\langle \epsilon \rangle$ denotes the mean local energy dissipation rate

$$\langle \epsilon \rangle = \frac{\nu}{2} \sum_{ij} \left\langle \left(\frac{\partial u_i}{\partial x_j} + \frac{\partial u_j}{\partial x_i} \right)^2 \right\rangle . \quad (54)$$

We formally consider the *turbulent limit* $\nu \to 0$. In this limit the mean local energy dissipation rate $\langle \epsilon \rangle$ has to be constant. Thus, one can estimate that the gradients have to behave like

$$\frac{\partial u_i}{\partial x_j} \approx \frac{1}{\sqrt{\nu}} . \quad (55)$$

The relation (53) can be obtained as a balance equation for the density of the mean kinetic energy from the evolution equation for the velocity increment (42) by scalar multiplication with $\mathbf{v}$, and subsequent averaging. The correlations involving the pressure term drop out due to homogeneity.

The resulting equation reads:

$$S^3(r) - 6\nu \frac{\mathrm{d}}{\mathrm{d}r} S^2(r) = -\frac{4}{5} \langle \epsilon \rangle r . \quad (56)$$

This equation relates the second- and third-order structure functions and, at first glance, seems to be underdetermined. However, for small values of r we can neglect the third-order structure function, since $S^3(r) \approx r^3$. The second-order structure function is then given by

$$S^2(r) = \frac{\langle \epsilon \rangle}{15\nu} r^2 . \quad (57)$$

The range of validity of this law is denoted as the *dissipative range* and is close to the Kolmogorov length and smaller. It is dominated by viscous structures.

In the second regime, whose existence is inferred by the -4/5th law, the third-order structure function dominates, leading to

$$S^3(r) = -\frac{4}{5} \langle \epsilon \rangle r . \quad (58)$$

This defines the so-called *inertial range* and is located be-

tween the Taylor length and the integral length. For data from a free jet experiment with high Reynolds numbers we show in Fig. 7 the third-order structure function in different presentations, which clearly demonstrate the existence of the inertial range.
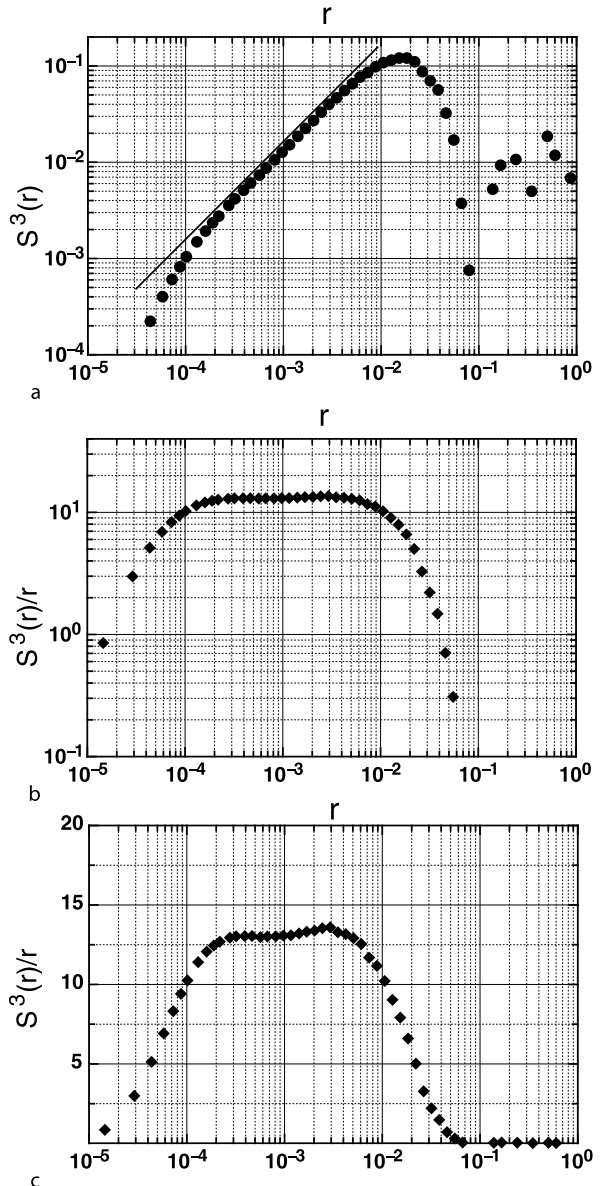
Fluid Dynamics, Turbulence, Figure 7
Third-order structure function $S^3(r)$ (**a**) of cryogenic free jet measurements with Re = 210.000 and the compensated structure functions $S^3(r)/r$ in a semi logarithmic plot (**b**) and in a linear plot (**c**), showing in more detail the quality of the present scaling behavior. (Note here we used the absolute values for $S^3(r)$, thus values are all positive). [9]

Phenomenological Theories of Turbulence

Kolmogorov's Theory K41:
Selfsimilarity in the Inertial Range

The inertial range is characterized by the decay of eddies, which can be assumed to be self-similar. As a consequence, the probability distribution of the longitudinal velocity increment v at scale r, $f(v, r)$, should have the form

$$f(v, r) = \frac{1}{\sqrt{\langle v(r)^2 \rangle}} F\left(\frac{v}{\sqrt{\langle v(r)^2 \rangle}}\right), \tag{59}$$

where $F(\xi)$ is a universal function in the range of scales $\lambda < r < L$. As a consequence, the nth order moments should have the form

$$\langle v^n(r) \rangle = \int dv \, v^n \frac{1}{\sqrt{\langle v(r)^2 \rangle}} F\left(\frac{v}{\sqrt{\langle v(r)^2 \rangle}}\right)$$
$$= \langle v(r)^2 \rangle^{n/2} \int dw \, w^n F(w) = \langle v(r)^2 \rangle^{n/2} V_n. \tag{60}$$

Since the r-dependence of the 3rd-order moment is known in the limit of high Reynolds number,

$$\langle v^3(r) \rangle = V_3 \langle \epsilon \rangle r = \langle v(r)^2 \rangle^{3/2} V_3, \tag{61}$$

we obtain

$$\langle v(r)^2 \rangle = K (\langle \epsilon \rangle r)^{2/3}. \tag{62}$$

The K41 assumption of self-similarity of the velocity increment statistics in the inertial range leads to the following fractal scaling behavior of the nth order moments

$$\langle |v(r)|^n \rangle = K_n (\langle \epsilon \rangle r)^{n/3}. \tag{63}$$

The corresponding probability distribution is entirely determined by the constants K_n. The scaling exponents $\zeta_n = N/3$ are linear functions of n.

Failure of K41: Intermittency

Kolmogorov's hypothesis on the self-similarity of the statistics in the inertial range has been tested experimentally as well as numerically. Figure 8 exhibits the probability distribution of the scaled variable $\tilde{v}$

$$\tilde{v} = \frac{v}{\sigma(r)} \quad \text{with } \sigma(r) = \sqrt{\langle v(r)^2 \rangle}. \tag{64}$$

Because of K41 theory all probability distributions should collapse for values of r taken from the inertial range. Both, experimental and numerical results show clear deviations from this behavior. Although the scaling exponents of
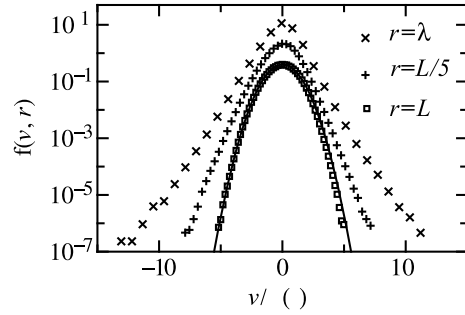
Fluid Dynamics, Turbulence, Figure 8
Probability density functions (pdf) for velocity increments on three different length scales $r = L, L/5, L/30 \approx \lambda$. The pdfs are shifted along the ordinate for a better representation. On the largest scale L a Gaussian distribution is fitted for comparison. Towards smaller scales the deviations from the Gaussian form become obvious ($R_\lambda = 180$). For further details see [79]

the second-order structure function is close to the K41 value 2/3, a characteristic change of shape of the probability distribution can be detected. The change of the pdf with scale r in the inertial range is a signature of the phenomenon called *intermittency*. Consequently, the structure functions do not scale as suggested by the theory of Kolmogorov. This has been experimentally documented by [1] and has been discussed by several groups.

Kolmogorov's Theory K62

In a famous note in the volume on hydrodynamics in his Course on Theoretical Physics Landau [47] remarked that the formula of the K41 theory contains the mean value of the local energy dissipation rate, $\langle \epsilon \rangle$. However, this quantity is a strongly fluctuating quantity in space, due to the strong spatial variations of the velocity gradient, as can be seen from Fig. 9. According to Landau these fluctuations should show up in the structure functions. Instead of the K41-result (63) he suggested use of the following representation

$$\langle |v(r)|^n \rangle = \tilde{K}_n \langle \epsilon_r^{n/3} \rangle r^{n/3} \tag{65}$$
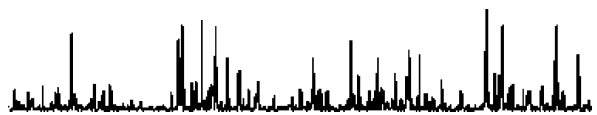
Fluid Dynamics, Turbulence, Figure 9
Spatial distribution of the local energy dissipation rate. The quantity strongly fluctuates in space. Data is obtained from the experiment [79]

where the quantity ϵ_r denotes the local energy dissipation rate averaged over a sphere of radius r. This reasoning leads to an extension of the K41 formula for the probability distribution in the form

$$f(v, r) = \int d\epsilon_r \, p\left(\epsilon_r, \frac{r}{L}\right) \frac{1}{(\epsilon r)^{2/3}} F\left(\frac{v}{(\epsilon_r r)^{2/3}}\right) . \quad (66)$$

In 1962 Kolmogorov [44] suggested use of a lognormal distribution for the local energy dissipation rate ϵ_r, whose variance is $\mu \ln(L/r)$. As a result, the structure functions scale like

$$\langle |v(r)|^n \rangle = \tilde{K}_n r^{n/3} (L/r)^{\mu n(n-3)} . \quad (67)$$

The experimental value for μ can be obtained from a fit of the K62 formula to experimental data. However, we note that the formula can only be valid for small values of n, since the scaling exponents ζ_n have to be a monotonously increasing function of the order n of the structure function.

The Multifractal Model

The multifractal model was introduced by U. Frisch and G. Parisi. For a detailed description we refer the reader to [26]. The basic idea is to view a turbulent field to be composed of regions where the velocity increment field is assumed to be characterized by a scaling index h:

$$v(r, t) = \beta \left(\frac{r}{L}\right)^h . \quad (68)$$

The structure functions are then given by

$$S^N(r) = \beta_N \int dh P(h, r) \left(\frac{r}{L}\right)^{Nh} , \quad (69)$$

where $P(r, h)dh$ is the probability to find increments for a certain scale r with scaling exponent h. Assuming self-similarity $P(h, r)$ should have the form

$$P(h, r) = \left(\frac{r}{L}\right)^{3-D(h)} . \quad (70)$$

Consequently, we have

$$S^N(r) = \beta_N \int dh \left(\frac{r}{L}\right)^{Nh+3-D(h)} \approx r^{\zeta_N} , \quad (71)$$

where the evaluation of the integral with the method of steepest descend yields

$$\zeta_N = \text{Min}[Nh + 3 - D(h)] . \quad (72)$$

The scaling indices ζ_N are related to the dimension $D(h)$ via a Legendre transform. Recently, an extension of the

multifractal model has been presented by Chevillard et al. [10]. This approach is essentially based on the representation of the probability distribution due to [8]. They succeeded to obtain a model for the symmetric part of the probability function of the longitudinal velocity increment, which is valid both in the integral as well as the dissipative scale. This gives a reasonable approximation to the experimentally determined probability distribution $f(v, r)$.

As we have seen, the statistics of the longitudinal velocity increment for a single scale r can be modeled in various ways in order to describe the deviations from the fractal scaling behavior predicted by the phenomenological theory of Kolmogorov formulated in 1941. The major shortcoming of these approaches, however, is the fact that they contain no information on the joint statistics of the velocity fields at different scales and times. However, due to the presumed energy cascade, the velocity increments on different scales have to be correlated.

Multiscale Analysis of Turbulent Fields

The spatial correlation of the velocity of turbulent fields has been examined in two ways. From a field theoretic point of view, pursued by V. L'vov and I. Procaccia, the existence of so-called fusion rules have been hypothesized. For this approach we refer the reader to the survey [45] and the work cited therein. A phenomenological approach has been performed by Friedrich and Peinke [22,23]. In this approach notions from the theory of stochastic processes have been used in order to characterize multiscale statistics of velocity increments. Relations to the fusion rule approach has been discussed in [15]. In the following we shall discuss the phenomenological approach to multiscale statistics of turbulence.

Statistics Across Scales

In order to address the spatial signatures of the cascading process underlying stationary, homogeneous, and isotropic turbulence it is necessary to consider the probability distributions for velocity increments at different scales:

$$f^N(\mathbf{v}_1, r_1; \mathbf{v}_2, r_2; \ldots; \mathbf{v}_N, r_N) . \quad (73)$$

Thereby, the quantities $\mathbf{v}_i$ are velocity increments with respect to a common point of reference, $\mathbf{v}_i(\mathbf{r}_i, t) = \mathbf{u}(\mathbf{x} + \mathbf{r}_i, t) - \mathbf{u}(\mathbf{x}, t)$ and different distances $\mathbf{r}_i$.

In the following we shall consider the longitudinal velocity increments and locations $\mathbf{r}_i$ positioned along a straight line. These quantities are easily accessible from experimental data.

In case that the longitudinal velocity increments v_i of different scales r_i are statistically independent, the N-point probability distribution simply factorizes:

$$
f^N(v_1, r_1; v_2, r_2; \ldots; v_N, r_N) \\
= f^1(v_1, r_1) f^1(v_2, r_2) \ldots f^1(v_N, r_N). \quad (74)
$$

Because of the cascading process, which involves the dynamics of velocity increments at different scales, this relationship cannot hold true. However, provided the cascade is generated locally by the nonlinear interaction of velocity increments at neighboring scales, one may expect that the two scale probability distributions or the conditional probability distribution

$$
p(v_1, r_1 | v_2, r_2) = \frac{f^2(v_1, r_1; v_2, r_2)}{f^1(v_2, r_2)} \quad (75)
$$

contains the most important information on the N-scale probability distribution (73).

Markovian Properties

One may wonder whether the knowledge of the conditional probability distribution suffices to reconstruct the N-scale probability distribution (73) in the form

$$
f^N(v_1, r_1; \ldots; v_N, r_N) \\
= p(v_1, r_1 | v_2, r_2) \ldots p(v_{N-1}, r_{N-1} | v_N, r_N) \cdot f(v_N, r_N). \quad (76)
$$

In this case, the probability distribution f^N defines a Markov chain.

The question, whether Markovian properties in scale exist for fully developed turbulence, has been pursued in several ways. First of all, a necessary condition for the existence of Markovian properties is the validity of the Chapman–Kolmogorov equation:

$$
p(v_1, r_1 | v_3, r_3) = \int dv_2 \, p(v_1, r_1 | v_2, r_2) p(v_2, r_2 | v_3, r_3). \quad (77)
$$

This approach has been pursued in [24]. Second, one can validate the Markovian properties by a direct inspection of the conditional probability distribution (75). Defining the conditional probability distribution

$$
p(v_1, r_1 | v_2, r_2; v_3, r_3) = \frac{f^3(v_1, r_1; v_2, r_2; v_3, r_3)}{f^2(v_2, r_2; v_3, r_3)}, \quad (78)
$$

the Markovian property can be assessed by comparing the conditional probability distributions

$$
p(v_1, r_1 | v_2, r_2; v_3, r_3) = p(v_1, r_1 | v_2, r_2). \quad (79)
$$

Summarizing the outcomes of the experimental investigations one can state that the Markovian property can be empirically validated provided the differences of the scales $r_1 - r_2, r_2 - r_3$ are not too small. This statement can be made even more precise. As has been shown in [51], the Markovian property breaks down provided the scale differences are smaller than the Taylor microscale [see Eq. (49)]. This finding attributes a statistical definition to the Taylor length scale. Because of the memory effect we have now called this length Markov–Einstein coherence length L_{mar} [51].

Estimation of the Conditional Probability Distribution

Markov processes are defined through their conditional probability distributions. If one considers the statistics on scales larger compared to the Markov–Einstein length, one may perform the limit $L_{mar} \to 0$ and consider the process to be continuous in scale r, the conditional probability distribution obeys a Fokker–Planck equation of the form

$$
-\frac{\partial}{\partial r} p(v, r | v_0 r_0) \\
= \left[-\frac{\partial}{\partial v} D^1(v, r) + \frac{\partial^2}{\partial v^2} D^2(v, r) \right] p(v, r | v_0 r_0), \quad (80)
$$

where the statistics are determined by the drift function $D^1(v, r)$ and the diffusion function $D^2(v, r)$. We remind the reader that in the definition of $p(v, r | v_0, r_0)$ we have used $r_0 > r$, which leads to the minus sign in the Fokker–Planck equation.

The drift function and the diffusion function have been determined empirically using methods of data analysis of stochastic processes, for further details see [74]. We also refer the reader to [57].

Path-Integral Representation of the N-Scale Probability Distribution

Since the joint N-point probability distribution can be constructed by the conditional probability function due to the Markovian property one can write down a path integral formula for the turbulent cascade in the form

$$
F[v(r)] = Z^{-1} \exp\left[-\int dr' \frac{\left[-\frac{dv(r')}{dr'} - D^1(v(r'), r') \right]^2}{D^2(v(r'), r')} \right]. \quad (81)
$$

This probability distribution is the analog to the Gibbs distribution describing the statistics of systems in thermodynamic equilibrium.
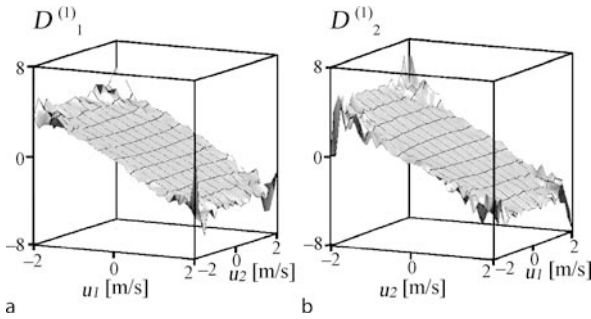
Fluid Dynamics, Turbulence, Figure 10
The u_1 and u_2 dependence of the drift vector for the scale $r = L/4$. **a** The drift coefficient $D_1^{(1)}$, **b** the drift coefficient $D_2^{(1)}$. Note that the vertical axis is rotated for better comparison between **a** and **b**. Both coefficients are linear functions in **u** [79]

Statistics of Longitudinal and Transversal Components

Recently, an analysis of the joint statistics of the longitudinal and transversal components of the velocity increments has been performed [78,79]. The result is a Fokker–Planck equation of the form

$$
- r \frac{\partial}{\partial r} p(\mathbf{u}, r | \mathbf{u}_0, r_0)
$$
$$
= \left\{ -\frac{\partial}{\partial u_i} D_i^{(1)}(\mathbf{u}, r) + \frac{\partial^2}{\partial u_i \partial u_j} D_{ij}^{(2)}(\mathbf{u}, t) \right\}
$$
$$
\cdot p(\mathbf{u}, r | \mathbf{u}_0, r_0) . \quad (82)
$$

Here, $\mathbf{u} = (u_1, u_2)$ denotes the increment vector of the longitudinal and transversal components, respectively. As mentioned above and worked out in more detail in another contribution to this Encyclopedia of Complexity and System Science [25] both drift and diffusion terms $D^{(1)}$, $D^{(2)}$ can be estimated directly from given data. Typ-

ical results are shown in Figs. 10 and 11. The drift terms turn out to be linear,

$$
D_i^{(1)}(\mathbf{u}, r) = d_i(r) u_i . \quad (83)
$$

Furthermore, the diffusion matrix can be approximated by low-order polynomials

$$
D_{ij}^{(2)}(\mathbf{u}, r) = d_{ij}(r) + \sum_k d_{ij;k} u_k + \sum_{kl} d_{ij;kl}(r) u_k u_l . \quad (84)
$$

Knowing the drift and diffusion coefficient it is possible to solve the Fokker–Planck equation numerically, which can be taken as a selfconsistent verification of the estimation procedure. In Fig. 12 it is shown that the numerical solution of the Fokker–Planck equation reproduces the probability distribution directly obtained from the data quite well.

It is quite interesting to notice that the resulting moment equations reproduce Karman's equation, which is a relation between the longitudinal and transversal velocity increments

$$
\langle u_2(r)^2 \rangle = \frac{1}{2r} \frac{d}{dr} r^2 \langle u_1(r)^2 \rangle . \quad (85)
$$

Karman's equation is a direct consequence of the incompressibility of the fluid motion. It can be interpreted as a low-order Taylor expansion of $\langle u_2(r)^2 \rangle = \langle u_1(\frac{3}{2}r)^2 \rangle$ which leads to the proposition that the complex structures of longitudinal and transversal velocity increments mainly differ in a different scale parametrization of the cascade, $r \to \frac{3}{2}r$.

Using these findings the different scaling exponents found for longitudinal and transversal increments can be explained. In Fig. 13 the longitudinal and transversal
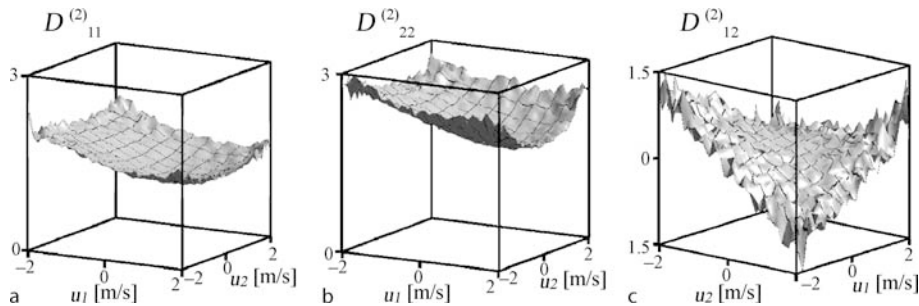
Fluid Dynamics, Turbulence, Figure 11
The u_1 and u_2-dependence of the diffusion matrix for the scale $r = L/4$. **a** The coefficient $D_{11}^{(2)}$, **b** the coefficient $D_{22}^{(2)}$. It can be seen that the diagonal coefficients are not constant but have a parabolic form, which is more pronounced for the $D_{22}^{(2)}$ coefficient (multiplicative noise). Both coefficients are symmetric under reflection with respect to $u_2 \to -u_2$, but not for $u_1 \to -u_1$. **c** The saddle-formed off-diagonal coefficient $D_{12}^{(2)}$ [79]
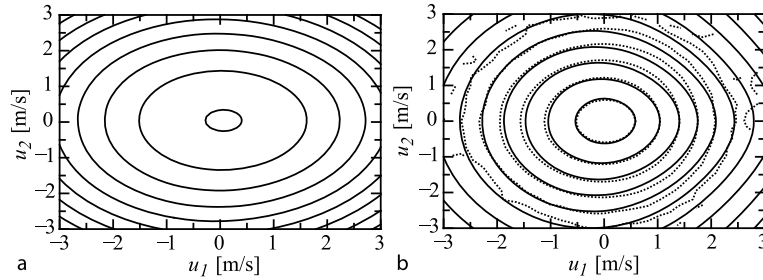
Fluid Dynamics, Turbulence, Figure 12

Solution of the Fokker–Planck equation. **a** Contour plot of the initial condition in logarithmic scale. The simulation starts at the integral length $r = L$ with a Gaussian distribution fitted to the data. **b** The contour plots in logarithm scale of the simulated probability distribution on the scale $r = 2\lambda$. The distance between the contour lines is chosen in logarithmic scale and corresponds to a factor 10. *Dashed lines* indicate the probability distribution calculated directly from data, the *full lines* are the simulation ones. The simulation reproduces well the properties of the data [79]
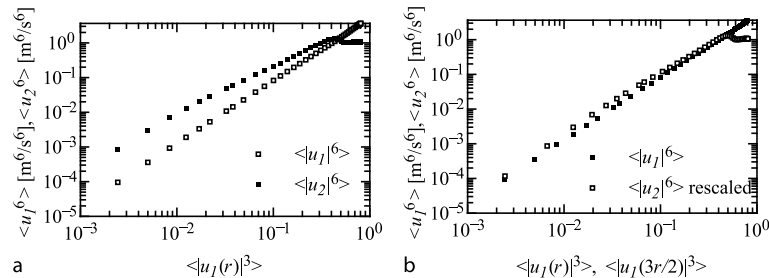
Fluid Dynamics, Turbulence, Figure 13

The sixth-order longitudinal (*black squares*) and transversal (*white squares*) structure function in **a** ESS representation and **b** ESST representation [79]

structure function are shown as a function of the third-order structure function, which should reveal the scaling behavior, too, as the third-order structure function is expected to depend linearly on r. (This presentation is called ESS – extended self-similarity [7]). Most interestingly, the difference in the scaling exponents vanishes if the 3/2 rescaling is applied, see Fig. 13b. The result indicates that proper scaling behavior is not detected and can only be valid approximately.

The Markov analysis of turbulent velocity fields yields a closed phenomenological description of the spatial complexity of turbulence. We note that a similar multiscale analysis has been performed for a turbulent passive scalar field [85].

Lagrangian Fluid Dynamics

Because of experimental progress in the detection of the Lagrangian path of passive tracer particles [46,63] interest in the formulation of stochastic processes describing Lagrangian motion in turbulent fluids has been renewed.

One of the first descriptions of such a process is due to Obukhov [67], who suggested to model the Lagrangian acceleration as a white noise process. He formulated the corresponding evolution equation for the probability distribution $f(\mathbf{x}, \mathbf{u}, t)$ specifying the Lagrangian path in form of a Fokker–Planck equation. The results, which can be derived from the Fokker–Planck equation indicate scaling behavior for the Lagrangian velocity increments and the distance traveled by the particle during time t:

$$\begin{aligned}\langle(\mathbf{u}(t) - \mathbf{u}(0))^2\rangle &= c\langle\epsilon\rangle t\,,\\ \langle(\mathbf{x}(t) - \mathbf{x}(0))^2\rangle &= d(\langle\epsilon\rangle t)^3\,.\end{aligned} \tag{86}$$

This scaling behavior is inferred from dimensional considerations along the theory of Kolmogorov (K41). Therefore, it is not a surprise that the experimentally observed probability distributions deviate from the Gaussian shape required by the Oboukhov model [63]. A modification of Oboukhov's model has been suggested in [21]. Recently, Lagrangian particle statistics has been modeled on the basis of a simple vortex model [87], reproducing qualitatively

the intermittent characteristics of Lagrangian velocity increments [63].

The investigations of the Lagrangian statistics of turbulent flows have benefited a lot from the investigation of passive tracers in disordered flows. Especially the treatment of the Kraichnan model has led to considerable insights into the statistics of particles in flows. We refer the reader to the review article [19].

Future Directions

Although a fundamental understanding is still lacking turbulence research has always led to new concepts and scientific ideas, which substantially have influenced the development of modern science. This will especially hold true for the future, where one may expect a major breakthrough due to combined efforts of experimental, numerical, and analytical work. The following points for future direction of research are only a rather subjective listing of the authors.

Fine Structure of Turbulence: A question which should be investigated experimentally in more detail is whether and how the fine structure is influenced by the mechanism of the generation of turbulence. This will also involve turbulent flows close to walls or in pipes. From the experimental side high precision measurements are required.

Geometrical and Topological Aspects of Turbulence: There is a variational formulation of ideal hydrodynamics, emphasized by V.I. Arnold [5]. We expect that research based on topological and geometrical reasoning will yield further important insights into the spatio-temporal organization of fluid flow.

Statistical Properties of the Lagrangian Map: Topological and geometrical aspects are intimately related with the properties of the Lagrangian map. A further analysis of the trajectories of several Lagrangian particles is currently under consideration. A highly interesting question is whether the stochastic processes of the particle motion in the inertial range can be assessed.

Further Reading

For further reading we suggest the monographs and proceedings [5,13,14,26,33,61,62,73,80].

Acknowledgments

Bibliography

1. Anselmet F, Gagne Y, Hopfinger EJ, Antonia RA (1984) High order structure functions in turbulent shear flows. J Fluid Mech 140:63
2. Antonia RA, Ould-Rouis M, Zhu Y, Anselmet F (1997) Fourth-order moments of longitudinal- and transverse-velocity structure functions. Europhys Lett 37:85
3. Aref H (1983) Integrable, chaotic, and turbulent vortex motion in two-dimensional flows. Ann Rev Fluid Mech 15:345
4. Aref H (2007) Vortices and polynomials. Fluid Dyn Res 39:5
5. Arnold VI, Khesin BA (1999) Topological methods in hydrodynamics. Springer, Berlin
6. Aronson D, Löfdahl LL (1993) The plane wake of a cylinder. Measurements and inferences on the turbulence modelling. Phys Fluids A 5:1433
7. Benzi R, Ciliberto S, Baudet C, Chavarria GR, Tripiccione R (1993) Extended self-similarity in the dissipation range of fully-developed turbulence. Europhys Lett 24:275
8. Castaing B, Gagne Y, Hopfinger EJ (1990) Velocity probability density functions of high Reynolds number turbulence. Phys D 46:177
9. Data provided by Chabaud B, Chanal O, CNRS Gernoble, France
10. Chevillard L, Roux SG, Leveque E, Mordant N, Pinton J-F, Arneodo A (2006) Phys D 218:77
11. Chorin AJ, Marsden GE (2000) A mathematical introduction to fluid mechanics. Springer, Berlin
12. Cuypers Y, Maurel A, Petitjeans P (2003) Vortex burst as a source of turbulence. Phys Rev Lett 91:194502
13. Darrigol O (2005) Worlds of Flow. Oxford University Press, Oxford
14. Davidson PA (2004) Turbulence. Oxford University Press, Oxford
15. Davoudi J, Tabar MRR (1999) Theoretical Model for Kramers-Moyal's description of Turbulence Cascade. Phys Rev Lett 82:1680
16. Doering CR, Gibbon JD (1995) Applied Analysis of the Navier-Stokes Equations. Cambridge University Press, Cambridge
17. Eule S, Friedrich R (2005) A note on a random driven Burgers equation. Phys Lett A 351:238
18. Eyink GL, Sreenivasan KR (2006) Onsager and the theory of hydrodynamic turbulence. Rev Mod Phys 78:87
19. Falkovich G, Gawedzki K, Vergassola M (2001) Particles and fields in fluid turbulence. Rev Mod Phys 73:797
20. Foias C, Rosa R, Manley O, Temam R (2001) Navier–Stokes Equation and Turbulence. Cambridge University Press, Cambridge
21. Friedrich R (2003) Statistics of Lagrangian velocities in turbulent flows. Phys Rev Lett 90:084501
22. Friedrich R, Peinke J (1997) Description of a turbulent cascade by a Fokker–Planck equation. Phys Rev Lett 78:863
23. Friedrich R, Peinke J (1997) Statistical properties of a turbulent cascade. Phys D 102:147
24. Friedrich R, Zeller J, Peinke J (1998) A Note in Three Point Statistics of Velocity Increments in Turbulence. Europhys Lett 41:153
25. Friedrich R, Peinke J, Tabar MRR (2008) Importance of Fluctuations: Complexity in the View of Stochastic Processes within this issue. Springer, Berlin
26. Frisch U (1995) Turbulence. The legacy of Kolmogorov. Cambridge University Press, Cambridge

27. Fursikov AV (1999) The closure problem of the Friedmann–Keller infinite chain of moment equations, corresponding to the Navier–Stokes system. In: Gyr A, Kinzelbach W, Tsinober A (eds) Fundamental problematic issues in turbulence. Birkhäuser, Basel

28. Galanti B, Gibbon JD, Heritage M (1997) Vorticity alignment results for the three-dimensional Euler and Navier–Stokes equations. Nonlinearity 10:1675

29. Gilbert AD (1993) A cascade interpretation of Lundgrens stretched spiral vortex model for turbulent fine structure. Phys Fluids A5:2831

30. Grafke T, Homann H, Dreher J, Grauer R (2008) Numerical simulations of possible finite time singularities in the incompressible Euler equations: comparison of numerical methods. Phys D 237:1932

31. Grassi V, Leo R a, Soliani G, Tempesta P (2000) Vortices and invariant surfaces generated by symmetries for the 3D Navier–Stokes equations. Physica A 286:79

32. Grauer R, Marliani C, Germaschewski K (1998) Adaptive mesh refinement for singular solutions of the incompressible Euler equations. Phys Rev Lett 84:4850

33. Gyr A, Kinzelbach W, Tsinober A (1999) Fundamental problematic issues in turbulence. Birkhäuser, Basel

34. Haken H (1983) Synergetics, An Introduction. Springer, Berlin

35. Haken H (1987) Advanced Synergetics. Springer, Berlin

36. Hamlington PE, Schumacher J, Dahm WJA (2008) Local and nonlocal strain rate fields and vorticity alignment in turbulent flows. Phys Rev E 77:026303

37. Hatakeyama N, Kambe T (1997) Statistical laws of random strained vortices in turbulence. Phys Rev Lett 79:1257

38. Hopf E (1957) Statistical hydromechanics and functional calculus. J Rat Mech Anal 1:87

39. Jimenez J, Wray AA, Saffman PG, Rogallo RS (1993) The structure of intense vorticity in isotropic turbulence. J Fluid Mech 255:65

40. Jovanovic J (2004) The statistical dynamics of turbulence. Springer, Berlin

41. Kirchhoff G (1883) Vorlesungen über mathematische Physik, vol 1, 3rd edn. Teubner, Leipzig

42. Kolmogorov AN (1941) Dissipation of energy in locally isotropic turbulence. Dokl Akad Nauk SSSR 32:19

43. Kolmogorov AN (1941) The local structure of turbulence in incompressible viscous fluid for very large Reynold's numbers. Dokl Akad Nauk SSSR 30:301

44. Kolmogorov AN (1962) A refinement of previous hypotheses concerning the local structure of turbulence in a viscous incompressible fluid at high Reynolds number. J Fluid Mech 13:82

45. L'vov V, Procaccia I Hydrodynamic Turbulence: a 19th Century Problem with a Challenge for the 21st Centurs. arXiv:chao-dyn 96006015

46. La Porta A, Voth G, Crawford AM, Alexander J, Bodenschatz E (2000) Fluid particle acceleration in fully developped turbulence. Nature 409:1017; La Porta A, Voth G, Moisy F, Bodenschatz E (2001) Using cavitation to measure statistics of low-pressure events in large Reynolds-number turbulence. Phys Fluids 12:1485

47. Landau L, Lifshitz EM (1981) Lehrbuch der Theoretischen Physik, Bd. 6, Hydrodynamik. Akademie, Berlin

48. Langner M, Peinke J, Rauh A () Langevin analysis with applica-

tion to a Rayleigh–Benard convection experiment. Int J Nonlinear Dyn Syst Chaos (in press)

49. Lesieur M (1997) Turbulence in fluids. Kluwer Academic Publishers, Dordrecht

50. Lesieur M, Yaglom A, David F (2000) New trends in turbulence, LesHouches Summer School. Springer, Berlin

51. Lück S, Renner C, Peinke J, Friedrich R (2006) The Markov–Einstein coherence length – a new meaning for the Taylor length in turbulence. Phys Lett A 359:335

52. Lundgren TS (1969) Distribution functions in the statistical theory of turbulence. Phys Fluids 10:969

53. Lundgren TS (1982) Strained spiral vortex model for turbulent fine structure. Phys Fluids 25:2193

54. Lundgren TS (1993) A small-scale turbulence model. Phys Fluids A5:1472

55. Manneville P (2004) Instabilities, chaos and turbulence. Imperial College Press, London

56. Marchioro C, Pulvirenti M (1984) Vortex methods in two-dimensional fluid dynamics. Lecture Notes in Physics. Springer, Berlin

57. Marcq P, Naert AA (2001) Langevin equation for turbulent velocity increments. Phys Fluids 13:2590

58. Martin PC, Siggia ED, Rose HA (1973) Statistical dynamics of classical systems. Phys Rev A8:423

59. McComb WD (1990) The physics of fluid turbulence. Clarendon Press, Oxford

60. Moffat HK, Kida S, Ohkitani K (1994) Stretched vortices—the sinews of turbulence. J Fluid Mech 259:231

61. Monin AS, Yaglom AM (1971) Statistical Fluid Mechanics: Mechanics of Turbulence, vol 1. MIT Press, Cambridge, MA

62. Monin AS, Yaglom AM (1975) Statistical Fluid Mechanics: Mechanics of Turbulence, vol 2. MIT Press, Cambridge, MA

63. Mordant N, Metz P, Michel O, Pinton J-F (2001) Measurement of Lagrangian velocity in fully developed turbulence. Phys Rev Lett 87:214501

64. Newton PK (2001) The N-Vortex problem. Springer, New York

65. Oberlack M (2000) Symmetrie, Invarianz und Selbstähnlichkeit in der Turbulenz. Shaker, Aachen

66. Oberlack M, Busse FH (2002) Theories of turbulence. Springer, Wien

67. Obukhov AM (1959) Description of turbulence in terms of Lagrangian variables. Adv Geophys 6:113

68. Onsager L (1949) Statistical hydrodynamics. Suppl Nuovo Cimento 6:279

69. Ott S, Mann J (2000) An experimental investigation of the relative diffusion of particle pairs in three-dimensional turbulent flow. J Fluid Mech 402:207

70. Ottino JM (1989) The Kinematics of Mixing: Stretching, Chaos, and Transport. Cambridge University Press, Cambridge

71. Piquet J (1999) Turbulent flows, models and physics. Springer, Berlin

72. Pope SB (1985) Lagrangian PDF methods for turbulent flows. Annu Rev Fluid Mech 26:23; Pope SB (1994) Pdf methods for turbulent reactive flows. Prog Energy Combust Sci 11:119

73. Pope SB (2000) Turbulent flows. Cambridge University Press, Cambridge

74. Renner C, Peinke J, Friedrich R (2001) Experimental indications for Markov properties of small scale turbulence. J Fluid Mech 433:383

75. Renner C, Peinke J, Friedrich R, Chanal O, Chabaud B (2002) Universality of small scale turbulence. Phys Rev Lett 89:124502

76. Reynolds O (1883) Philos Trans R Soc Lond 174:935

77. Richardson LF (1926) Atmospheric diffusion shown on a distance-neighbour graph. Proc R Soc Lond A 110:709

78. Siefert M, Peinke J (2004) Different cascade speeds for longitudinal and transverse velocity increments of small-scale turbulence. Phys Rev E 70:015302R

79. Siefert M, Peinke J (2006) Joint multi-scale statistics of longitudinal and transversal increments in small-scale wake turbulence. J Turbul 7:1

80. Sreenivasan KR, Antonia RA (1997) The phenomenology of small-scale turbulence. Annu Rev Fluid Mech 29:435–472

81. Sturman R, Ottino JM, Wiggins S (2006) The mathematical foundations of mixing. In: Cambridge Monographs of Applied and Computational Mathematics, No 22

82. Taylor GI (1921) Diffusion by continuous movement. Proc Lond Math Soc Ser 20(2):196

83. Temam R (2007) Infinite dimensional dynamical systems in mechanics and physics. Springer, Heidelberg

84. Townsend AA (1951) On the fine structure of turbulence. Proc R Soc Lond A 208:534

85. Tutkun M, Mydlarski L (2004) Markovian properties of passive scalar increments in grid-generated turbulence. New J Phys 6:49

86. Ulinich FR, Ya Lyubimov B (1969) Statistical theory of turbulence of an incompressible fluid at large Reynolds numbers. Zh Exper Teor Fiz 55:951

87. Wilczek M, Jenko F, Friedrich R Lagrangian particle statistics in turbulent flows from a simple vortex model. Phys Rev E (to appear)

88. Calculation performed by Wilczek M (Münster)

89. Wyld HW (1961) Formulation of the theory of turbulence in incompressible fluids. Ann Phys 14:143

90. http://www.claymath.org/millenium/

Food Webs

JENNIFER A. DUNNE[1,2]

[1] Santa Fe Institute, Santa Fe, USA
[2] Pacific Ecoinformatics and Computational Ecology Lab, Berkeley, USA

Article Outline

Glossary

Connectance(C) The proportion of possible links in a food web that actually occur. There are many algorithms for calculating connectance. The simplest and most widely used algorithm, sometimes referred to as "directed connectance," is links per species^2 (L/S^2), where S^2 represents all possible directed feeding interactions among S species, and L is the total number of actual feeding links. Connectance ranges from ~ 0.03 to 0.3 in food webs, with a mean of ~ 0.10 to 0.15.

Consumer-resource interactions A generic way of referring to a wide variety of feeding interactions, such as predator-prey, herbivore-plant or parasite-host interactions. Similarly, "consumer" refers generically to anything that consumes or preys on something else, and "resource" refers to anything that is consumed or preyed upon. Many taxa are both consumers and resources within a particular food web.

Food web The network of feeding interactions among diverse co-occurring species in a particular habitat.

Trophic species (S) Defined within the context of a particular food web, a trophic species is comprised of a set of taxa that share the same set of consumers and resources. A particular trophic species is represented by a single node in the network, and that node is topologically distinct from all other nodes. "Trophic species" is a convention introduced to minimize bias due to uneven resolution in food web data and to focus analysis and modeling on functionally distinct network components. S is used to denote the number of trophic species in a food web. The terms "trophic species," "species," and "taxa" will be used somewhat interchangeably throughout this article to refer to nodes in a food web. "Original species" will be used specifically to denote the taxa found in the original dataset, prior to trophic species aggregation.

Definition of the Subject

Food webs refer to the networks of feeding ("trophic") interactions among species that co-occur within particular habitats. Research on food webs is one of the few subdisciplines within ecology that seeks to quantify and analyze direct and indirect interactions among diverse species, rather than focusing on particular types of taxa. Food webs ideally represent whole communities including plants, bacteria, fungi, invertebrates and vertebrates. Feeding links represent transfers of biomass and encompass

a variety of trophic strategies including detritivory, herbivory, predation, cannibalism and parasitism. At the base of every food web are one or more types of autotrophs, organisms such as plants or chemoautotrophic bacteria, which produce complex organic compounds from an external energy source (e. g., light) and simple inorganic carbon molecules (e. g., CO_2). Food webs also have a detrital component—non-living particulate organic material that comes from the body tissues of organisms. Feeding-mediated transfers of organic material, which ultimately trace back to autotrophs or detritus via food chains of varying lengths, provide the energy, organic carbon and nutrients necessary to fuel metabolism in all other organisms, referred to as heterotrophs.

While food webs have been a topic of interest in ecology for many decades, some aspects of contemporary food web research fall within the scope of the broader cross-disciplinary research agenda focused on complex, "real-world" networks, both biotic and abiotic [2,83,101]. Using the language of graph theory and the framework of network analysis, species are represented by vertices (nodes) and feeding links are represented by edges (links) between vertices. As with any other network, the structure and dynamics of food webs can be quantified, analyzed and modeled. Links in food webs are generally considered directed, since biomass flows from a resource species to a consumer species (A → B). However, trophic links are sometimes treated as undirected, since any given trophic interaction alters the population and biomass dynamics of both the consumer and resource species (A ↔ B). The types of questions explored in food web research range from "Do food webs from different habitats display universal topological characteristics, and how does their structure compare to that of other types of networks?" to "What factors promote different aspects of stability of complex food webs and their components given internal dynamics and external perturbations?" Two fundamental measures used to characterize food webs are S, the number of species or nodes in a web, and C, connectance—the proportion of possible feeding links that are actually realized in a web ($C = L/S^2$, where L is the number of observed directed feeding links, and S^2 is the number of possible directed feeding interactions among S taxa).

This article focuses on research that falls at the intersection of food webs and complex networks, with an emphasis on network structure augmented by a brief discussion of dynamics. This is a subset of a wide variety of ecological research that has been conducted on feeding interactions and food webs. Refer to the "Books and Reviews" in the bibliography for more information about a broader range of research related to food webs.

Introduction: Food Web Concepts and Data

The concept of food chains (e. g., grass is eaten by grasshoppers which are eaten by mice which are eaten by owls; A →B → C → D) goes back at least several hundred years, as evidenced by two terrestrial and aquatic food chains briefly described by Carl Linnaeus in 1749 [42]. The earliest description of a food web may be the mostly detrital-based feeding interactions observed by Charles Darwin in 1832 on the island of St. Paul, which had only two bird species (Darwin 1939, as reported by Egerton [42]):

> By the side of many of these *[tern]* nests a small flying-fish was placed; which, I suppose, had been brought by the male bird for its partner … quickly a large and active crab (Craspus), which inhabits the crevices of the rock, stole the fish from the side of the nest, as soon as we had disturbed the birds. Not a single plant, not even a lichen, grows on this island; yet it is inhabited by several insects and spiders. The following list completes, I believe, the terrestrial fauna: a species of Feronia and an acarus, which must have come here as parasites on the birds; a small brown moth, belonging to a genus that feeds on feathers; a staphylinus (Quedius) and a woodlouse from beneath the dung; and lastly, numerous spiders, which I suppose prey on these small attendants on, and scavengers of the waterfowl.

The earliest known diagrams of generalized food chains and food webs appeared in the late 1800s, and diagrams of specific food webs, began appearing in the early 1900s, for example the network of insect predators and parasites on cotton-feeding weevils ("the boll weevil complex," [87]). By the late 1920s, diagrams and descriptions of terrestrial and marine food webs were becoming more common (e. g., Fig. 1 from [103], see also [48,104]). Charles Elton introduced the terms "food chain" and "food cycle" in his classic early textbook, *Animal Ecology* [43]. By the time Eugene Odum published a later classic textbook, *Fundamentals of Ecology* [84], the term "food web" was starting to replace "food cycle."

From the 1920s to the 1980s, dozens of system-specific food web diagrams and descriptions were published, as well as some webs that were more stylized (e. g., [60]) and that quantified link flows or species biomasses. In 1977, Joel Cohen published the first comparative studies of empirical food web network structure using up to 30 food webs collected from the literature [23,24]. To standardize the data, he transformed the diagrams and descriptions of webs in the literature into binary matrices with m rows and n columns [24]. Each column is headed by the number of
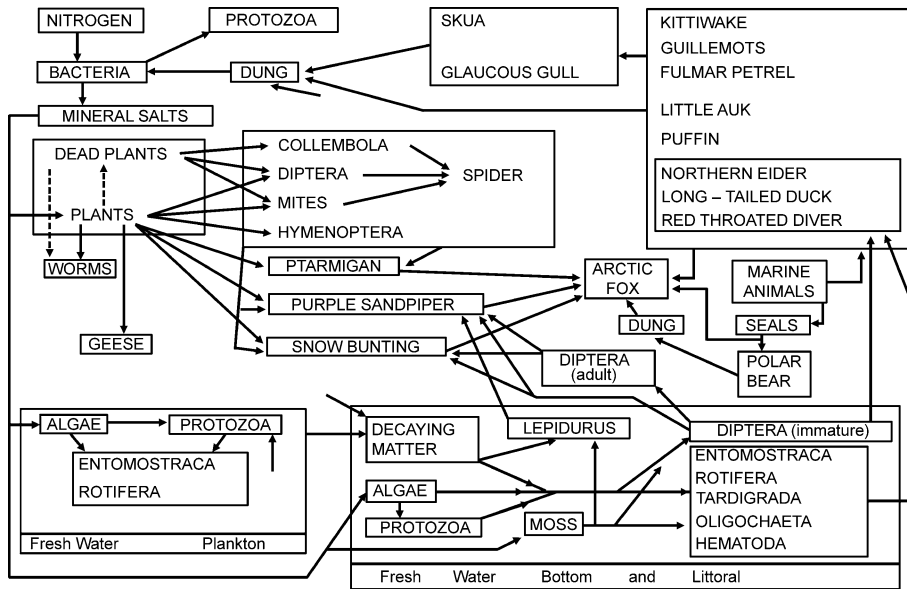
Food Webs, Figure 1
A diagram of a terrestrial Arctic food web, with a focus on nitrogen cycling, for Bear Island, published in 1923 [103]

one of the consumer taxa in a particular web, and each row is headed by the number of one of the resource taxa for that web. If w_{ij} represents the entry in the ith row and the jth column, it equals 1 if consumer j eats resource i or 0 if j does not eat i. This matrix-based representation of data is still often used, particularly in a full S by S format (where S is the number of taxa in the web), but for larger datasets a compressed two- or three-column notation for observed links is more efficient (Fig. 2).

By the mid-1980s, those 30 initial webs had expanded into a 113-web catalog [30] which included webs mostly culled from the literature, dating back to the 1923 Bear Island food web ([103], Fig. 1). However, it was apparent that there were many problems with the data. Most of the 113 food webs had very low diversity compared to the biodiversity known to be present in ecosystems, with a range of only 5 to 48 species in the original datasets and 3 to 48 trophic species. This low diversity was largely due to very uneven resolution and inclusion of taxa in most of these webs. The webs were put together in many different ways and for various purposes that did not include comparative, quantitative research. Many types of organisms were aggregated, underrepresented, or missing altogether, and in a few cases animal taxa had no food chains connecting them to basal species. In addition, cannibalistic links were purged when the webs were compiled into the 113-web catalog. To many ecologists, these food webs looked like little more than idiosyncratic cartoons of much richer and more complex species interactions found in

natural systems, and they appeared to be an extremely unsound foundation on which to build understanding and theory [86,92].

Another catalog of "small" webs emerged in the late 1980s, a set of 60 insect-dominated webs with 2 to 87 original species (mean = 22) and 2 to 54 trophic species (mean = 12) [102]. Unlike the 113-web catalog, these webs are highly taxonomically resolved, mostly to the species level. However, they are still small due to their focus, in most cases, on insect interactions in ephemeral microhabitats such as phytotelmata (i. e., plant-held aquatic systems such as water in tree holes or pitcher plants) and singular detrital sources (e. g., dung paddies, rotting logs, animal carcasses). Thus, while the 113-web catalog presented food webs for communities at fairly broad temporal and spatial scales, but with low and uneven resolution, the 60-web catalog presented highly resolved but very small spatial and temporal slices of broader communities. These two very different catalogs were compiled into ECOWeB, the "Ecologists Co-Operative Web Bank," a machine readable database of food webs that was made available by Joel Cohen in 1989 [26]. The two catalogs, both separately and together as ECOWeB, were used in many studies of regularities in food web network structure, as discussed in the next Sect. "Early Food Web Structure Research".

A new level of detail, resolution and comprehensiveness in whole-community food web characterization was presented in two seminal papers in 1991. Gary Polis [92] published an enormous array of data for taxa found in the

a — Partial matrix format

	3	4	5	6
1	1	1	0	0
2	1	1	0	1
3	1	0	1	0
4	0	0	0	1
5	0	0	1	1
6	0	0	1	1

b — Full matrix format

	1	2	3	4	5	6
1	0	0	1	1	0	0
2	0	0	1	1	0	1
3	0	0	1	0	1	0
4	0	0	0	0	0	1
5	0	0	0	0	1	1
6	0	0	0	0	1	1

c — Two-column format

3	1
3	2
3	3
4	1
4	2
5	3
5	5
5	6
6	2
6	4
6	5
6	6

d — Three-column format

3	1	3
4	1	2
5	3	
5	5	6
6	2	
6	4	6

Food Webs, Figure 2
Examples of formats for standardized notation of binary food web data. A hypothetical web with 6 taxa and 12 links is used. *Numbers 1–6 correspond to the different taxa.* **a** Partial matrix format: the 1s or 0s inside the matrix denote the presence or absence of a feeding link between a consumer (whose numbers 3–6 head columns) and a resource (whose numbers 1–6 head rows); **b** Full matrix format: similar to **a**, but all 6 taxa are listed at the heads of columns and rows; **c** Two-column format: a consumer's number appears in the *first column*, and one of its resource's numbers appears in the *second column*; **d** Three-column format: similar to **c**, but where there is a third number, the second and third numbers refer to a range of resource taxa. In this hypothetical web, taxa numbers 1 and 2 are basal taxa (i. e., taxa that do not feed on other taxa—autotrophs or detritus), and taxa numbers 3, 5, and 6 have cannibalistic links to themselves

Coachella Valley desert (California). Over two decades, he collected taxonomic and trophic information on at least 174 vascular plant species, 138 vertebrate species, 55 spider species, thousands of insect species including parasitoids, and unknown numbers of microorganisms, acari, and nematodes. He did not compile a complete food web including all of that information, but instead reported a number of detailed subwebs (e. g., a soil web, a scorpion-focused web, a carnivore web, etc.), each of which was more diverse than most of the ECOWeB webs. On the basis of the subwebs and a simplified, aggregated 30-taxa web of the whole community, he concluded that "… *most catalogued webs are oversimplified caricatures of actual communities … [they are] grossly incomplete representations of communities in terms of both diversity and trophic connections.*"

At about the same time, Neo Martinez [63] published a detailed food web for Little Rock Lake (Wisconsin) that he compiled explicitly to test food web theory and patterns (see Sect. "Early Food Web Structure Research"). By piecing together diversity and trophic information from multiple investigators actively studying various types of taxa in the lake, he was able to put together a relatively complete and highly resolved food web of 182 taxa, most identified to the genus, species, or ontogentic life-stage level, including fishes, copepods, cladocera, rotifers, diptera and other insects, mollusks, worms, porifera, algae, and cyanobacteria. In later publications, Martinez modified the original dataset slightly into one with 181 taxa. The 181 taxa web aggregates into a 92 trophic-species web, with nearly 1000 links among the taxa (Fig. 3). This dataset, and the accompanying analysis, set a new standard for food web empiricism and analysis. It still stands as the best whole-community food web compiled, in terms of even, detailed, comprehensive resolution.

Since 2000, the use of the ECOWeB database for comparative analysis and modeling has mostly given way to a focus on a smaller set of more recently published food webs [10,37,39,99,110]. These webs, available through www.foodwebs.org or from individual researchers, are compiled for particular, broad-scale habitats such as St. Mark's Estuary [22], Little Rock Lake [63], the island of St. Martin [46], and the Northeast U.S. Marine Shelf [61]. Most of the food webs used in contemporary comparative research are still problematic—while they generally are more diverse and/or evenly resolved than the earlier webs, most could still be resolved more highly and evenly. Among several issues, organisms such as parasites are usually left out (but see [51,59,67,74]), microorganisms are either missing or highly aggregated, and there is still a tendency to resolve vertebrates more highly than lower level organisms. An important part of future food web research is the compilation of more inclusive, evenly resolved, and well-defined datasets. Meanwhile, the careful selection and justification of datasets to analyze is an important part of current research that all too often is ignored.
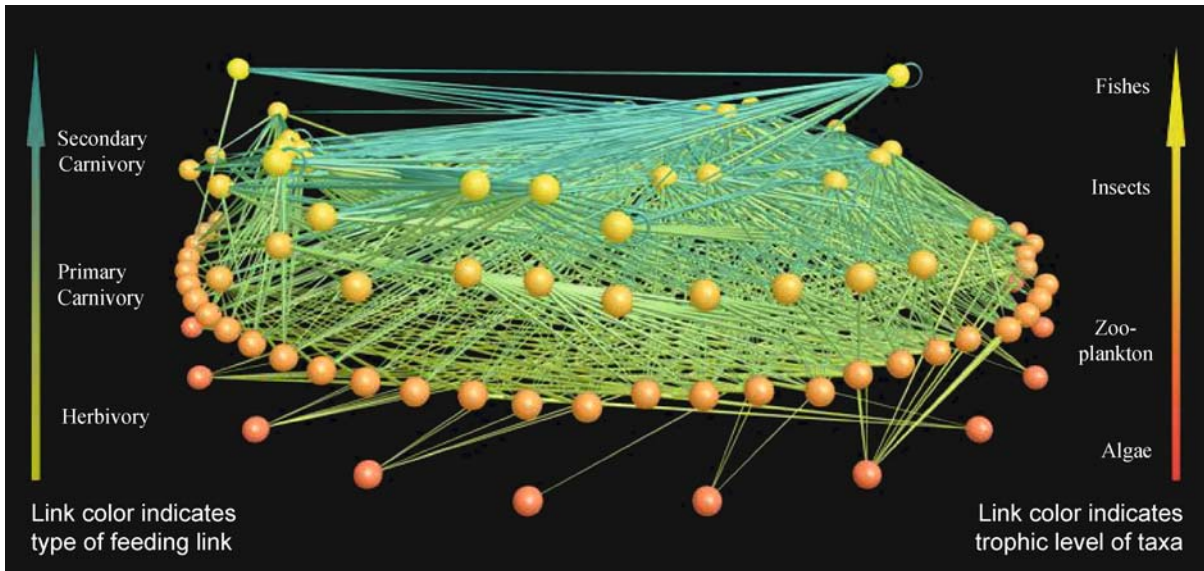
Food Webs, Figure 3
Food web of Little Rock Lake, Wisconsin [63]. 997 feeding links among 92 trophic species are shown. Image produced with Food-Web3D, written by R.J. Williams, available at the Pacific Ecoinformatics and Computational Ecology Lab (www.foodwebs.org)

How exactly are food web data collected? In general, the approach is to compile as complete a species list as possible for a site, and then to determine the diets of each species present at that site. However, researchers have taken a number of different approaches to compiling food webs. In some cases, researchers base their food webs on observations they make themselves in the field. For example, ecologists in New Zealand have characterized the structure of stream food webs by taking samples from particular patches in the streams, identifying the species present in those samples, taking several individuals of each species present, and identifying their diets through gut-content analysis [106]. In other cases, researchers compile food web data by consulting with experts and conducting literature searches. For example, Martinez [63] compiled the Little Rock Lake (WI) food web by drawing on the expertise of more than a dozen biologists who were specialists on various types of taxa and who had been working at Little Rock Lake for many years. Combinations of these two approaches can also come into play—for example, a researcher might compile a relatively complete species list through field-based observations and sampling, and then assign trophic habits to those taxa through a combination of observation, consulting with experts, and searching the literature and online databases.

It is important to note that most of the webs used for comparative research can be considered "cumulative" webs. Contemporary food web data range from time- and space-averaged or "cumulative" (e. g., [63]) to more finely resolved in time (e. g., seasonal webs—[6]) and/or space (e. g., patch-scale webs—[106]; microhabitat webs—[94]). The generally implicit assumption underlying cumulative food web data is that the set of species in question co-exist within a habitat and individuals of those species have the opportunity over some span of time and space to interact directly. To the degree possible, such webs document who eats whom among all species within a macrohabitat, such as a lake or meadow, over multiple seasons or years, including interactions that are low frequency or represent a small proportion of consumption. Such cumulative webs are used widely for comparative research to look at whether there are regularities in food web structure across habitat (see Sect. "Food Webs Compared to Other Networks" and Sect. "Models of Food Web Structure"). More narrowly defined webs at finer scales of time or space, or that utilize strict evidence standards (e. g., recording links only through gut content sampling), have been useful for characterizing how such constraints influence perceived structure within habitats [105,106], but are not used as much to look for cross-system regularities in trophic network structure.

Early Food Web Structure Research

The earliest comparative studies of food web structure were published by Joel Cohen in 1977. Using data from the first 30-web catalog, one study focused on the ratio of predators to prey in food webs [23], and the other in-

vestigated whether food webs could be represented by single dimension interval graphs [24], a topic which continues to be of interest today (see Sect. "Food Webs Compared to Other Networks"). In both cases, he found regularities—(1) a ratio of prey to predators of $\sim 3/4$ regardless of the size of the web, and (2) most of the webs are interval, such that all species in a food web can be placed in a fixed order on a line such that each predator's set of prey forms a single contiguous segment of that line. The prey-predator ratio paper proved to be the first salvo in a quickly growing set of papers that suggested that a variety of food web properties were "scale-invariant." In its strong sense, scale invariance means that certain properties have constant values as the size (S) of food webs change. In its weak sense, scale-invariance refers to properties not changing systematically with changing S. Other scale-invariant patterns identified include constant proportions of top species (Top, species with no predators), intermediate species (Int, species with both predators and prey), and basal species (Bas, species with no prey), collectively called "species scaling laws" [12], and constant proportions of T-I, I-B, T-B, and I-I links between T, I, and B species, collectively called "link scaling laws" [27]. Other general properties of food webs were thought to include: food chains are short [31,43,50,89]; cycling/looping is rare (e. g., A $\rightarrow$ B $\rightarrow$ C $\rightarrow$ A; [28]); compartments, or subwebs with many internal links that have few links to other subwebs, are rare [91]; omnivory, or feeding at more than one trophic level, is uncommon [90]; and webs tend to be interval, with instances of intervality decreasing as S increases [24,29,116]. Most of these patterns were reported for the 113-web catalog [31], and some of the regularities were also documented in a subset the 60 insect-dominated webs [102].

Another, related prominent line of early comparative food web research was inspired by Bob May's work from the early 1970s showing that simple, abstract communities of interacting species will tend to transition sharply from local stability to instability as the complexity of the system increases—in particular as the number of species (S), the connectance (C) *or* the average interaction strength (i) increase beyond critical values [69,70]. He formalized this as a criterion that ecological communities near equilibrium will tend to be stable if $i(SC)^{1/2} < 1$. This mathematical analysis flew in the face of the intuition of many ecologists (e. g., [44,50,62,84]) who felt that increased complexity (in terms of greater numbers of species and links between them) in ecosystems gives rise to stability.

May's criterion and the general question of how diversity is maintained in communities provided a framework within which to analyze some readily accessible em-

pirical data, namely the numbers of links and species in food webs. Assuming that average interaction strength (i) is constant, May's criterion suggests that communities can be stable given increasing diversity (S) as long as connectance (C) decreases. This can be empirically demonstrated using food web data in three similar ways, by showing that 1) C hyperbolically declines as S increases, so that the product SC remains constant, 2) the ratio of links to species (L/S), also referred to as link or linkage density, remains constant as S increases, or 3) L plotted as a function of S on a log-log graph, producing a power-law relation of the form $L = \alpha S^{\beta}$, displays an exponent of $\beta = 1$ (the slope of the regression) indicating a linear relationship between L and S. These relationships were demonstrated across food webs in a variety of studies (see detailed review in [36]), culminating with support from the 113-web catalog and the 60 insect-dominated web catalog. Cohen and colleagues identified the "link-species scaling law" of $L/S \approx 2$ using the 113 web catalog (i. e., there are two links per species on average in any given food web, regardless of its size) [28,30], and SC was reported as "roughly independent of species number" in a subset of the 60 insect-dominated webs [102].

However, these early conclusions about patterns of food web structure began to crumble with the advent of improved data and new analysis methods that focused on the issues of species aggregation, sampling effort, and sampling consistency [36]. Even before there was access to improved data, Tom Schoener [93] set the stage for critiques of the conventional paradigm in his Ecological Society of America MacArthur Award lecture, in which he explored the ramifications of a simple conceptual model based on notions of "generality" (what Schoener referred to as "generalization") and "vulnerability." He adopted the basic notion underlying the "link-species scaling law": that how many different taxa something can eat is constrained, which results in the number of resource taxa per consumer taxon (generality) holding relatively steady with increasing S. However, he further hypothesized that the ability of resource taxa to defend again consumers is also constrained, such that the number of consumer taxa per resource taxon (vulnerability) should increase with increasing S. A major consequence of this conceptual model is that total links per species (L/S, which includes links to resources and consumers) and most other food web properties should display scale-dependence, not scale-invariance. A statistical reanalysis of a subset of the 113-web catalog supported this contention as well as the basic assumptions of his conceptual model about generality and vulnerability.

Shortly thereafter, more comprehensive, detailed datasets, like the ones for Coachella Valley [92] and Little

Rock Lake [63], began to appear in the literature. These and other new datasets provided direct empirical counterpoints to many of the prevailing notions about food webs: their connectance and links per species were much higher than expected from the "link-species scaling law," food chains could be quite long, omnivory and cannibalism and looping could be quite frequent, etc. In addition, analyzes such as the one by Martinez [63], in which he systematically aggregated the Little Rock Lake food web taxa and links to generate small webs that looked like the earlier data, demonstrated that "*most published food web patterns appear to be artifacts of poorly resolved data.*" Comparative studies incorporating newly available data further undermined the whole notion of "scale invariance" of most properties, particularly L/S (e. g., [65,66]).

For many researchers, the array of issues brought to light by the improved data and more sophisticated analyzes was enough for them to turn their back on structural food web research. A few hardy researchers sought to build new theory on top of the improved data. "Constant connectance" was suggested as an alternative hypothesis to constant L/S (the "link-species scaling law"), based on a comparative analysis of the relationship of L to S across a subset of available food webs including Little Rock Lake [64]. The mathematical difference between constant C and constant L/S can be simply stated using a log-log graph of links as a function of species (Fig. 4). If a power law exists of the form $L = \alpha S^{\beta}$, in the case of the link-species scaling law $\beta = 1$, which means that $L = \alpha S$, $L/S = \alpha$, indicating constant L/S. In the case of constant connectance, $\beta = 2$ and thus $L = \alpha S^2$, $L/S^2 = \alpha$, indicating constant C (L/S^2). Constant connectance means that L/S increases as a fixed proportion of S. One ecological interpretation of constant connectance is that consumers are likely to exploit an approximately constant fraction of available prey species, so as diversity increases, links per species increases [108].

Given the $L = \alpha S^{\beta}$ framework, $\beta = 2$ was reported for a set of 15 webs derived from an English pond [108], and $\beta = 1.9$ for a set of 50 Adirondack lakes [65], suggesting connectance may be constant across webs within a habitat or type of habitat. Across habitats, the picture is less clear. While $\beta = 2$ was reported for a small subset of the 5 most "credible" food webs then available from different habitats [64], several analyzes of both the old ECOWeB data and the more reliable newer data suggest that the exponent lies somewhere between 1 and 2, suggesting that C declines non-linearly with S (Fig. 4, [27,30,36,64,79,93]). For example, Schoener's reanalysis of the 113-web catalog suggested that $\beta = 1.5$, indicating that $L^{2/3}$ is proportional to S. A recent analysis of 19 recent trophic-species food
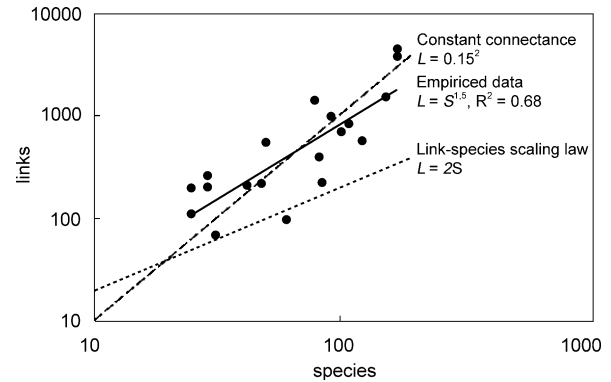
Food Webs, Figure 4
The relationship of links to species for 19 trophic-species food webs from a variety of habitats (*black circles*). The *solid line* shows the log-log regression for the empirical data, the *dashed line* shows the prediction for constant connectance, and the *dotted line* shows the prediction for the link-species scaling law (reproduced from [36], Fig. 1)

webs with S of 25 to 172 also reported $\beta = 1.5$, with much scatter in the data (Fig. 4).

A recent analysis has provided a possible mechanistic basis for the observed constrained variation in C (~ 0.03 to 0.3 in cumulative community webs) as well as the scaling of C with S implied by β intermediate between 1 and 2 [10]. A simple diet breadth model based on optimal foraging theory predicts both of these patterns across food webs as an emergent consequence of individual foraging behavior of consumers. In particular, a contingency model of optimal foraging is used to predict mean diet breadth for S animal species in a food web, based on three parameters for an individual of species j: (1) net energy gain from consumption of an individual of species i, (2) the encounter rate of individuals of species i, and (3) the handling time spent attacking an individual of species i. This allows estimation of C for the animal portion of food webs, once data aggregation and cumulative sampling, well-known features of empirical datasets, are taken into account. The model does a good job of predicting values of C observed in empirical food webs and associated patterns of C across food webs.

Food Web Properties

Food webs have been characterized by a variety of properties or metrics, several of which have been mentioned previously (Sect. "Early Food Web Structure Research"). Many of these properties are quantifiable just using the basic network structure ("topology") of feeding interactions. These types of topological properties have been used to

evaluate simple models of food web structure (Sect. "Food Web Properties"). Any number of properties can be calculated on a given network—ecologists tend to focus on properties that are meaningful within the context of ecological research, although other properties such as path length (**Path**) and clustering coefficient (**Cl**) have been borrowed from network research [109]. Examples of several types of food web network structure properties, with common abbreviations and definitions, follow.

Fundamental Properties: These properties characterize very simple, overall attributes of food web network structure.

S: number of nodes in a food web
L: number of links in a food web
L/S: links per species
C, or L/S^2: connectance, or the proportion of possible links that are realized

Types of Taxa: These properties characterize what proportion or percentage of taxa within a food web fall into particular topologically defined roles.

Bas: percentage of basal taxa (taxa without resources)
Int: percentage of intermediate taxa (taxa with both consumers and resources)
Top: percentage of top taxa (taxa with no consumers)
Herb: percentage of herbivores plus detritivores (taxa that feed on autotrophs or detritus)
Can: percentage of cannibals (taxa that feed on their own taxa)
Omn: percentage of omnivores (taxa that feed that feed on taxa at different trophic levels
Loop: percentage of taxa that are in loops, food chains in which a taxon occur twice (e. g., A → B → C → A)

Network Structure: These properties characterize other attributes of network structure, based on how links are distributed among taxa.

TL: trophic level averaged across taxa. Trophic level represents how many steps energy must take to get from an energy source to a taxon. Basal taxa have TL = 1, and obligate herbivores are TL = 2. TL can be calculated using many different algorithms that take into account multiple food chains that can connect higher level organisms to basal taxa (Williams and Martinez 2004).
ChLen: mean food chain length, averaged over all species
ChSD: standard deviation of ChLen
ChNum: log number of food chains

LinkSD: normalized standard deviation of links (# links per taxon)
GenSD: normalized standard deviation of generality (# resources per taxon)
VulSD: normalized standard deviation of vulnerability (# consumers per taxon)
MaxSim: mean across taxa of the maximum trophic similarity of each taxon to other taxa
Ddiet: the level of diet discontinuity—the proportion of triplets of taxa with an irreducible gap in feeding links over the number of possible triplets [19]—a local estimate of intervality
Cl: clustering coefficient (probability that two taxa linked to the same taxon are linked)
Path: characteristic path length, the mean shortest set of links (where links are treated as undirected) between species pairs

The previous properties (most of which are described in [110] and [39]) each provide a single metric that characterizes some aspect of food web structure. There are other properties, such as **Degree Distribution**, which are not single-number properties. "Degree" refers to the number of links that connect to a particular node, and the degree distribution of a network describes (in the format of a function or a graph) the total number of nodes in a network that have a given degree for each level of degree (Subsect. "Degree Distribution"). In food web analysis, **LinkSD**, **GenSD**, and **VulSD** characterize the variability of different aspects of degree distribution. Many food web structure properties are correlated with each other, and vary in predictable ways with *S* and/or *C*. This provides opportunities for topological modeling that are discussed below (Sect. "Models of Food Web Structure").

In addition to these types of metrics based on networks with unweighted links and nodes, it is possible to calculate a variety of metrics for food webs with nodes and/or links that are weighted by measures such as biomass, numerical abundance, frequency, interaction strength, or body size [11,33,67,81]. However, few food web datasets are "enriched" with such quantitative data and it remains to be seen whether such approaches are primarily a tool for richer description of particular ecosystems or whether they can give rise to novel generalities, models and predictions. One potential generality was suggested by a study of interaction strengths in seven soil food webs, where interaction strength reflects the size of the effects of species on each other's dynamics near equilibrium. Interaction strengths appear to be organized such that long loops contain many weak links, a pattern which enhances stability of complex food webs [81].

Food Webs Compared to Other Networks

Small-World Properties

How does the structure of food webs compare to that of other kinds of networks? One common way that various networks have been compared is in terms of whether they are "small-world" networks. Small-world networks are characterized by two of the properties described previously, characteristic path length (**Path**) and clustering coefficient (**Cl**) [109]. Most real-world networks appear to have high clustering, like what is seen on some types of regular spatial lattices (such as a planar triangular lattice, where many of a node's neighbors are neighbors of one another), but have short path lengths, like what is seen on "random graphs" (i.e., networks in which links are distributed randomly). Food webs do display short path lengths that are similar to what is seen in random webs (Table 1, [16,37,78,113]). On average, taxa are about two links from other taxa in a food web ("two degrees of separation"), and path length decreases with increasing connectance [113].

However, clustering tends to be quite low in many food webs, closer to the clustering expected on a random network (Table 1). This relatively low clustering in food webs appears consistent with their small size compared to most other kinds of networks studied, since the ratio of clustering in empirical versus comparable random networks increases linearly with the size of the network (Fig. 5).

Degree Distribution

In addition to small-world properties, many real-world networks appear to display power-law degree distributions [2]. Whereas regular graphs have the same number of links per node, and random graphs display a Poisson degree distribution, many empirical networks, both biotic and abiotic, display a highly skewed power-law ("scale-free") degree distribution, where most nodes have few links and a few nodes have many links. However, some empirical networks display less-skewed distributions such as exponential distributions [4]. Most empirical food webs display exponential or uniform degree distributions, not power-law distributions [16,37], and it has been suggested that normalized degree distributions in food webs follow universal functional forms [16] although there is a quite a bit of scatter when a wide range of data are considered (Fig. 6, [37]). Variable degree distributions, like what is seen in individual food webs, could result from simple mechanisms. For example, exponential and uniform food web degree distributions are generated by a model that combines (1) random immigration to local webs from a randomly linked regional set of taxa, and (2) random extinctions in the local webs [5]. The general lack of power-

Food Webs, Table 1
Topological properties of empirical and random food webs, listed in order of increasing connectance. *Path* refers to characteristic path length, and *Cl* refers to the clustering coefficient. *Path$_r$* and *Cl$_r$* refer to the mean *D* and *Cl* for 100 random webs with the same *S* and *C*. Modified from [37] Table 1

Food Web	S	C(L/S²)	L/S	Path	Path$_r$	Cl	Cl$_r$	Cl/Cl$_r$
Grassland	61	0.026	1.59	3.74	3.63	0.11	0.03	3.7
Scotch Broom	85	0.031	2.62	3.11	2.82	0.12	0.04	3.0
Ythan Estuary 1	124	0.038	4.67	2.34	2.39	0.15	0.04	3.8
Ythan Estuary 2	83	0.057	4.76	2.20	2.19	0.16	0.06	2.7
El Verde Rainforest	155	0.063	9.74	2.20	1.95	0.12	0.07	1.4
Canton Creek	102	0.067	6.83	2.27	2.01	0.02	0.07	0.3
Stony Stream	109	0.070	7.61	2.31	1.96	0.03	0.07	0.4
Chesapeake Bay	31	0.071	2.19	2.65	2.40	0.09	0.09	1.0
St. Marks Seagrass	48	0.096	4.60	2.04	1.94	0.14	0.11	1.3
St. Martin Island	42	0.116	4.88	1.88	1.85	0.14	0.13	1.1
Little Rock Lake	92	0.118	10.84	1.89	1.77	0.25	0.12	2.1
Lake Tahoe	172	0.131	22.59	1.81	1.74	0.14	0.13	1.1
Mirror Lake	172	0.146	25.13	1.76	1.72	0.14	0.15	0.9
Bridge Brook Lake	25	0.171	4.28	1.85	1.68	0.16	0.19	0.8
Coachella Valley	29	0.312	9.03	1.42	1.43	0.43	0.32	1.3
Skipwith Pond	25	0.315	7.88	1.33	1.41	0.33	0.33	1.0
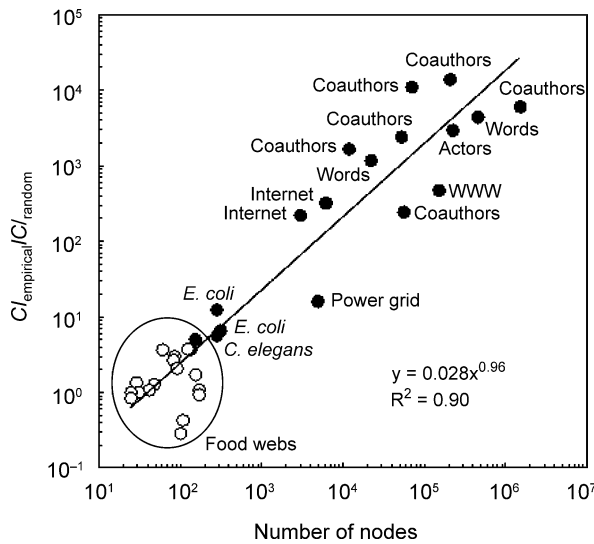
Food Webs, Figure 5
Trends in clustering coefficient across networks. The ratio of clustering in empirical networks ($CI_{empirical}$) to clustering in random networks with the same number of nodes and links (CI_{random}) is shown as a function of the size of the network (number of nodes). Reproduced from [37], Fig. 1

Food Webs, Figure 6
Log-log overlay plot of the cumulative distributions of links per species in 16 food webs. The link data are normalized by the average number of links/species in each web. If the distributions followed a power law, the data would tend to follow a straight line. Instead, they follow a roughly exponential shape. Reproduced from [37], Fig. 3

law degree distributions in food webs may result partly from the small size and large connectance of such networks, which limits the potential for highly skewed distributions. Many of the networks displaying power-law degree distributions are much larger and much more sparsely connected than food webs.

Other Properties

Assortative mixing, or the tendency of nodes with similar degree to be linked to each other, appears to be a pervasive phenomenon in a variety of social networks [82]. However, other kinds of networks, including technological and biological networks, tend to show disassortative mixing, where nodes with high degree tend to link to nodes with low degree. Biological networks, and particularly two food webs examined, show strong disassortativity [82]. Some of this may relate to a finite-size effect in systems like food webs that have limits on how many links are recorded between pairs of nodes. However, in food webs it may also result from the stabilizing effects of having feeding specialists linked to feeding generalists, as has been suggested for plant-animal pollination and frugivory (fruit-eating) networks ([7], Sect. "Ecological Networks").

Another aspect of structure that has been directly compared across several types of networks including food webs are "motifs," defined as "recurring, significant patterns of interconnections" [77]. A variety of networks (transcriptional gene regulation, neuron connectivity, food webs, two types of electronic circuits, the World Wide Web) were scanned for all possible subgraphs that could be constructed out of 3 or 4 nodes (13 and 199 possible subgraphs, respectively). Subgraphs that appeared significantly more often in empirical webs than in their randomized counterparts (i. e., networks with the same number of nodes and links, and the same degree for each node, but with links otherwise randomly distributed) were identified. For the seven food webs examined, there were two "consensus motifs" shared by most of the webs—a three-node food chain, and a four-species diamond where a predator has two prey, which in turn prey on the same species (Fig. 7). The four-node motif was shared by two other types of networks (neuron connectivity, one type of electronic circuit), and nothing shared the three-node chain. The WWW and food web networks appear most dissimilar to other types of networks (and to each other) in terms of significant motifs.

Complex networks can be decomposed into minimum spanning trees (MST). A MST is a simplified version of a network created by removing links to minimize the distance between nodes and some destination. For example, a food web can be turned into MST by adding an "environment" node that all basal taxa link to, tracing the shortest food chain from each species to the environment node, and removing links that do not appear in the shortest chains. Given this algorithm, a MST removes links that
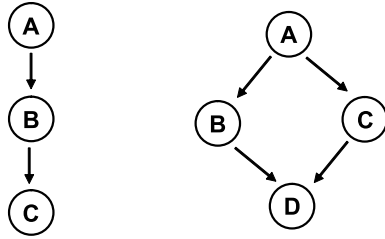
Food Webs, Figure 7
The two 3 or 4-node network motifs found to occur significantly more often than expected in most of seven food webs examined. There is one significant 3-node motif (out of 13 possible motifs), a food chain of the form A eats B eats C. There is one significant 4-node motif (out of 199 possible motifs), a trophic diamond ("bi-parallel") of the form A eats B and C, which both eat D

occur in loops and retains a basic backbone that has a tree-like structure. In a MST, the quantity A_i is defined as the number of nodes in a subtree rooted at node i, and can be regarded as the transportation rate through that node. C_i is defined as the integral of A_i (i. e., the sum of A_i for all nodes rooted at node i) and can be regarded as the transportation cost at node i. These properties can be used to plot C_i versus A_i for each node in a networks, or to plot whole-system C_o versus A_o across multiple networks, to identify whether scaling relationships of the form $C(A) \propto A^n$ are present, indicating self-similarity in the structure of the MST (see [18] for review). In a food web MST, the most efficient configuration is a star, where every species links directly to the environment node, resulting in an exponent of 1, and the least efficient configuration is a single chain, where resources have to pass through each species in a line, resulting in an exponent of 2. It has been suggested that food webs display a universal exponent of 1.13 [18,45], reflecting an invariant functional food web property relating to very efficient resource transportation within an ecosystem. However, analyzes based on a larger set of webs (17 webs versus 7) suggest that exponents for C_i as a function of A_i range from 1.09 to 1.26 and are thus not universal, that the exponents are quite sensitive to small changes in food web structure, and that the observed range of exponent values would be similarly constrained in any network with only 3 levels, as is seen in food web MSTs [15].

Models of Food Web Structure

An important area of research on food webs has been whether their observed structure, which often appears quite complex, can emerge from simple rules or models. As with other kinds of "real-world" networks, models that assign links among nodes randomly, according to fixed probabilities, fail to reproduce the network structure of empirically observed food webs [24,28,110]. Instead, several models that combine stochastic elements with simple link assignment rules have been proposed to generate and predict the network structure of empirical food webs.

The models share a basic formulation [110]. There are two empirically quantifiable parameters: (1) S, the number of trophic species in a food web, and (2) C, the connectance of a food web, defined as links per species squared, L/S^2. Thus, S specifies the number of nodes in a network, and C specifies the number of links in a network with S nodes. Each species is assigned a "niche value" n_i drawn randomly and uniformly from the interval [0,1]. The models differ in the rules used to distribute links among species. The link distribution rules follow in the order the models were introduced in the literature:

Cascade Model (Cohen and Newman [28]): Each species has the fixed probability $P = 2CS/(S-1)$ of consuming species with niche values less than its own. This creates a food web with hierarchical feeding, since it does not allow feeding on taxa with the same niche value (cannibalism) or taxa with higher niche values (looping/cycling). This formulation [110] is a modified version of the original cascade model that allows L/S, equivalent to the CS term in the probability statement above, to vary as a tunable parameter, rather than be fixed as a constant [28].

Niche Model (Williams and Martinez [110], Fig. 8): Each species consumes all species within a segment of the [0,1] interval whose size r_i is calculated using the feeding range width algorithm described below. The r_i's center c_i is set at a random value drawn uniformly from the interval $[r_i/2, n_i]$ or $[r_i/2, 1 - r_i/2]$ if $n_i > 1 - r_i/2$, which places c_i equal to or lower than the niche value n_i and keeps the r_i segment within [0,1]. The c_i rule relaxes the strict feeding hierarchy of the cascade model and allows for the possibility of cannibalism and looping. Also, the r_i rule ensures that species feed on a contiguous range of species, necessarily creating interval graphs (i. e., species can be lined up along a single interval such that all of their resource species are located in contiguous segments along the interval).

Feeding range width algorithm: The value of $r_i = xn_i$, where $0 < x < 1$ is randomly drawn from the probability density function $p(x) = \beta(1-x)^{b-1}$ (the beta distribution), where $\beta = (1/2C) - 1$ to obtain a C close to the desired C.
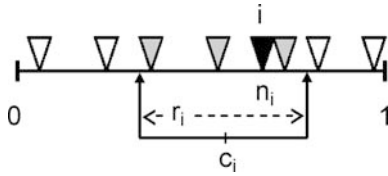
Food Webs, Figure 8
Graphical representation of the niche model: Species i feeds on 4 taxa including itself and one with a higher niche value

Nested-Hierarchy Model (Cattin et al. [19]): Like the niche model, the number of prey items for each species is drawn randomly from a beta distribution that constrains C close to a target value. Once the number of prey items for each species is set, those links are assigned in a multistep process. First, a link is randomly assigned from species i to a species j with a lower n_i. If j is fed on by other species, the next feeding links for i are selected randomly from the pool of prey species fed on by a set of consumer species defined as follows: they share at least one prey species and at least one of them feeds on j. If more links need to be distributed, they are then randomly assigned to species without predators and with niche values $< n_i$, and finally to those with niche value $\geq n_i$. These rules were chosen to relax the contiguity rule of the niche model and to allow for trophic habit overlap among taxa in a manner which the authors suggest evokes phylogenetic constraints.

Generalized Cascade Model (Stouffer et al. [99]): Species i feeds on species j if $n_j \leq n_i$ with a probability drawn from the interval [0,1] using the beta or an exponential distribution. This model combines the beta distribution introduced in the niche model with the hierarchical, non-contiguous feeding of the cascade model.

These models have been evaluated with respect to their relative fit to empirical data in a variety of ways. In a series of six papers published from 1985 to 1990 with the common title "A stochastic theory of community food webs," the cascade model was proposed as a means of explaining "*the phenomenology of observed food web structure, using a minimum of hypotheses*" [31]. This was not the first simple model proposed for generating food web structure [25,88,89,116], but it was the most well-developed model. Cohen and colleagues also examined several model variations, most of which performed poorly. While the cascade model appeared to generate structures that qualitatively fit general patterns in the data from the 113-web catalog, subsequent statistical analyzes suggested that the fit between the model and that early data was

poor [93,96,97]. Once improved data began to emerge, it became clear that some of the basic assumptions built in to the cascade model, such as no looping and minimal overlap and no clustering of feeding habits across taxa, are violated by common features of multi-species interactions.

The niche model was introduced in 2000, along with a new approach to analysis: numerical simulations to compare statistically the ability of the niche model, the cascade model, and one type of random network model to fit empirical food web data [110]. Because of stochastic variation in how species and links are distributed in any particular model web, analysis begins with the generation of hundreds to thousands of model webs with the same S and similar C as an empirical food web of interest. Model webs that fall within 3% of the target C are retained. Model-generated webs occasionally contain species with no links to other species, or species that are trophically identical. Either those webs are thrown out, or those species are eliminated and replaced, until every model web has no disconnected or identical species. Also, each model web must contain at least one basal species. These requirements ensure that model webs can be sensibly comparable to empirical trophic-species webs.

Once a set of model webs are generated with the same S and C as an empirical web, model means and standard deviations are calculated for each food web property of interest, which can then be compared to empirical values. Raw error, the difference between the value of an empirical property and a model mean for that property, is normalized by dividing it by the standard deviation of the property's simulated distribution. This approach allows assessment not only of whether a model over- or under-estimates empirical properties as indicated by the raw error, but also to what degree a model's mean deviates from the empirical value. Normalized errors within ± 2 are considered to indicate a good fit between the model prediction and the empirical value. This approach has also been used to analyze network motifs [77] (Subsect. "Other Properties").

The initial niche model analyzes examined seven more recent, diverse food webs ($S = 24$ to 92) and up to 12 network structure properties for each web [110]. The random model (links are distributed randomly among nodes) performs poorly, with an average normalized error (ANE) of 27.1 (SD $= 202$). The cascade model performs better, with ANE of -3.0 (SD $= 14.1$). The niche model performs an order of magnitude better than the cascade model, with ANE of 0.22 (SD $= 1.8$). Only the niche model falls within ± 2 ANE considered to show a good fit to the data. Not surprisingly, there is variability in how all three models fit different food webs and properties. For example, the niche

model generally overestimates food-chain length. Specific mismatches are generally attributable either to limitations of the models or biases in the data [110]. A separate test of the niche and cascade models with three marine food webs, a type of habitat not included in the original analysis, obtained similar results [39]. These analyzes demonstrate that the structure of food webs is far from random and that simple link distribution rules can yield apparently complex network structure, similar to that observed in empirical data. In addition, the analyzes suggest that food webs from a variety of habitats share a fundamentally similar network structure, and that the structure is scale-dependent in predictable ways with S and C.

The nested-hierarchy model [19] and generalized cascade model [99], variants of the niche model, do not appear to improve on the niche model, and in fact may be worse at representing several aspects of empirical network structure. Although the nested-hierarchy model breaks the intervality of the niche model and uses a complicated-sounding set of link distribution rules to try to mimic phylogenetic constraints on trophic structure, it "*generates webs characterized by the same universal distributions of numbers of prey, predators, and links*" as the niche model [99]. Both the niche and nested-hierarchy models have a beta distribution at their core. The beta distribution is reasonably approximated by an exponential distribution for $C < 0.12$ [99], and thus reproduces the exponential degree distributions observed in many empirical webs, particularly those with average or less-than-average C [37]. The generalized cascade model was proposed as a simplified model that would return the same distributions of taxa and links as the niche and nested-hierarchy models. It is defined using only two criteria: (1) taxa form a totally ordered set—this is fulfilled by the arrangement of taxa along a single "niche" interval or dimension, and (2) each species has an exponentially decaying probability of preying on a given fraction of species with lower niche values [99].

Although the generalized cascade model does capture a central tendency of successful food web models, only some food web properties are derivable from link distributions (e. g., Top, Bas, Can, VulSD, GenSD, Clus). There are a variety of food web structure properties of interest that are not derivable from degree distributions (e. g., Loop, Omn, Herb, TL, food-chain statistics, intervality statistics). The accurate representation of these types of properties may depend on additional factors, for example the contiguous feeding ranges specified by the niche model but absent from the cascade, nested-hierarchy, and generalized cascade models. While it is known that empirical food webs are not interval, until recently it was not clear how non-interval they are. Intervality is a brittle property that

is broken by a single gap in a single feeding range (i. e., a single missing link in a food web), and trying to arrange species in a food web into their most interval ordering is a computationally challenging problem. A more robust measure of intervality in food webs has been developed, in conjunction with the use of simulated annealing to estimate the most interval ordering of empirical food webs [100]. This analysis suggests that complex food webs "*do exhibit a strong bias toward contiguity of prey, that is, toward intervality*" when compared to several alternative "null" models, including the generalized cascade model. Thus, the intervality assumption of the niche model, initially critiqued as a flaw of the model [19], helps to produce a better fit to empirical data than the non-interval alternate models.

Structural Robustness of Food Webs

A series of papers have examined the response of a variety of networks including the Internet and WWW web pages [1] and metabolic and protein networks [52,53] to the simulated loss of nodes. In each case, the networks, all of which display highly skewed power-law degree distributions, appear very sensitive to the targeted loss of highly-connected nodes but relatively robust to random loss of nodes. When highly-connected nodes are removed from scale-free networks, the average path length tends to increase rapidly, and the networks quickly fragment into isolated clusters. In essence, paths of information flow in highly skewed networks are easily disrupted by loss of nodes that are directly connected to an unusually large number of other nodes. In contrast, random networks with much less skewed Poisson degree distributions display similar responses to targeted loss of highly-connected nodes versus random node loss [101].

Within ecology, species deletions on small ($S < 14$) hypothetical food web networks as well as a subset of the 113-web catalog have been used to examine the reliability of network flow, or the probability that sources (producers) are connected to sinks (consumers) in food webs [54]. The structure of the empirical webs appears to conform to reliable flow patterns identified using the hypothetical webs, but that result is based on low diversity, poorly resolved data. The use of more highly resolved data with node knock-out algorithms to simulate the loss of species allows assessment of potential secondary extinctions in complex empirical food webs. Secondary extinctions result when the removal of taxa results in one or more consumers losing all of their resource taxa. Even though most food webs do not have power-law degree distributions, they show similar patterns of robustness to other networks: re-

moval of highly-connected species results in much higher rates of secondary extinctions than random loss of species ([38,39,95], Fig. 9). In addition, loss of high-degree species results in more rapid fragmentation of the webs [95]. Protecting basal taxa from primary removal increases the robustness of the web (i. e., fewer secondary extinctions occur) ([38], Fig. 9). While removing species with few links generally results in few secondary extinctions, in a quarter of the food webs examined, removing low-degree species results in secondary extinctions comparable to or greater than what is seen with removal of high-degree species [38]. This tends to occur in webs with relatively high C.

Beyond differential impacts of various sequences of species loss in food webs, food web 'structural robustness' can be defined as the fraction of primary species loss that induces some level of species loss (primary + secondary extinctions) for a particular trophic-species web. Analysis of R_{50} (i. e., what proportion of species have to be removed to achieve $\geq$ 50% total species loss) across multiple food webs shows that robustness increases approximately logarithmically with increasing connectance ([38,39], Fig. 9, 10). In essence, from a topological perspective food webs with more densely interconnected taxa are better protected from species loss, since it takes greater species loss for consumers to lose all of their resources.

It is also potentially important from a conservation perspective to identify particular species likely to result in the greatest number of secondary extinctions through their loss. The loss of a particular highly-connected species may or may not result in secondary extinctions. One
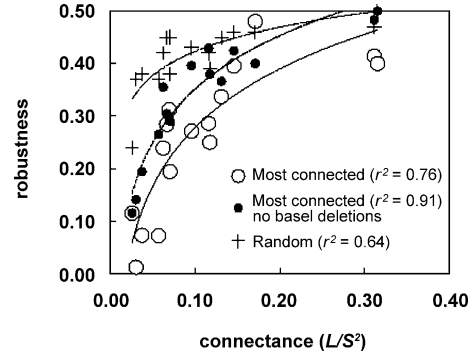
Food Webs, Figure 10
The proportion of primary species removals required to induce a total loss (primary removals plus secondary extinctions) of 50% of the species in each of 16 food webs ("robustness," see the shorter red line of Fig. 9 for visual representation) as a function of the connectance of each web. Logarithmic fits to the three data sets are shown, with a *solid line* for the most connected deletion order, a *long dashed line* for the most connected with basal species preserved deletion order, and a *short dashed* line for random deletion order. The maximum possible y value is 0.50. The equations for the fits are: $y = 0.162 \ln(x) + 0.651$ for most connected species removals, $y = 0.148 \ln(x) + 0.691$ for most connected species removals with basal species preserved, and $y = 0.067 \ln(x) + 0.571$ for random species removals. Reproduced from [38], Fig. 2

way to identify critical taxa is to reduce the topological structure of empirical food webs into linear pathways that define the essential chains of energy delivery in each web. A particular species can be said to "dominate" other
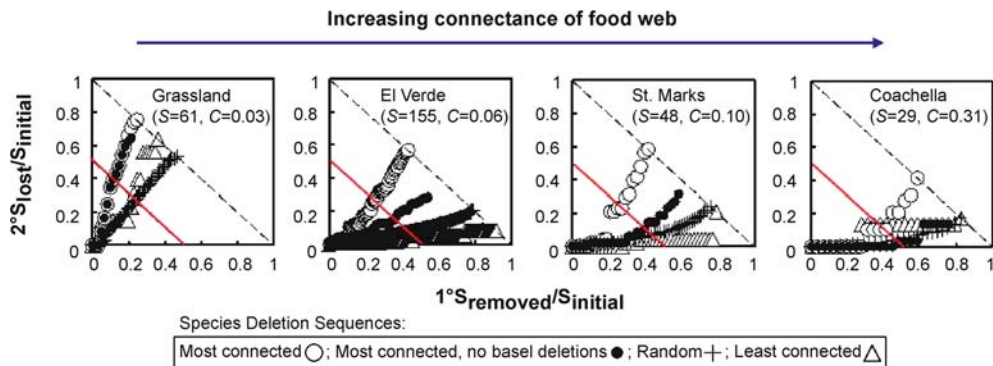
Food Webs, Figure 9
Secondary extinctions resulting from primary species loss in 4 food webs ordered by increasing connectance (*C*). The *y-axis* shows the cumulative secondary extinctions as a fraction of initial S, and the *x-axis* shows the primary removals of species as a fraction of initial *S*. 95% error bars for the random removals fall within the size of the symbols and are not shown. For the most connected (*circles*), least connected (*triangles*), and random removal (*plus symbols*) sequences, the data series end at the *black diagonal dashed line*, where primary removals plus secondary extinctions equal *S* and the web disappears. For the most connected species removals with basal species preserved (*black dots*), the data points end when only basal species remain. The shorter *red diagonal lines* show the points at which 50% of species are lost through combined primary removals and secondary extinctions ("robustness" or R_{50})

species if it passes energy to them along a chain in the dominator tree. The higher the number of species that a particular species dominates, the greater the secondary extinctions that may result from its removal [3]. This approach has the advantage of going beyond assessment of direct interactions to include indirect interactions.

As in food webs, the order of pollinator loss has an effect on potential plant extinction patterns in plant-pollinator networks [75] (see Sect. "Ecological Networks"). Loss of plant diversity associated with targeted removal of highly-connected pollinators is not as extreme as comparable secondary extinctions in food webs, which may be due to pollinator redundancy and the nested topology of those networks.

While the order in which species go locally extinct clearly affects the potential for secondary extinctions in ecosystems, the focus on high-degree, random, or even dominator species does not provide insight on ecologically plausible species loss scenarios, whether the focus is on human perturbations or natural dynamics. The issue of what realistic natural extinction sequences might look like has been explored using a set of pelagic-focused food webs for 50 Adirondack lakes [49] with up to 75 species [98]. The geographic nestedness of species composition across the lakes is used to derive an ecologically plausible extinction sequence scenario, with the most restricted taxa the most likely to go extinct. This sequence is corroborated by the pH tolerances of the species. Species removal simulations show that the food webs are highly robust in terms of secondary extinctions to the "realistic" extinction order and highly sensitive to the reverse order. This suggests that nested geographical distribution patterns coupled with local food web interaction patterns appear to buffer effects of likely species losses. This highlights important aspects of community organization that may help to minimize biodiversity loss in the face of a naturally changing environment. However, anthropogenic disturbances may disrupt the inherent buffering of how taxa are organized geographically and trophically, reducing the robustness of ecosystems.

Food Web Dynamics

Analysis of the topology of food webs has proven very useful for exploring basic patterns and generalities of "who eats whom" in ecosystems. This approach seeks to identify "*the most universal, high-level, persistent elements of organization*" [35] in trophic networks, and to leverage understanding of such organization for thinking about ecosystem robustness. However, food webs are inherently dynamical systems, since feeding interactions in-

volve biomass flows among species whose "stocks" can be characterized by numbers of individuals and/or aggregate population biomass. All of these stocks and flows change through time in response to direct and indirect trophic and other types of interactions. Determining the interplay among network structure, network dynamics, and various aspects of stability such as persistence, robustness, and resilience in complex "real-world" networks is one of the great current challenges in network research [101]. It is particularly important in the study of ecosystems, since they face a variety of anthropogenic perturbations such as climate change, habitat loss, and invasions, and since humans depend on them for a variety of "ecosystem services" such as supply of clean water and pollination of crops [34].

Because it is nearly impossible to compile detailed, long-term empirical data for dynamics of more than two interacting species, most research on species interaction dynamics relies on analytical or simulation modeling. Most modeling studies of trophic dynamics have focused narrowly on predator-prey or parasite-host interactions. However, as the previous sections should make clear, in natural ecosystems such interaction dyads are embedded in diverse, complex networks, where many additional taxa and their direct and indirect interactions can play important roles for the stability of focal species as well as the stability or persistence of the broader community. Moving beyond the two-species population dynamics modeling paradigm, there is a strong tradition of research that looks at interactions among 3–8 species, exploring dynamics and simple variations in structure in slightly more complex systems (see reviews in [40,55]). However, these interaction modules still present a drastic simplification of the diversity and structure of natural ecosystems. Other dynamical approaches have focused on higher diversity model systems [69], but ignore network structure in order to conduct analytically tractable analyzes.

Researchers are increasingly integrating dynamics with complex food web structure in modeling studies that move beyond small modules. The Lotka–Volterra cascade model [20,21,32] was an early incarnation of this type of integration. As its name suggests, the Lotka–Volterra cascade model runs classic L-V dynamics, including a nonsaturating linear functional response, on sets of species interactions structured according to the cascade model [28]. The cascade model was also used to generate the structural framework for a dynamical food web model with a linear functional response [58] used to study the effects of prey-switching on ecosystem stability. Improving on aspects of biological realism of both dynamics and structure, a bioenergetic dynamical model with nonlinear functional responses [119] was used in conjunction with em-

pirically-defined food web structure among 29 species to simulate the biomass dynamics of a marine fisheries food web [117,118]. This type of nonlinear bioenergetic dynamical modeling approach has been integrated with niche model network structure and used to study more complex networks [13,14,68,112]. A variety of types of dynamics are observed in these non-linear models, including equilibrium, limit cycle, and chaotic dynamics, which may or may not be persistent over short or long time scales (Fig. 11). Other approaches model ecological and evolutionary dynamics to assemble species into networks, rather than imposing a particular structure on them. These models, which typically employ an enormous amount of parameters, are evaluated as to whether they generate plausible persistence, diversity, and network structure (see review by [72]). All of these approaches are generally used to examine stability, characterized in a diversity of ways, in ecosystems with complex structure and dynamics [71,85].

While it is basically impossible to empirically validate models of integrated structure and dynamics for complex ecological networks, in some situations it is possible to draw interesting connections between models and data at more aggregated levels. This provides opportunities to move beyond the merely heuristic role that such models generally play. For example, nonlinear bioenergetic models of population dynamics parametrized by biological rates allometrically scaled to populations' average body masses have been run on various types of model food web structures [14]. This approach has allowed the comparison of trends in two different measures of food web stability, and how they relate to consumer-resource body-size ratios and to initial network structure. One measure of stability is the fraction of original species that display persistent dynamics, i. e., what fraction of species do not go extinct in the model when it is run over many time steps ("species persistence"). Another measure of stability is how variable the densities of all of the persistent species are ("population stability")—greater variability across all the species indicates decreased stability in terms of population dynamics.

Brose and colleagues [14] ran the model using different hypothetical consumer-resource body-size ratios that range from 10^{-2} (consumers are 100 times smaller than their resources) to 10^5 (consumers are 100,000 times larger than their resources) (Fig. 12). Species persistence increases dramatically with increasing body-size ratios, until inflection points are reached at which persistence shifts to high levels ($\sim$ 0.80) of persistence (Fig. 12a). However, population stability decreases with increasing body-size ratios until inflection points are reached that show the lowest stability, and then increases again beyond those
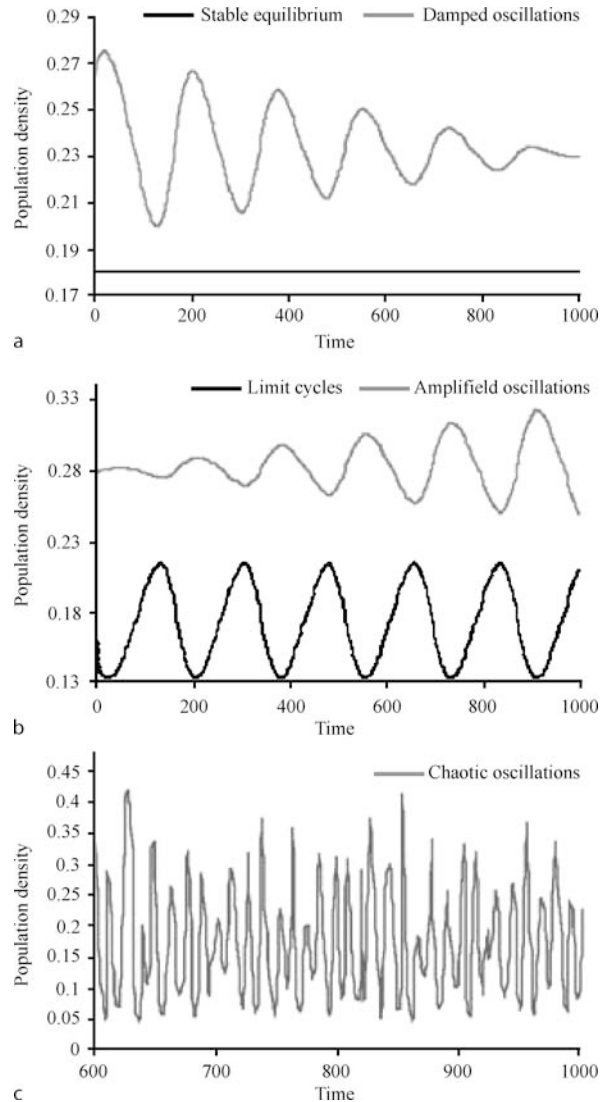
Food Webs, Figure 11
5 different types of population dynamics shown as time series of population density (from [40], Fig. 10.1). The types of dynamics shown include a stable equilibrium, damped oscillations, limit cycles, amplified oscillations, and chaotic oscillations

points (Fig. 12b). In both cases, the inflection points correspond to empirically observed consumer-resource body-size ratios, both for webs parametrized to represent invertebrate dominated webs, and for webs parametrized to represent ectotherm vertebrate dominated webs. Thus, across thousands of observations, invertebrate consumer-resource body size ratios are $\sim 10^1$, and ectotherm vertebrate consumer-resource body size ratios are $\sim 10^2$, which correspond to the model's inflection points for species persistence and population stability (Fig. 12). It is interesting
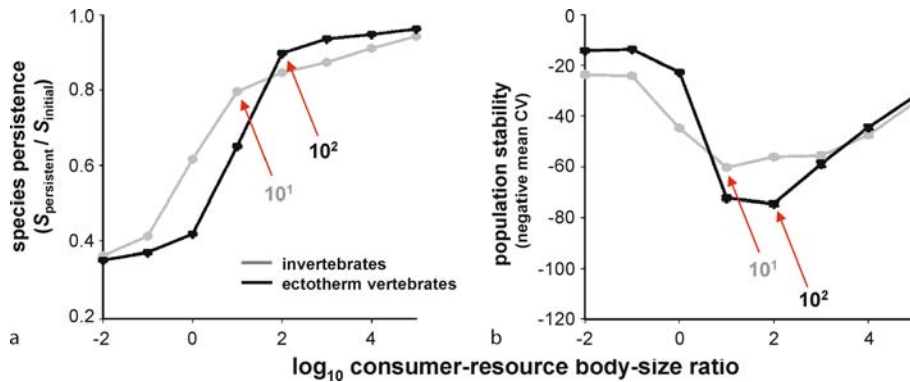
Food Webs, Figure 12
a shows the fraction of species that display persistent dynamics as a function of consumer-resource body-size ratios for model food webs parametrized for invertebrates (*gray line*) and ectotherm vertebrates (*black line*). The inflection points for shifts to high-persistence dynamics are indicated by *red arrows* for both curves, and those inflection points correspond to empirically observed consumer-resource body size ratios for invertebrate dominated webs (10^1—consumers are on average 10 times larger than their resources) and ectotherm vertebrate dominated webs (10^2—consumers are on average 100 times larger than their resources). **b** shows results for population stability, the mean of how variable species population biomasses are in persistent webs. In this case, the inflection points for shifts to low population stability are indicated by red arrows, and those inflection points also correspond to the empirically observed body-size ratios for consumers and resources. Figure adapted from [14]

to note that high species persistence is coupled to low population stability—i. e., an aspect of increased stability of the whole system (species persistence) is linked to an aspect of decreased stability of components of that system (population stability). It is also interesting to note that in this formulation, using initial cascade versus niche model structure had little impact on species persistence or population stability [14], although other formulations show increased persistence when dynamics are initiated with niche model versus other structures [68]. How structure influences dynamics, and vice-versa, is an open question.

Ecological Networks

This article has focused on food webs, which generally concern classic predator-herbivore-primary producer feeding interactions. However, the basic concept of food webs can be extended to a broader framework of "ecological networks" that is more inclusive of different components of ecosystem biomass flow, and that takes into consideration different kinds of species interactions that are not classic "predator-prey" interactions. Three examples are mentioned here. First, parasites have typically been given short shrift in traditional food webs, although exceptions exist (e. g., [51,67,74]). This is changing as it becomes clear that parasites are ubiquitous, often have significant impacts on predator-prey dynamics, and may be the dominant trophic habitat in most food webs, potentially altering our understanding of structure and dynam-

ics [59]. The dynamical models described previously have been parametrized with more conventional, non-parasite consumers in mind. An interesting open question is how altering dynamical model parameters such as metabolic rate, functional response, and consumer-resource body size ratios to reflect parasite characteristics will affect our understanding of food web stability.

Second, the role of detritus, or dead organic matter, in food webs has yet to be adequately resolved in either structural or dynamical approaches. Detritus has typically been included as one or several separate nodes in many binary-link and flow-weighted food webs. In some cases, it is treated as an additional "primary producer," while in other cases both primary producers and detritivores connect to it. Researchers must think much more carefully about how to include detritus in all kinds of ecological studies [80], given that it plays a fundamental role in most ecosystems and has particular characteristics that differ from other food web nodes: it is non-living organic matter, all species contribute to detrital pools, it is a major resource for many species, and the forms it takes are extremely heterogeneous (e. g., suspended organic matter in water columns; fecal material; rotting trees; dead animal bodies; small bits of plants and molted cuticle, skin, and hair mixed in soil; etc.).

Third, there are many interactions that species participate in that go beyond strictly trophic interactions. Plant-animal mutualistic networks, particularly pollination and seed dispersal or "frugivory" networks, have re-

ceived the most attention thus far. They are character-
ized as "bipartite" (two-level) graphs, with links from an-
imals to plants, but no links among plants or among ani-
mals [7,9,56,57,73,107]. While both pollination and seed
dispersal do involve a trophic interaction, with animals
gaining nutrition from plants during the interactions, un-
like in classic predator-prey relationships a positive bene-
fit is conferred upon both partners in the interaction. The
evolutionary and ecological dynamics of such mutualis-
tic relationships place unique constraints on the network
structure of these interactions and the dynamical stabil-
ity of such networks. For example, plant-animal mutualis-
tic networks are highly nested and thus asymmetric, such
that generalist plants and generalist animals tend to in-
teract among themselves, but specialist species (whether
plants or animals) also tend to interact with the most gen-
eralist species [7,107]. When simple dynamics are run on
these types of "coevolutionary" bipartite networks, it ap-
pears that the asymmetric structure enhances long-term
species coexistence and thus biodiversity maintenance [9].

Future Directions

Food web research of all kinds has expanded greatly over
the last several years, and there are many opportunities
for exciting new work at the intersection of ecology and
network structure and dynamics. In terms of empiricism,
there is still a paucity of detailed, evenly resolved com-
munity food webs in every habitat type. Current theory,
models, and applications need to be tested against more
diverse, more complete, and more highly quantified data.
In addition, there are many types of datasets that could be
compiled which would support novel research. For exam-
ple, certain kinds of fossil assemblages may allow the com-
pilation of detailed paleo food webs, which in turn could
allow examination of questions about how and why food
web structure does or does not change over deep time or
in response to major extinction events. Another example is
data illustrating the assembly of food [41] webs in partic-
ular habitats over ecological time. In particular, areas un-
dergoing rapid successional dynamics would be excellent
candidates, such as an area covered by volcanic lava flows,
a field exposed by a retreating glacier, a hillside denuded by
an earth slide, or a forest burned in a large fire. This type of
data would allow empirically-based research on the topo-
logical dynamics of food webs. Another empirical frontier
is the integration of multiple kinds of ecological interac-
tion data into networks with multiple kinds of links—for
example, networks that combine mutualistic interactions
such as pollination and antagonistic interactions such as
predator-prey relationships. In addition, more spatially

explicit food web data can be compiled across microhab-
itats or linked macrohabitats [8]. Most current food web
data is effectively aspatial even though trophic interactions
occur within a spatial context. More could also be done to
collect food web data based on specific instances of trophic
interactions. This was done for the insects that live inside
the stems of grasses in British fields. The web includes
multiple grass species, grass herbivores, their parasitoids,
hyper-parasitoids, and hyper-hyper parasitoids [67]. Dis-
section of over 160,000 grass stems allowed detailed quan-
tification of the frequency with which the species ($S = 77$
insect plus 10 grass species) and different trophic interac-
tions ($L = 126$) were observed.

Better empiricism will support improved and novel
analysis, modeling, and theory development and testing.
For example, while food webs appear fundamentally dif-
ferent in some ways from other kinds of "real-world" net-
works (e. g., they don't display power-law degree distribu-
tions), they also appear to share a common core network
structure that is scale-dependent with species richness and
connectance in predicable ways, as suggested by the suc-
cess of the niche and related models. Some of the dispar-
ity with other kinds of networks, and the shared structure
across food webs, may be explicable through finite-size ef-
fects or other methodological or empirical constraints or
artifacts. However, aspects of these patterns may reflect at-
tributes of ecosystems that relate to particular ecological,
evolutionary, or thermodynamic mechanisms underlying
how species are organized in complex bioenergetic net-
works of feeding interactions. Untangling artifacts from
attributes [63] and determining potential mechanisms un-
derlying robust phenomenological patterns (e. g., [10]) is
an important area of ongoing and future research. As
a part of this, there is much work to be done to continue
to integrate structure and dynamics of complex ecologi-
cal networks. This is critical for gaining a more compre-
hensive understanding of the conditions that underlie and
promote different aspects of stability, at different levels of
organization, in response to external perturbations and to
endogenous short- and long-term dynamics.

As the empiricism, analysis and modeling of food web
structure and dynamics improves, food web network re-
search can play a more central and critical role in con-
servation and management [76]. It is increasingly appar-
ent that an ecological network perspective, which encom-
passes direct and indirect effects among interacting taxa,
is critical for understanding, predicting, and managing the
impacts of species loss and invasion, habitat conversion,
and climate change. Far too often, critical issues of ecosys-
tem management have been decided on extremely limited
knowledge of one or a very few taxa. For example, this

has been an ongoing problem in fisheries science. The narrow focus of most research driving fisheries management decisions has resulted in overly optimistic assessments of sustainable fishing levels. Coupled with climate stressors, over-fishing appears to be driving steep, rapid declines in diversity of common predator target species, and probably many other kinds of associated taxa [114]. Until we acknowledge that species of interest to humans are embedded within complex networks of interactions that can produce unexpected effects through the interplay of direct and indirect effects, we will continue to experience negative outcomes from our management decisions [118]. An important part of minimizing and mitigating human impacts on ecosystems also involves research that explicitly integrates human and natural dynamics. Network research provides a natural framework for analyzing and modeling the complex ways in which humans interact with and impact the world's ecosystems, whether through local foraging or large-scale commercial harvesting driven by global economic markets.

These and other related research directions will depend on efficient management of increasingly dispersed and diversely formatted ecological and environmental data. Ecoinformatic tools—the technologies and practices for gathering, synthesizing, analyzing, visualizing, storing, retrieving and otherwise managing ecological knowledge and information—are playing an increasingly important role in the study of complex ecosystems, including food web research [47]. Indeed, ecology provides an excellent testbed for developing, implementing, and testing new information technologies in a biocomplexity research context (e. g., Semantic Prototypes in Research Ecoinformatics/SPiRE: spire.umbc.edu/us/; Science Environment for Ecological Knowledge/SEEK: seek.ecoinformatics.org; Webs on the Web/WoW: www.foodwebs.org). Synergistic ties between ecology, physics, computer science and other disciplines will dramatically increase the efficacy of research that takes advantage of such interdisciplinary approaches, as is currently happening in food web and related research.

Bibliography

Primary Literature

1. Albert R, Jeong H, Barabási AL (2000) Error and attack tolerance of complex networks. Nature 406:378–382
2. Albert R, Barabási AL (2002) Statistical mechanics of complex networks. Rev Mod Phys 74:47–97
3. Allesina S, Bodini A (2004) Who dominates whom in the ecosystem? Energy flow bottlenecks and cascading extinctions. J Theor Biol 230:351–358
4. Amaral LAN, Scala A, Berthelemy M, Stanley HE (2000) Classes of small-world networks. Proc Nat Acad Sci USA 97:11149–11152
5. Arii K, Parrott L (2004) Emergence of non-random structure in local food webs generated from randomly structured regional webs. J Theor Biol 227:327–333
6. Baird D, Ulanowicz RE (1989) The seasonal dynamics of the Chesapeake Bay ecosystem. Ecol Monogr 59:329–364
7. Bascompte J, Jordano P, Melián CJ, Olesen JM (2003) The nested assembly of plant-animal mutualistic networks. Proc Natl Acad Sci USA 100:9383–9387
8. Bascompte J, Melián CJ, Sala E (2005) Interaction strength combinations and the overfishing of a marine food web. Proc Natl Acad Sci 102:5443–5447
9. Bascompte J, Jordano P, Olesen JM (2006) Asymmetric coevolutionary networks facilitate biodiversity maintenance. Science 312:431–433
10. Beckerman AP, Petchey OL, Warren PH (2006) Foraging biology predicts food web complexity. Proc Natl Acad Sci USA 103:13745–13749
11. Bersier L-F, Banašek-Richter C, Cattin M-F (2002) Quantitative descriptors of food web matrices. Ecology 83:2394–2407
12. Briand F, Cohen JE (1984) Community food webs have scale-invariant structure. Nature 398:330–334
13. Brose U, Williams RJ, Martinez ND (2003) Comment on "Foraging adaptation and the relationship between food-web complexity and stability". Science 301:918b
14. Brose U, Williams RJ, Martinez ND (2006) Allometric scaling enhances stability in complex food webs. Ecol Lett 9:1228–1236
15. Camacho J, Arenas A (2005) Universal scaling in food-web structure? Nature 435:E3-E4
16. Camacho J, Guimerà R, Amaral LAN (2002) Robust patterns in food web structure. Phys Rev Lett 88:228102
17. Camacho J, Guimerà R, Amaral LAN (2002) Analytical solution of a model for complex food webs. Phys Rev Lett E 65:030901
18. Cartoza CC, Garlaschelli D, Caldarelli G (2006) Graph theory and food webs. In: Pascual M, Dunne JA (eds) Ecological Networks: Linking Structure to Dynamics in Food Webs. Oxford University Press, New York, pp 93–117
19. Cattin M-F, Bersier L-F, Banašek-Richter C, Baltensperger M, Gabriel J-P (2004) Phylogenetic constraints and adaptation explain food-web structure. Nature 427:835–839
20. Chen X, Cohen JE (2001) Global stability, local stability and permanence in model food webs. J Th Biol 212:223–235
21. Chen X, Cohen JE (2001) Transient dynamics and food web complexity in the Lotka–Volterra cascade model. Proc Roy Soc Lond B 268:869–877
22. Christian RR, Luczkovich JJ (1999) Organizing and understanding a winter's seagrass foodweb network through effective trophic levels. Ecol Model 117:99–124
23. Cohen JE (1977) Ratio of prey to predators in community food webs. Nature 270:165–167
24. Cohen JE (1977) Food webs and the dimensionality of trophic niche space. Proc Natl Acad Sci USA 74:4533–4563
25. Cohen JE (1978) Food Webs and Niche Space. Princeton University Press, NJ
26. Cohen JE (1989) Ecologists Co-operative Web Bank (ECOWeB™). Version 1.0. Machine Readable Data Base of Food Webs. Rockefeller University, NY

27. Cohen JE, Briand F (1984) Trophic links of community food webs. Proc Natl Acad Sci USA 81:4105–4109

28. Cohen JE, Newman CM (1985) A stochastic theory of community food webs: I. Models and aggregated data. Proc R Soc Lond B 224:421–448

29. Cohen JE, Palka ZJ (1990) A stochastic theory of community food webs: V. Intervality and triangulation in the trophic niche overlap graph. Am Nat 135:435–463

30. Cohen JE, Briand F, Newman CM (1986) A stochastic theory of community food webs: III. Predicted and observed length of food chains. Proc R Soc Lond B 228:317–353

31. Cohen JE, Briand F, Newman CM (1990) Community Food Webs: Data and Theory. Springer, Berlin

32. Cohen JE, Luczak T, Newman CM, Zhou Z-M (1990) Stochastic structure and non-linear dynamics of food webs: qualitative stability in a Lotka–Volterra cascade model. Proc R Soc Lond B 240:607–627

33. Cohen JE, Jonsson T, Carpenter SR (2003) Ecological community description using the food web, species abundance, and body size. Proc Natl Acad Sci USA 100:1781–1786

34. Daily GC (ed) (1997) Nature's services: Societal dependence on natural ecosystems. Island Press, Washington DC

35. Doyle J, Csete M (2007) Rules of engagement. Nature 446:860

36. Dunne JA (2006) The network structure of food webs. In: Pascual M, Dunne JA (eds) Ecological Networks: Linking Structure to Dynamics in Food Webs. Oxford University Press, New York, pp 27–86

37. Dunne JA, Williams RJ, Martinez ND (2002) Food web structure and network theory: the role of connectance and size. Proc Natl Acad Sci USA 99:12917–12922

38. Dunne JA, Williams RJ, Martinez ND (2002) Network structure and biodiversity loss in food webs: robustness increases with connectance. Ecol Lett 5:558–567

39. Dunne JA, Williams RJ, Martinez ND (2004) Network structure and robustness of marine food webs. Mar Ecol Prog Ser 273:291–302

40. Dunne JA, Brose U, Williams RJ, Martinez ND (2005) Modeling food-web dynamics: complexity-stability implications. In: Belgrano A, Scharler U, Dunne JA, Ulanowicz RE (eds) Aquatic Food Webs: An Ecosystem Approach. Oxford University Press, New York, pp 117–129

41. Dunne JA, Williams RJ, Martinez ND, Wood RA, Erwing DE (2008) Compilation and network analyses of Cambrian food webs. PLoS Biology 5:e102. doi:10.1371/journal.pbio.0060102

42. Egerton FN (2007) Understanding food chains and food webs, 1700–1970. Bull Ecol Soc Am 88(1):50–69

43. Elton CS (1927) Animal Ecology. Sidgwick and Jackson, London

44. Elton CS (1958) Ecology of Invasions by Animals and Plants. Chapman & Hall, London

45. Garlaschelli D, Caldarelli G, Pietronero L (2003) Universal scaling relations in food webs. Nature 423:165–168

46. Goldwasser L, Roughgarden JA (1993) Construction of a large Caribbean food web. Ecology 74:1216–1233

47. Green JL, Hastings A, Arzberger P, Ayala F, Cottingham KL, Cuddington K, Davis F, Dunne JA, Fortin M-J, Gerber L, Neubert M (2005) Complexity in ecology and conservation: mathematical, statistical, and computational challenges. Bioscience 55:501–510

48. Hardy AC (1924) The herring in relation to its animate environment. Part 1. The food and feeding habits of the herring with special reference to the East Coast of England. Fish Investig Ser II 7:1–53

49. Havens K (1992) Scale and structure in natural food webs. Science 257:1107–1109

50. Hutchinson GE (1959) Homage to Santa Rosalia, or why are there so many kinds of animals? Am Nat 93:145–159

51. Huxham M, Beany S, Raffaelli D (1996) Do parasites reduce the chances of triangulation in a real food web? Oikos 76:284–300

52. Jeong H, Tombor B, Albert R, Oltvia ZN, Barabási A-L (2000) The large-scale organization of metabolic networks. Nature 407:651–654

53. Jeong H, Mason SP, Barabási A-L, Oltavia ZN (2001) Lethality and centrality in protein networks. Nature 411:41

54. Jordán F, Molnár I (1999) Reliable flows and preferred patterns in food webs. Ecol Ecol Res 1:591–609

55. Jordán F, Scheuring I (2004) Network ecology: topological constraints on ecosystem dynamics. Phys Life Rev 1:139–229

56. Jordano P (1987) Patterns of mutualistic interactions in pollination and seed dispersal: connectance, dependence asymmetries, and coevolution. Am Nat 129:657–677

57. Jordano P, Bascompte J, Olesen JM (2003) Invariant properties in co-evolutionary networks of plant-animal interactions. Ecol Lett 6:69–81

58. Kondoh M (2003) Foraging adaptation and the relationship between food-web complexity and stability. Science 299:1388–1391

59. Lafferty KD, Dobson AP, Kurls AM (2006) Parasites dominate food web links. Proc Nat Acad Sci USA 103:11211–11216

60. Lindeman RL (1942) The trophic-dynamic aspect of ecology. Ecology 23:399–418

61. Link J (2002) Does food web theory work for marine ecosystems? Mar Ecol Prog Ser 230:1–9

62. MacArthur RH (1955) Fluctuation of animal populations and a measure of community stability. Ecology 36:533–536

63. Martinez ND (1991) Artifacts or attributes? Effects of resolution on the Little Rock Lake food web. Ecol Monogr 61:367–392

64. Martinez ND (1992) Constant connectance in community food webs. Am Nat 139:1208–1218

65. Martinez ND (1993) Effect of scale on food web structure. Science 260:242–243

66. Martinez ND (1994) Scale-dependent constraints on food-web structure. Am Nat 144:935–953

67. Martinez ND, Hawkins BA, Dawah HA, Feifarek BP (1999) Effects of sampling effort on characterization of food-web structure. Ecology 80:1044–1055

68. Martinez ND, Williams RJ, Dunne JA (2006) Diversity, complexity, and persistence in large model ecosystems. In: Pascual M, Dunne JA (eds) Ecological Networks: Linking Structure to Dynamics in Food Webs. Oxford University Press, New York, pp 163–185

69. May RM (1972) Will a large complex system be stable? Nature 238:413–414

70. May RM (1973) Stability and Complexity in Model Ecosystems. Princeton University Press, Princeton. Reprinted in 2001 as a "Princeton Landmarks in Biology" edition

71. McCann KS (2000) The diversity-stability debate. Nature 405:228–233

72. McKane AJ, Drossel B (2006) Models of food-web evolution. In: Pascual M, Dunne JA (eds) Ecological Networks: Linking

Structure to Dynamics in Food Webs. Oxford University Press, New York, pp 223–243

73. Memmott J (1999) The structure of a plant-pollinator network. Ecol Lett 2:276–280

74. Memmott J, Martinez ND, Cohen JE (2000) Predators, parasitoids and pathogens: species richness, trophic generality and body sizes in a natural food web. J Anim Ecol 69:1–15

75. Memmott J, Waser NM, Price MV (2004) Tolerance of pollination networks to species extinctions. Proc Royal Soc Lond Series B 271:2605–2611

76. Memmott J, Alonso D, Berlow EL, Dobson A, Dunne JA, Sole R, Weitz J (2006) Biodiversity loss and ecological network structure. In: Pascual M, Dunne JA (eds) Ecological Networks: Linking Structure to Dynamics in Food Webs. Oxford University Press, New York, pp 325–347

77. Milo R, Shen-Orr S, Itzkovitz S, Kashtan N, Chklovskii D, Alon U (2002) Network motifs: simple building blocks of complex networks. Science 298:763–764

78. Montoya JM, Solé RV (2002) Small world patterns in food webs. J Theor Biol 214:405–412

79. Montoya JM, Solé RV (2003) Topological properties of food webs: from real data to community assembly models. Oikos 102:614–622

80. Moore JC, Berlow EL, Coleman DC, de Ruiter PC, Dong Q, Hastings A, Collin Johnson N, McCann KS, Melville K, Morin PJ, Nadelhoffer K, Rosemond AD, Post DM, Sabo JL, Scow KM, Vanni MJ, Wall DH (2004) Detritus, trophic dynamics and biodiversity. Ecol Lett 7:584–600

81. Neutel AM, Heesterbeek JAP, de Ruiter PC (2002) Stability in real food webs: weak links in long loops. Science 296:1120–1123

82. Newman MEJ (2002) Assortative mixing in networks. Phys Rev Lett 89:208701

83. Newman M, Barabasi A-L, Watts DJ (eds) (2006) The Structure and Dynamics of Networks. Princeton University Press, Princeton

84. Odum E (1953) Fundamentals of Ecology. Saunders, Philadelphia

85. Pascual M, Dunne JA (eds) (2006) Ecological Networks: Linking Structure to Dynamics in Food Webs. Oxford University Press, New York

86. Paine RT (1988) Food webs: road maps of interactions or grist for theoretical development? Ecology 69:1648–1654

87. Pierce WD, Cushamn RA, Hood CE (1912) The insect enemies of the cotton boll weevil. US Dept Agric Bull 100:9–99

88. Pimm SL (1982) Food Webs. Chapman and Hall, London. Reprinted in 2002 as a 2nd edition by University of Chicago Press

89. Pimm SL (1984) The complexity and stability of ecosystems. Nature 307:321–326

90. Pimm SL, Lawton JH (1978) On feeding on more than one trophic level. Nature 275:542–544

91. Pimm SL, Lawton JH (1980) Are food webs divided into compartments? J Anim Ecol 49:879–898

92. Polis GA (1991) Complex desert food webs: an empirical critique of food web theory. Am Nat 138:123–155

93. Schoener TW (1989) Food webs from the small to the large. Ecology 70:1559–1589

94. Schoenly K, Beaver R, Heumier T (1991) On the trophic relations of insects: a food web approach. Am Nat 137:597–638

95. Solé RV, Montoya JM (2001) Complexity and fragility in ecological networks. Proc R Soc Lond B 268:2039–2045

96. Solow AR (1996) On the goodness of fit of the cascade model. Ecology 77:1294–1297

97. Solow AR, Beet AR (1998) On lumping species in food webs. Ecology 79:2013–2018

98. Srinivasan UT, Dunne JA, Harte H, Martinez ND (2007) Response of complex food webs to realistic extinction sequences. Ecology 88:671–682

99. Stouffer DB, Camacho J, Guimera R, Ng CA, Amaral LAN (2005) Quantitative patterns in the structure of model and empirical food webs. Ecology 86:1301–1311

100. Stouffer DB, Camacho J, Amaral LAN (2006) A robust measure of food web intervality. Proc Nat Acad Sci 103:19015–19020

101. Strogatz SH (2001) Exploring complex networks. Nature 410:268–275

102. Sugihara G, Schoenly K, Trombla A (1989) Scale invariance in food web properties. Science 245:48–52

103. Summerhayes VS, Elton CS (1923) Contributions to the ecology of Spitzbergen and Bear Island. J Ecol 11:214–286

104. Summerhayes VS, Elton CS (1928) Further contributions to the ecology of Spitzbergen and Bear Island. J Ecol 16:193–268

105. Thompson RM, Townsend CR (1999) The effect of seasonal variation on the community structure and food-web attributes of two streams: implications for food-web science. Oikos 87:75–88

106. Thompson RM, Townsend CR (2005) Food web topology varies with spatial scale in a patchy environment. Ecology 86:1916–1925

107. Vásquez DP, Aizen MA (2004) Asymmetric specialization: a pervasive feature of plant-pollinator interactions. Ecology 85:1251–1257

108. Warren PH (1989) Spatial and temporal variation in the structure of a freshwater food web. Oikos 55:299–311

109. Watts DJ, Strogatz SH (1998) Collective dynamics of 'small-world' networks. Nature 393:440–442

110. Williams RJ, Martinez ND (2000) Simple rules yield complex food webs. Nature 404:180–183

111. Williams RJ, Martinez ND (2004) Trophic levels in complex food webs: theory and data. Am Nat 163:458–468

112. Williams RJ, Martinez ND (2004) Diversity, complexity, and persistence in large model ecosystems. Santa Fe Institute Working Paper 04-07-022

113. Williams RJ, Berlow EL, Dunne JA, Barabási AL, Martinez ND (2002) Two degrees of separation in complex food webs. Proc Natl Acad Sci USA 99:12913–12916

114. Worm B, Sandow M, Oschlies A, Lotze HK, Myers RA (2005) Global patterns of predator diversity in the open oceans. Science 309:1365–1369

115. Yodzis P (1980) The connectance of real ecosystems. Nature 284:544–545

116. Yodzis P (1984) The structure of assembled communities II. J Theor Biol 107:115–126

117. Yodzis P (1998) Local trophodynamics and the interaction of marine mammals and fisheries in the Benguela ecosystem. J Anim Ecol 67:635–658

118. Yodzis P (2000) Diffuse effects in food webs. Ecology 81:261–266

119. Yodzis P, Innes S (1992) Body-size and consumer-resource dynamics. Am Nat 139:1151–1173.

Books and Reviews

Belgrano A, Scharler U, Dunne JA, Ulanowicz RE (eds) (2005) Aquatic Food Webs: An Ecosystem Approach. Oxford University Press, Oxford

Berlow EL, Neutel A-M, Cohen JE, De Ruiter P, Ebenman B, Emmerson M, Fox JW, Jansen VAA, Jones JI, Kokkoris GD, Logofet DO, McKane AJ, Montoya J, Petchey OL (2004) Interaction strengths in food webs: issues and opportunities. J Animal Ecol 73:585–598

Borer ET, Anderson K, Blanchette CA, Broitman B, Cooper SD, Halpern BS (2002) Topological approaches to food web analyses: a few modifications may improve our insights. Oikos 99:397–401

Christensen V, Pauly D (1993) Trophic Models of Aquatic Ecosystems. ICLARM, Manila

Cohen JE, Beaver RA, Cousins SH, De Angelis DL, et al (1993) Improving food webs. Ecology 74:252–258

Cohen JE, Briand F, Newman CM (1990) Community Food Webs: Data and Theory. Springer, Berlin

DeAngelis DL, Post WM, Sugihara G (eds) (1983) Current Trends in Food Web Theory. ORNL-5983, Oak Ridge Natl Laboratory

Drossel B, McKane AJ (2003) Modelling food webs. In: Bornholt S, Schuster HG (eds) Handbook of Graphs and Networks: From the Genome to the Internet. Wiley-VCH, Berlin

Hall SJ, Raffaelli DG (1993) Food webs: theory and reality. Advances in Ecological Research 24:187–239

Lawton JH (1989) Food webs. In Cherett JM, (ed) Ecological Concepts. Blackwell Scientific, Oxford

Lawton JH, Warren PH (1988) Static and dynamic explanations for patterns in food webs. Trends in Ecology and Evolution 3:242–245

Martinez ND (1995) Unifying ecological subdisciplines with ecosystem food webs. In Jones CG, Lawton JH, (eds) Linking Species and Ecosystems. Chapman and Hall, New York

Martinez ND, Dunne JA (1998) Time, space, and beyond: scale issues in food-web research. In Peterson D, Parker VT, (eds) Ecological Scale: Theory and Applications. Columbia University Press, New York

May RM (1983) The structure of food webs. Nature 301:566–568

May RM (2006) Network structure and the biology of populations. Trends Ecol Evol 21:394–399

Montoya JM, Pimm SL, Sole RV (2006) Ecological networks and their fragility. Nature 442:259–264

Moore J, de Ruiter P, Wolters V (eds) (2005) Dynamic Food Webs: Multispecies Assemblages, Ecosystem Development and Environmental Change. Academic Press, Elsevier, Amsterdam

Pimm SL, Lawton JH, Cohen JE (1991) Food web patterns and their consequences. Nature 350:669–674

Polis GA, Winemiller KO, (eds) (1996) Food Webs: Integration of Patterns & Dynamics. Chapman and Hall

Polis GA, Power ME, Huxel GR, (eds) (2003) Food Webs at the Landscape Level. University of Chicago Press

Post DM (2002) The long and short of food-chain length. Trends Ecol Evol 17:269–277

Strong DR (ed) (1988) Food web theory: a ladder for picking strawberries. Special Feature. Ecology 69:1647–1676

Warren PH (1994) Making connections in food webs. Trends Ecol Evol 9:136–141

Woodward G, Ebenman B, Emmerson M, Montoya JM, Olesen JM, Valido A, Warren PH (2005) Body size in ecological networks. Trends Ecol Evol 20:402–409

Foraging Robots

ALAN FT WINFIELD
University of the West of England, Bristol, UK

Article Outline

Glossary

Autonomy In robotics autonomy conventionally refers to the degree to which a robot is able to make its own decisions about which actions to take next. Thus a fully autonomous robot would be capable of carrying out its entire mission or function without human control or intervention. A semi-autonomous robot would have a degree of autonomy but require some human supervision.

Behavior-based control Behavior-based control describes a class of robot control systems characterized by a set of conceptually independent task achieving modules, or behaviors. All task achieving modules are able to access the robot's sensors and when a particular module becomes active it is able to temporarily take control of the robot's actuators [2].

Braitenberg vehicle In robotics a Braitenberg vehicle is a conceptual mobile robot in which simple sensors are connected directly to drive wheels. Thus if, for instance, a front-left-side sensor is connected to the right-side drive wheel and vice-versa, then if the sensors are light sensitive the robot will automatically steer towards a light source [11].

Finite state machine In the context of this article a finite state machine (FSM) is a model of robot behavior which has a fixed number of states. Each state represents a particular set of actions or behaviors. The robot can be in only one of these states at any given instant in time and transitions between states may be triggered by either external or internal events.

Odometry Odometry refers to the technique of self-localization in which a robot measures how far it has

traveled by, for instance, counting the revolutions of its wheels. Odometry suffers the problem that wheel-slip leads to cumulative errors so odometric position estimates are generally inaccurate and of limited value unless combined with other localization techniques.

Robot In this article the terms *robot* and *mobile robot* are used interchangeably. A mobile robot is a man-made device or vehicle capable of (1) sensing its environment and (2) purposefully moving through and acting upon or within that environment. A robot may be fully autonomous, semi-autonomous or tele-operated.

Swarm intelligence The term swarm intelligence describes the purposeful collective behaviors observed in nature, most dramatically in social insects. Swarm intelligence is the study of those collective behaviors, in both natural and artificial systems of multiple agents, and how they emerge from the local interactions of the agents with each other and with their environment [8,19].

Tele-operation A robot is said to be tele-operated if it is remotely controlled by a human operator.

Definition of the Subject

Foraging robots are mobile robots capable of searching for and, when found, transporting objects to one or more collection points. Foraging robots may be single robots operating individually, or multiple robots operating collectively. Single foraging robots may be remotely tele-operated or semi-autonomous; multiple foraging robots are more likely to be fully autonomous systems. In robotics foraging is important for several reasons: firstly, it is a metaphor for a broad class of problems integrating exploration, navigation and object identification, manipulation and transport; secondly, in multi-robot systems foraging is a canonical problem for the study of robot-robot cooperation, and thirdly, many actual or potential real-world applications for robotics are instances of foraging robots, for instance cleaning, harvesting, search and rescue, landmine clearance or planetary exploration.

Introduction

Foraging is a benchmark problem for robotics, especially for multi-robot systems. It is a powerful benchmark problem for several reasons: (1) sophisticated foraging observed in social insects, recently becoming well-understood, provides both inspiration and system level models for artificial systems. (2) Foraging is a complex task involving the coordination of several – each also difficult – tasks including efficient exploration (searching) for food or prey, physical collection (harvesting) of food or prey al-

most certainly requiring physical manipulation, transport of the food or prey, homing or navigation whilst carrying the food or prey back to a nest site, and deposition of the food item in the nest before returning to foraging. (3) Effective foraging requires cooperation between individuals involving either communication to signal to others where food or prey may be found (e. g. pheromone trails, or direction giving) and/or cooperative transport of food items too large for a single individual to transport.

There are, at the time of writing, very few types of foraging robots successfully employed in real-world applications. Most foraging robots are to be found in research laboratories or, if they are aimed at real-world applications, are at the stage of prototype or proof-of-concept. The reason for this is that foraging is a complex task which requires a range of competencies to be tightly integrated within the physical robot and, although the principles of robot foraging are now becoming established, many of the sub-system technologies required for foraging robots remain very challenging. In particular, sensing and situational awareness; power and energy autonomy; actuation, locomotion and safe navigation in unknown physical environments and proof of safety and dependability all remain difficult problems in robotics.

This article therefore focuses on describing and defining the principles of robot foraging. The majority of examples will necessarily be of laboratory systems not aimed at solving real-world applications but designed to model, illuminate and demonstrate those principles. The article proceeds as follows. Section "An Abstract Model of Robot Foraging" describes an abstract model of robot foraging, using a finite state machine (FSM) description to define the discrete sub-tasks, or states, that constitute foraging. The FSM method will be used throughout this article. The section then develops a taxonomy of robot foraging. Section "Single Robot Foraging" describes the essential design features that are a requirement of any foraging robot, whether operating singly or in a multi-robot team, and the technologies currently available to implement those features; the section then outlines a number of examples of single-robot foraging, including robots that are commercially available. Section "Multi-Robot (Collective) Foraging" then describes the development and state-of-the-art in multi-robot (collective) foraging; strategies for cooperation are described including information sharing, cooperative transport and division of labor (task allocation), the section then reviews approaches to the mathematical modeling of multi-robot foraging. The article concludes in Sect. "Future Directions" with a discussion of future directions in robot foraging and an outline of the technical challenges that remain to be solved.

An Abstract Model of Robot Foraging

Foraging, by humans or animals, is the act of searching (widely) for and collecting (or capturing) food for storage or consumption. Robot foraging is defined more broadly as searching for and collecting *any* objects, then returning those objects to a collection point. Of course if the robot(s) are searching for energy resources for themselves then robot foraging will have precisely the same meaning as human or animal foraging. In their definitive review paper on cooperative mobile robotics Cao et al. state simply "In foraging, a group of robots must pick up objects scattered in the environment" [14]. Østergaard et al. define foraging as "a two-step repetitive process in which (1) robots search a designated region of space for certain objects, and (2) once found these objects are brought to a goal region using some form of navigation" [54].

Figure 1 shows a Finite State Machine (FSM) representation of a foraging robot (or robots). In the model the robot is in always in one of four states: *searching*, *grabbing*, *homing* or *depositing*. Implied in this model is, firstly, that the environment or search space contains more than one of the target objects; secondly, that there is a single collection point (hence this model is sometimes referred to as central-place foraging), and thirdly, that the process continues indefinitely. The four states are defined as follows.

1. Searching. In this state the robot is physically moving through the search space using its sensors to locate and recognize the target items. At this level of abstraction we do not need to state how the robot searches: it could, for instance, wander at random, or it could employ a systematic strategy such as moving alternately left and right in a search pattern. The fact that the robot has to search at all follows from the pragmatic real-world assumptions that either the robot's sensors are of short range and/or the items are hidden (behind occluding obstacles for instance); in either event we must assume that the robot cannot find items simply by staying in one place and scanning the whole environment with its sensors. Object identification or recognition could require one of a wide range of sensors and techniques. When the robot finds an item it changes state from searching to grabbing. If the robot fails to find the target item then it remains in the searching state forever; searching is therefore the 'default' state.
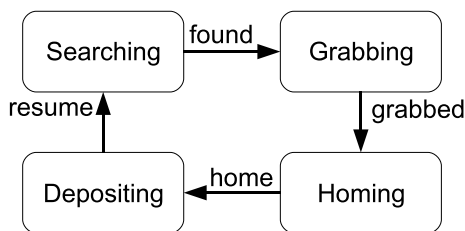
2. Grabbing. In this state the robot physically captures and grabs the item ready to transport it back to the home region. Here we assume that the item is capable of being grabbed and conveyed by a single robot (the case of larger items that require cooperative transport by more than one robot will be covered later in this article). As soon as the item has been grabbed the robot will change state to homing.

3. Homing. In this state the robot must move, with its collected object, to a home or nest region. Homing clearly requires a number of stages, firstly, determination of the position of the home region relative to where the robot is now, secondly, orientation toward that position and, thirdly, navigation to the home region. Again there are a number of strategies for homing: one would be to re-trace the robot's path back to the home region using, for instance, odometry or by following a marker trail; another would be to home in on a beacon with a long range beacon sensor. When the robot has successfully reached the home region it will change state to depositing.

4. Depositing. In this state the robot deposits or delivers the item in the home region, and then immediately changes state to searching and hence resumes its search.

There are clearly numerous variations on this basic foraging model. Some are simplifications: for instance if a robot is searching for one or a known fixed number of objects then the process will not loop indefinitely. Real robots do not have infinite energy and so a model of practical foraging would need to take account of energy management. However, many variations entail either complexity within one or more of the four basic states (consider, for instance, objects that actively evade capture – a predator-prey model of foraging), or complexity in the interaction or cooperation between robots in multi-robot foraging. Thus the basic model stands as a powerful top-level abstraction.

A Taxonomy of Robot Foraging

Oster and Wilson classify the foraging strategies of social insects into five types, summarized in Table 1 [53]. Hölldobler and Wilson describe a more comprehensive taxonomy of insect foraging as a combination of strategies for (1) hunting, (2) retrieval and (3) defense [30]. However, since we will not be concerned in this article with

Foraging Robots, Figure 1
Finite State Machine for Basic Foraging

Foraging Robots, Table 1
Oster and Wilson's classification of insect foraging

Type	Description
I	Solitary insects find and retrieve prey singly
II	As I except that solitary foragers signal the location of food to other insects
III	Foragers depart the nest and follow 'trunk trails' before branching off to search unmarked terrain
IV	As II except that a group of insects assaults or retrieves the prey en-masse
V	Multiple insects forage as groups

defensive robot(s), then Oster and Wilson's classification is more than sufficient as a basis for consideration of robot foraging.

In robotics several taxonomies have been proposed for multi-robot systems. Dudek et al. define seven taxonomic axes: collective size; communications [range, topology and bandwidth]; collective reconfigurability; processing ability and collective composition [21]. Here collective size may be: single robot, pair of robots, limited (in relation to the size of the environment) or infinite (number of robots $N_r \gg 1$); communications range may be: none (i. e. robots do not communicate directly), near (robots have limited range communication) or infinite (any robot may communicate with any other). Collective reconfigurability refers to spatial organization and may be: static (robots are in a fixed formation); coordinated (robots may coordinate to alter their formation) or dynamic (spatial organization may change arbitrarily). Processing ability refers to the computational model of individuals, here Dudek et al. make the distinction between the general purpose computer which most practical robots will have, or simpler models including the finite state machine. Collective composition may be: identical (robots are both physically and functionally identical), homogeneous or heterogeneous. Dudek et al. makes the distinction – highly relevant to foraging robots – between tasks that are *traditionally single-agent*, tasks that are *traditionally multi-agent*, tasks that *require* multiple agents, or tasks that *may benefit* from multiple agents.

In contrast to Dudek's taxonomy which is based upon the characteristics of the robot(s), Balch characterizes tasks and rewards [3]. Balch's task taxonomy is particularly relevant to robot foraging because it leads naturally to the definition of performance metrics. Balch articulates six task axes: time; criteria; subject of action; resource limits; group movement and platform capabilities. Time and criteria are linked; time may be: limited (task performance is determined by how much can be achieved in the fixed time);

minimum (task performance is measured as time taken to complete the task); unlimited time, or synchronized (robots must synchronize their actions). Criteria refers to how performance is optimized over time; it may be finite (performance is summed over a finite number of time steps); average (performance is averaged over all time) or discounted (future performance is discounted geometrically). Subject of action may be: object- or robot-based, depending upon whether the movement or positioning of objects or robots, respectively, is important. Balch's fourth criterion is again relevant to foraging: resource limits which may be: limited (external resources, i. e. objects to be foraged, are limited); energy (energy consumption must be minimized); internally competitive (one robot's success reduces the likelihood of success of another), or externally competitive (if, for instance, one robot team competes against another). See also [24] for a formal analysis and taxonomy of task allocation.

Østergaard et al. [54] develop a simple taxonomy of foraging by defining eight characteristics each of which has two values:

1. number of robots: single or multiple;
2. number of sinks (collection points for foraged items): single or multiple;
3. number of source areas (of objects to be collected): single or multiple;
4. search space: unbounded or constrained;
5. number of types of object to be collected: single or multiple;
6. object placement: in fixed areas or randomly scattered;
7. robots: homogeneous or heterogeneous and
8. communication: none or with.

This taxonomy maps more closely (but not fully) onto the insect foraging taxonomy of Table 1, but fails to capture task performance criteria, nor does it specify the strategy for either searching for, physically collecting or retrieving objects. Tables 2 and 3 propose a more comprehensive taxonomy for robot foraging that incorporates the robot-centric and task/performance oriented features of Dudek et al. and Balch, respectively, with the environmental features of Østergaard et al., whilst mapping onto the insect foraging classification of Oster and Wilson. The four major axes are Environment, Robot(s), Performance and Strategy. Each major axis has several minor axes and each of these can take the values enumerated in the third column of Tables 2 and 3. The majority of the values are self-explanatory, those that are not are annotated. Table 3 suggests a mapping of Oster and Wilson's classification onto robot foraging strategies.

Foraging Robots, Table 2
A taxonomy of robot foraging, part A

Major Axis	Minor Axis	Value	Notes
Environment	Search space	Unbounded	
		Constrained	
	Source areas	Single limited	Fixed number of objects
		Single unlimited	Objects 're-grow'
		Multiple	
	Sinks	Single	Home, nest or collection point
		Multiple	
	Object types	Single static	One type of static object, food or 'prey'
		Multiple static	
		Single active	One type of prey which evades capture
	Object placement	Fixed known locations	
		Uniform distribution	
		Clustered	
Robot(s)	Number	Single	
		Multiple	
	Type	Homogeneous	
		Heterogeneous	
	Object sensing	Limited	Short-range sensing
		Unlimited	Unlimited-range sensing
	Localization	None	
		Relative	
		Absolute	
	Communications	None	
		Near	
		Infinite	
	Power	Limited	Robot can run out of energy
		Forage	Robot forages for own energy
		Unlimited	

Following Balch [3], we can formalize successful object collection and retrieval as follows:

$$F(O_i, t) = \begin{cases} 1 & \text{if object } O_i \text{ is in a sink at time } t \\ 0 & \text{otherwise} . \end{cases} \quad (1)$$

If the foraging task is performance time limited (Performance time = fixed) and the objective is to maximize the number of objects foraged within fixed time T, then we may define a performance metric for the number of objects collected in time T,

$$P = \sum_{i=1}^{N_o} F(O_i, t_0 + T) \quad (2)$$

where N_o is the number of objects available for collection and t_0 is the start time. A metric for the number of objects foraged per second is clearly, $P_t = P/T$. P as defined here

is independent of the number of robots. In order to measure the performance improvement of multi-robot foraging, for example the benefit gained by search or homing with trail following, recruitment or coordination (assuming the task can be completed by a single robot, grabbing = single and transport = single), then we may define the performance of a single robot P_s as defined in Eq. 2 and use this a baseline for the normalized performance P_m of a multi-robot system,

$$P_m = \frac{P}{N_r} \quad (3)$$

where N_r is the total number of robots. The efficiency of multi-robot foraging is then the ratio P_m/P_s.

Consider now that we wish instead to minimize the energy cost of foraging (Performance energy = minimum). If the energy cost of foraging object i is E_i, then we may define a performance metric for the number of objects for-

Foraging Robots, Table 3
A taxonomy of robot foraging, part B

Major Axis	Minor Axis	Value	Notes
Performance	Time	Fixed	Objects foraged per second
		Minimum	Minimize time to forage
		Unlimited	
	Energy	Fixed	Objects foraged per Joule
		Minimum	Minimize energy used
		Unlimited	
Strategy	Search	Random wander	
		Geometrical pattern	
		Trail following	Type III
		Follow other robots	
		In teams	Type V
	Grabbing	Single	
		Cooperative	Type IV
	Transport	Single	
		Cooperative	Type IV
	Homing	Self-navigation	
		Home on beacon	
		Follow trail	
	Recruitment	None	Type I
		Direct	Type II
		Indirect	
	Coordination	None	Type I
		Self-organized	Types II-V
		Master slave	
		Central control	

aged per Joule of energy,

$$P_e = \frac{N_o}{\sum_{i=1}^{N_o} E_i} \qquad (4)$$

then seek the foraging strategy that achieves the highest value for P_e.

Single Robot Foraging

The design of any foraging robot, whether operating alone or as part of a multi-robot team, will necessarily follow a similar basic pattern. The robot will require one or more *sensors*, with which it can both sense its environment for safe navigation and detect the objects or food-items it seeks; *actuators* for both locomotion through the environment and for physically collecting, holding then depositing its prey, and a *control system* to provide the robot with – at the very least – a set of basic reflex behaviors. Since robots are machines that perform work, which requires energy, then *power management* is important; if, for instance, the robot is foraging for its own energy then balancing its energy needs with the energy cost of foraging is clearly critical. Normally, a *communication* transceiver is also a requirement, either to allow remote tele-operation or monitoring or, in the case of multi-robot collective foraging, for robot-robot communications. A foraging robot is therefore a complex set of interconnected sub-systems and, although its system-level structure may follow a standard pattern, the shape and form of the robot will vary significantly depending upon its intended environment and application.

This section will review approaches and techniques for sensing, actuation, communications and control, within the context of robot foraging and with reference to research which focuses on advancing specific capabilities within each of these domains of interest. Then a number of examples of single robot foraging are given, including real-world applications.

Sensing

Obstacle Avoidance and Path Planning There are many sensors available to designers of foraging robots and

a comprehensive review can be found in [22]. A foraging robot will typically require short or medium range proximity sensors for obstacle avoidance, such as infra-red return-signal-intensity or ultrasonic- or laser-based time-of-flight systems. The most versatile and widely used device is the 2D or 3D scanning laser range finder which can provide the robot with a set of radial distance measurements and hence allow the robot to plan a safe path through obstacles [64].

Localization   All but the simplest foraging robots will also require sensors for localization, that is to enable the robot to estimate its own position in the environment. If external reference signals are available such as fixed beacons so that a robot can use radio trilateration to fix its position relative to those beacons, or a satellite navigation system such as the Global Positioning System (GPS), then localization is relatively straightforward. If no external infrastructure is available then a robot will typically make use of several sensors including odometry, an inertial measurement unit (IMU) and a magnetic compass, often combining the data from all of these sensors, including laser scanning data, to form an estimate of its position. Simultaneous Localization and Mapping (SLAM) is a well-known stochastic approach which typically employs Kalman filters to allow a robot (or a team of robots) to both fix their position relative to observed landmarks and map those landmarks with increasing confidence as the robot(s) move through the environment [18].

Object Detection   Vision is often the sensor of choice for object detection in laboratory experiments in foraging robots. If, for instance, the object of interest has a distinct color which stands out in the environment then standard image processing techniques can be used to detect then steer towards the object [31]. However, if the environment is visually cluttered, unknown or poorly illuminated then vision becomes problematical. Alternative approaches to object detection include, for instance, artificial odor sensors: Hayes et al. demonstrated a multi-robot approach to localization of an odor source [28]. An artificial whisker modeled on the Rat mystacial vibrissae has recently been demonstrated [56], such a sensor could be of particular value in dusty or smoky environments.

Actuation

Locomotion   The means of physical locomotion for a foraging robot can take many forms and clearly depends on the environment in which the robot is intended to operate. Ground robots typically use wheels, tracks or legs, although wheels are predominantly employed in proof-of-concept or demonstrator foraging robots. An introduction to the technology of robot mobility can be found in [63]. Flying robots (unmanned air vehicles – UAVs) are either fixed- or rotary-wing; for recent examples of work towards teams of flying robots see [13] (fixed-wing) and [51] (rotary-wing). Underwater robots (unmanned underwater vehicles – UUVs) generally use the same principles for propulsion as submersible remotely operated vehicles (ROVs), [70]. Whatever the means of locomotion important principles which apply to all foraging robots are that robot(s) must be able to (1) move with sufficient stability for the object detection sensors to be able to operate effectively and (2) position themselves with sufficient precision and stability to allow the object to be physically grabbed. These factors place high demands on a foraging robot's physical locomotion system, especially if the robot is required to operate in soft or unstable terrain.

Object Manipulation   The manipulation required of a foraging robot is clearly dependent on the form of the object and the way the object presents itself to the robot as it approaches. The majority of foraging experiments or demonstrations have simplified the problem of object manipulation by using objects that are, for instance, always the right way up (metal pucks or wooden sticks protruding from holes) so that a simple gripper mounted on the front of the robot is able to grasp the objects with reasonable reliability. However, in general a foraging robot would require the versatility of a robot arm (multi-axis manipulator) and general purpose gripper (hand) such that – with appropriate vision sensing – the robot can pick up the object regardless of its shape and orientation. This technology is well developed in tele-operated robots used for remote inspection and handling of dangerous materials or devices, see [62,66].

Communications

Communications is of fundamental importance to robot foraging. Only in the simplest case of a single robot foraging autonomously would communications be unnecessary. For single robot teleoperation radio communication between operator and robot is clearly an essential requirement. In multi-robot foraging robot-robot communication is frequently employed to improve multi-robot performance; all six axes of strategy in the taxonomy of Table 3: search, grabbing, transport, homing, recruitment and coordination may require some form of robot-robot communication. Arai et al. point out the important distinction between *explicit* and *implicit* communication [1].

Explicit Communication   Explicit communication applies when robots need to exchange information directly. The physical medium of communication is frequently (but not necessarily) radio, and wireless local area network (WLAN) technology is highly appropriate to terrestrial multi-robot systems, not least because a spatially distributed team of wireless networked robots naturally forms an *ad-hoc* network, which – providing the team maintains sufficient connectivity – allows any robot to communicate with any other via multiple hops, [69]. A method for linking wireless connectivity to locomotion in order to maintain connectivity is described in [52]; work that falls within the framework of *situated* communications proposed by Støy. Situated communication pertains when "both the physical properties of the signal that transfers the message and the content of the message contribute to its meaning" [65].

Implicit Communication   Implicit communication applies when robots communicate not directly but via the environment, also known as *stigmergic* communications. Thus one robot changes the environment and another senses the change and alters its behavior accordingly. Beckers et al., in one of the first demonstrations of self-organized multi-robot puck clustering, show that stigmergic communication alone can give rise to the desired overall group behavior [6]. However, in their study on multi-robot communication, Balch and Arkin show that while stigmergy may be sufficient to complete the task, direct communication can increase efficiency [4]. Trail following, in which a robot follows a short-lived trail left by other(s), is an example of implicit communication [59,60].

Control

From a control perspective the simplicity of the finite state machine for basic foraging, in Fig. 1, is deceptive. In principle, a very simple foraging robot could be built with basic hard-wired reflex actions such as obstacle avoidance and taxis toward the attractor object; such a robot is known as a Braitenberg vehicle, after his landmark work [11]. However, even simple foraging requires a complex set of competencies that would be impractical to implement except as a program on one or more embedded computers (microprocessors) in the robot. There are clearly many ways of building such a control program, but in the field of mobile robotics a number of robot control architectures have been defined. Such architectures mean that robot designers can approach the design of the control system in a principled way.
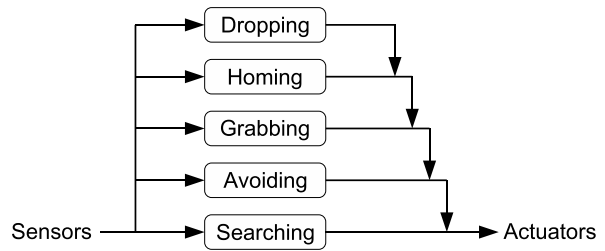
Foraging Robots, Figure 2
Subsumption control architecture for basic foraging

A widely adopted robot control architecture, first proposed and developed by Brooks, is the layered *subsumption* architecture known generically as behavior-based control [12]. Behavior-based control is particularly relevant to foraging robots since, like foraging, it is biologically inspired. In particular, as Arkin describes in [2], the principles of behavior-based control draw upon ethology – the study of animal behavior in the natural environment. Essentially behavior-based control replaces the functional modularity of earlier robot control architectures with task achieving modules, or behaviors. Matarić uses Brooks' behavior language (BL) to implement a set of basic behaviors for multi-robot foraging, as described in more detail below in Sect. Multi-Robot (Collective) Foraging, [46,47]. Refer to [14] for a comprehensive review of group control architectures for multi-robot systems.

Figure 2 shows the subsumption architecture for basic foraging (from Fig. 1), with the addition of *avoidance* for safely avoiding obstacles (including other robots in the case of multi-robot foraging). Each behavior runs in parallel and, when activated suppresses the output of the layer(s) below to take control of the robot's actuators.

Examples of Single Robot Foraging

A Soda-Can Collecting Robot   Possibly the first demonstration of autonomous single-robot foraging is Connell's soda-can collecting robot *Herbert*, [15]. Herbert's task was to wander safely through an office environment while searching for empty soda-cans; upon finding a soda-can Herbert would need to carefully grab the can with its hand and 2 degrees-of-freedom arm, then return to a waste basket to deposit it before resuming the search. Herbert therefore represents an implementation of exactly the basic foraging model of Fig. 1 and 2. However, two of the behaviors are not so straightforward. Both searching and homing require the robot to be able to navigate safely through a cluttered and unstructured 'real-world' environment, while grabbing is equally complex given the precision required

to safely reach and grab the soda-can. Thus Herbert's control system required around 40 low-level behaviors in order to realize foraging.

A Robot Predator   Arguably the first attempt to build a robot capable of actively predating for its own energy is the *Slugbot* of Holland and co-workers, [26,33]. The Slugbot (Fig. 3) solved the difficult problems of finding and collecting slugs in an energy efficient manner by means of, firstly, a long but light articulated arm which allows the robot to scan (in spiral fashion) a large area of ground for slugs without having to physically move the whole robot (which is much more costly in energy). Secondly, the special purpose gripper at the end of the arm is equipped with a camera which, by means of reflected red light and appropriate vision processing, is able to reliably detect and collect the slugs. An evolution of the Slugbot, the *Ecobot*, uses microbial fuel cell (MFC) technology to generate electrical energy directly from unrefined biomass [49].

Real-World Foraging Robots   Autonomous crop harvesting is an obvious real-world application of single-robot foraging. The Demeter system [57] has successfully demonstrated automated harvesting of cereal crops. Demeter uses a combination of GPS for coarse navigation and vision to sense the crop-line and hence fine-tune the harvester's steering to achieve a straight and even cut of the crop. The vision processing is challenging because it has to cope with a wide range of lighting conditions including – in conditions of bright sunlight – shadows cast onto the crop line by the harvester itself. In the field of automated agriculture a number of proof-of-concept robot harvesters have been demonstrated for cucumber, tomato and other fruits [34,35].

Robot lawn mowers and vacuum cleaners can similarly be regarded as simple forms of foraging robot and are notable because they are the only form of autonomous foraging robot in commercial production; in both cases the search task is simple because the grass, or dirt are not discrete objects to be found. The search problem for robot lawn movers and vacuum cleaners thus becomes the problem of energy efficient strategies for (1) safely covering the whole search space while avoiding obstacles and (2) homing and docking to a re-charging station. Robot lawn mowers typically require a wire to be installed at the perimeter of the lawn, thus delimiting the robot's working area, see [29] for a survey of commercial robot lawn mowers. A short account of the development of a vacuum cleaning robot is given in [58].

Although technically an *off-world* application, the planetary rover may be regarded as an instance of single-robot foraging in which the objects of interest (geological samples) are collected and analyzed within the robot. Autonomous sample-return robots would be true foragers [61]. The proof-of-concept robot astrobiologist *Zoë* forages – in effect – for evidence of life [67].

Multi-Robot (Collective) Foraging

Foraging is clearly a task that lends itself to multi-robot systems and, even if the task can be accomplished by a single robot, foraging should – with careful design of strategies for cooperation – benefit from multiple robots. Swarm intelligence is the study of natural and artificial systems of multiple agents in which there is no centralized or hierarchical command or control. Instead, global swarm behaviors emerge as a result of local interactions between the agents and each other, and between agents and the environment, [8]. Swarm robotics is concerned with the design of artificial robot swarms based upon the principles of swarm intelligence, thus control is completely distributed and robots, typically, must choose actions on the basis only of local sensing and communications, [7,16]. Swarm robotics is thus a sub-set of multi-robot systems and, in the taxonomy of Table 2 the strategy: coordination = self-organized.

Foraging Robots, Figure 3
The Slugbot: a proof-of-concept robot predator

Foraging is therefore a benchmark problem within swarm robotics, not least because of the strong cross-over between the study of self-organization in social insects and their artificial counterparts within swarm intelligence [19]. This section will therefore focus on examples of multi-robot foraging from within the field of swarm robotics. Three strategies for cooperation will be outlined: information sharing, physical cooperation and division of labor. The section will conclude with an outline of the problem of mathematical modeling of swarms of foraging robots.

Without Cooperation

Balch and co-workers describe the winners of the 'Office Cleanup Event' of the 1994 AAAI Mobile Robot Competition: a multi-robot trash-collecting team [5]. The robots were equipped with a vision system for recognition and distance estimation of trash items (primarily soda cans) and differentiation between trash items, wastebaskets and other robots. The robots did not communicate, but employed a collective strategy in which robots generate a strong repulsive force if they see each other while searching, and a weaker (but sufficient for avoidance) repulsive force while in other states; this had the effect of causing the robots to spread-out and hence search the environment more efficiently. Interestingly, Balch et al. found that the high density of trash in the competition favored a 'sit-and-spin' strategy to scan for trash items rather than the random wander approach of the original design. The FSM was essentially the same schema as shown in Fig. 1 except that since there could be a number of wastebaskets at unknown locations then 'homing' becomes 'search for nearest wastebasket'.

Strategies for Cooperation

Information Sharing   Matarić and Marjanovic provide what is believed to be the first description of a multi-robot foraging experiment using real (laboratory) robots in which there is no centralized control [47]. They describe a system of 20 identical 12″ 4-wheeled robots, equipped with: a two-pronged forklift for picking up, carrying and stacking metal pucks; proximity and bump sensors; radio transceivers for data communication and a sonar-based global positioning system. Matarić and Marjanovic extend the basic five state foraging model (wandering, grabbing, homing, dropping and avoiding), to introduce information sharing as follows. If a robot finds a puck it will grab it but also broadcast a radio message to tell other robots it has found a puck. Meanwhile, if another robot in the locale hears this message it will first enter state *tracking* to home in on the source of the message, then state *search-*
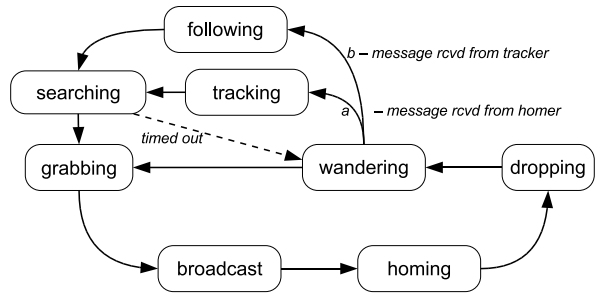
Foraging Robots, Figure 4
Finite State Machine for multi-robot foraging with recruitment – adapted from [47]

ing – a more localized form of wandering. The robot will return to wandering if it finds no puck within some time out period. Furthermore, while in state *tracking* a robot will also transmit a radio signal. If nearby robots hear this signal they will switch from wandering into *following* to pursue the tracking robot. Thus the tracking robot actively recruits additional robots as it seeks the original successful robot (a form of secondary swarming, [48]); when the tracking robot switches to searching its recruits will do the same. Figure 4 shows a simplified FSM. Within the taxonomy of Table 3 Strategy : recruitment = direct and indirect.

Physical Cooperation

Cooperative Grabbing   Consider the case of multi-robot foraging in which the object to be collected cannot be grabbed by a single robot working alone, in Table 3 this is Strategy: grabbing = cooperative. Ijspeert et al. describe an experiment in collaborative stick-pulling in which two robots must work together to pull a stick out of a hole [32,44]. Each *Khephera* robot is equipped with a gripper capable of grabbing and lifting the stick, but the hole containing the stick is too deep for one robot to be able to pull the stick out alone; one robot must pull the stick half-way then wait for another robot to grab the stick and lift it clear of the hole, see Fig. 5. Ijspeert and co-workers describe an elegant minimalist strategy which requires no direct communication between robots. If one robot finds a stick it will lift it and wait. If another finds the same stick it will also lift it, on sensing the force on the stick from the second robot the first robot will let go, hence allowing the second to complete the operation.

Cooperative Transport   Now consider the the situation in which the object to be collected is too large to be transported by a single robot, in Table 3 Strategy: transport =
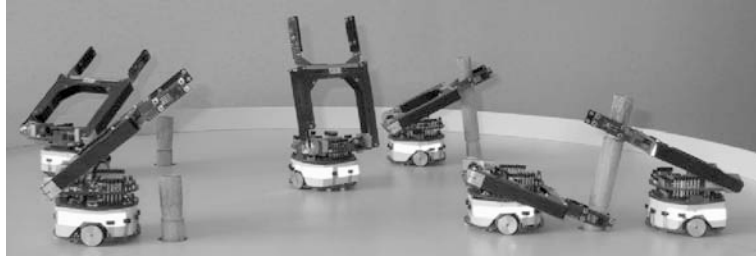
Foraging Robots, Figure 5
Cooperative grabbing: Khephera robots engaged in collective stick-pulling. With kind permission of A. Martinoli
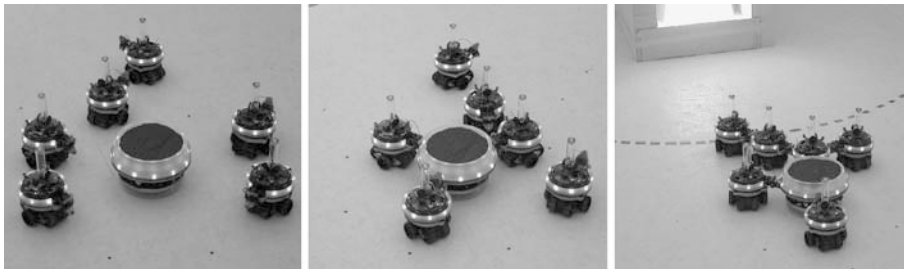
Foraging Robots, Figure 6
Cooperative transport by s-bots. (*Left*) s-bots approach the attractor object, (*middle*) s-bots start to grab the object, (*right*) s-bots collectively drag the object toward a beacon. With kind permission of M. Dorigo

cooperative. Parker describes the ALLIANCE group control architecture applied to an example of cooperative box-pushing by two robots [55].

Arguably the most accomplished demonstration of cooperative multi-robot foraging to date is within the *swarm-bot* project of Dorigo and co-workers [20]. The *s-bot* is a modular robot equipped with both a gripper and a gripping ring, which allows one robot to grip another [50]. Importantly, the robot is able to rotate its wheelbase independently of the gripping ring so that robots can grip each other at any arbitrary point on the circumference of the grip ring but then rotate and align their wheels in order to be able to move as a single unit (a *swarm-bot*). Groß et al. describe cooperative transport which uses visual signaling [27]. s-bots are attracted to the (large) object to be collected by its ring of red LEDs. The s-bot's LEDs are blue, but when an s-bot finds and grabs the attractor object it switches its LEDs to red. This increases the red light intensity to attract further s-bots which may grab either the object, or arbitrarily a robot already holding the object. The s-bots are then able to align and collectively move the object.

Division of Labor   In multi-robot foraging it is well know that overall performance (measured, for instance, as the number of objects foraged per robot in a given time interval), does not increase monotonically with increasing team size because of interference between robots (overcrowding), [4,25,38]. Division of labor in ant colonies has been well studied and in particular a response threshold model is described in [9] and [10]; in essence a threshold model means that an individual will engage in a task when the level of some task-associated stimulus exceeds its threshold.

For threshold-based multi-robot foraging with division of labor Fig. 7 shows a generalized finite state machine for each robot. In this foraging model the robot will not search endlessly. If the robot fails to find a food-item because, for instance, its searching time exceeds a maximum search time threshold T_s, or its energy level falls below a minimum energy threshold, then it will abandon its search and return home without food, shown as *failure*. Conversely *success* means food was found, grabbed and deposited. Note, however, that a robot might see a food-item but fail to grab it because, for instance, of competition with another robot for the same food-item. The robot now also has a *resting* state during which time it remains in the nest conserving energy. The robot will stop resting and leave home which might be according to some threshold criterion, such as its resting time exceeding the maximum rest time threshold T_r, or the overall nest energy falling below a given threshold.
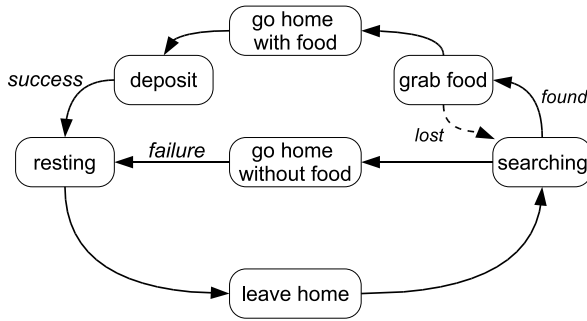
Foraging Robots, Figure 7
Finite State Machine for foraging with division of labor

Let us consider the special case of multi-robot foraging in which robots are foraging for their own energy. For an individual robot foraging costs energy, whereas resting conserves energy. We can formally express this as follows. Each robot consumes energy at A units per second while searching or retrieving and B units per second while resting, where $A > B$. Each discrete food item collected by a robot provides C units of energy to the swarm. The average food item retrieval time, is a function of the number of foraging robots x, and the density of food items in the environment, ρ, thus $t = f(x, \rho)$.

If there are N robots in the swarm, E_c is the energy consumed and E_r the energy retrieved, per second, by the swarm then

$$E_c = Ax + B(N - x) \tag{5}$$

$$E_r = Cx/t = \frac{Cx}{f(x, \rho)} . \tag{6}$$

The average energy income to the swarm, per second, is clearly the difference between the energy retrieved and the energy consumed,

$$E = E_r - E_c = \left(\frac{C}{f(x, \rho)} - (A - B) \right) x - BN \tag{7}$$

Equation 7 shows that maximizing the energy income to the swarm requires either increasing the number of foragers x or decreasing the average retrieval time $f(x, \rho)$. However, if we assume that the density of robots in the foraging area is high enough that interference between robots will occur then, for constant ρ, increasing x will increase $f(x, \rho)$. Therefore, for a given food density ρ there must be an optimal number of foragers x^*.

Krieger ad Billeter adopt a threshold-based approach to the allocation of robots to either foraging or resting; in their scheme each robot is allocated a fixed but randomly

chosen activation threshold [36]. While waiting in the nest each robot listens to a periodic radio broadcast indicating the nest-energy level E; when the nest-energy level falls below the robot's personal activation threshold then it leaves the nest and searches for food. It will continue to search until either its search is successful, or it runs out of energy and returns home; if its search is successful and it finds another food-item the robot will record its position (using odometry). On returning home the robot will radio its energy consumption thus allowing the nest to update its overall net energy. Krieger and Billeter show that team sizes of 3 or 6 robots perform better than 1 robot foraging alone, but larger teams of 9 or 12 robots perform less well. Additionally, they test a recruitment mechanism in which a robot signals to another robot waiting in the nest to follow it to the food source, in tandem. Krieger's approach is, strictly speaking, not fully distributed in that the nest is continuously tracking the average energy income E; the nest is – in effect – acting as a central coordinator.

Based upon the work of [17] on individual adaptation and division of labor in ants, Labella et al. describe a fully distributed approach that allows the swarm to self-organize to automatically find the optimal value x^* [37]. They propose a simple adaptive mechanism to change the ratio of foragers to resters by adjusting the probability of leaving home based upon successful retrieval of food. With reference to Fig. 7 the mechanism works as follows. Each robot will *leave home*, i. e. change state from resting to searching, with probability P_l. Each time the robot makes the *success* transition from deposit to resting, it increments its P_l value by a constant Δ multiplied by the number of consecutive successes, up to a maximum value $P_{\max}$. Conversely, if the robot's searching time is up, the transition *failure* in Fig. 7, it will decrement its P_l by Δ times the number of consecutive failures, down to minimum $P_{\min}$. Interestingly, trials with laboratory robots show that the same robots self-select as foragers or resters – the algorithm exploits minor mechanical differences that mean that some robots are better suited as foragers.

Recently Liu et al. have extended this fully distributed approach by introducing two additional adaptation rules [43]. As in the case of Labella et al. individual robots use internal cues (successful object retrieval), but Liu adds environmental cues (collisions with team mates while searching), and social cues (team mate success in object retrieval), to dynamically vary the time spent foraging or resting. Furthermore, Liu investigates the performance of a number of different adaptation strategies based on combinations of these three cues. The three cues increment or decrement the searching time and resting time thresholds T_s and T_r as follows (note that adjusting T_r is equivalent

Foraging Robots, Table 4
Foraging swarm strategy – cue combinations

	Internal cues	Social cues	Environment cues
S_1 (baseline)	×	×	×
S_2	✓	×	×
S_3	✓	✓	×
S_4	✓	✓	✓

to changing the probability of leaving the nest P_l):

1. Internal cues. If a robot successfully finds food it will reduce its own rest time T_r; conversely if the robot fails to find food it will increase its own rest time T_r.
2. Environment cues. If a robot collides with another robot while searching, it will reduce its T_s and increase its T_r times.
3. Social cues. When a robot returns to the nest it will communicate its food retrieval success or failure to the other robots in the nest. A successful retrieval will cause the other robots in the nest to increase their T_s and reduce their T_r times. Conversely failure will cause the other robots in the nest to reduce their T_s and increase their T_r times.

In order to evaluate the relative effect of these cues three different strategies are tested, against a baseline strategy of no cooperation. The strategy/cue combinations are detailed in Table 4.

Figures 8 and 9, from [43], show the number of active foragers and the instantaneous net swarm energy, respectively, for a swarm of eight robots. In both plots the food density in the environment is changed at time $t = 5000$ and again at time $t = 10000$ seconds. Figure 8 shows the swarm's ability to automatically adapt the number of active foragers in response to each of the step changes in food density. The baseline strategy S_1 shows of course that all eight robots are actively foraging continuously; S_2–S_4 however require fewer active foragers and strategies with social and environmental cues, S_3 and S_4, clearly show the best performance. Notice, firstly that the additional of social cues – communication between robots – significantly improves the rate at which the system can adapt the ratio of foragers to resters and, secondly, that the addition of environmental cues – collisions with other robots – brings only a marginal improvement. The rates of change of net swarm energy in Fig. 9 tell a similar story. Interestingly, however, we see very similar gradients for S_2–S_4 when the food density is high (on the RHS of the plot), but when the food density is medium or poor the rate of increase in net energy of strategies S_3 and S_4 is significantly better than

S_2. This result interestingly suggests that foraging robots benefit more from cooperation when food is scarce, than when food is plentiful.

Mathematical Modeling

A multi-robot system of foraging robots is typically a stochastic non-linear dynamical system and therefore challenging to mathematically model, but without such models any claims about the correctness of foraging algorithms are weak. Experiments in computer simulation or with real-robots (which provide in effect an 'embodied' simulation) allow limited exploration of the parameter space and can at best only provide weak inductive proof of correctness. Mathematical models on the other hand, allow analysis of the whole parameter space and discovery of optimal parameters. Ultimately, in real-world applications, validation of a foraging robot system for safety and dependability will require a range of formal approaches including mathematical modeling.

Martinoli and coworkers proposed a *microscopic* approach to study collective behavior of a swarm of robots engaged in cluster aggregation [45] and collaborative stick-pulling [32], in which a robot's interactions with other robots and the environment are modeled as a series of stochastic events, with probabilities determined by simple geometric considerations and systematic experiments with one or two real robots.

Lerman, Martinoli and co-workers have also developed the *macroscopic* approach, as widely used in physics, chemistry, biology and the social sciences, to directly describe the collective behavior of the robotic swarm. A class of macroscopic models have been used to study the effect of interference in a swarm of foraging robots [38] and collaborative stick-pulling [39,44]. A review of macroscopic models is given in [41]. More recently, Lerman et al. [40] successfully expanded the macroscopic probabilistic model to study dynamic task allocation in a group of robots engaged in a puck collecting task, in which the robots need to decide whether to pick up red or green pucks based on observed local information.

A Macroscopic Mathematical Model of Multi-Robot Foraging with Division of Labor   Recently Liu et al. have applied the macroscopic approach to develop a mathematical model for foraging with division of labor (as described above in Section "Division of Labor"), [42]. The finite state machine of Fig. 7 is extended in order to describe the probabilistic behavior of the whole swarm, resulting in a probabilistic finite state machine (PFSM). In Fig. 10 each state represents the average number of robots
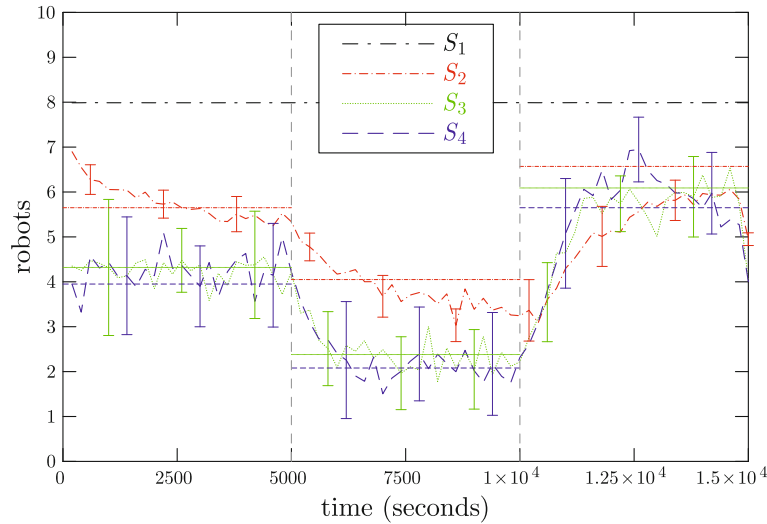
Foraging Robots, Figure 8
Number of foraging robots x in a foraging swarm of N = 8 robots with self-organized division of labor. S_1 is the baseline (no cooperation strategy); S_2, S_3 and S_4 are three different cooperation strategies (see Table 4). Food density changes from 0.03 (medium) to 0.015 (poor) at t = 5000, then from 0.015 (poor) to 0.045 (rich) at t = 10000. Each plot is the average of 10 runs
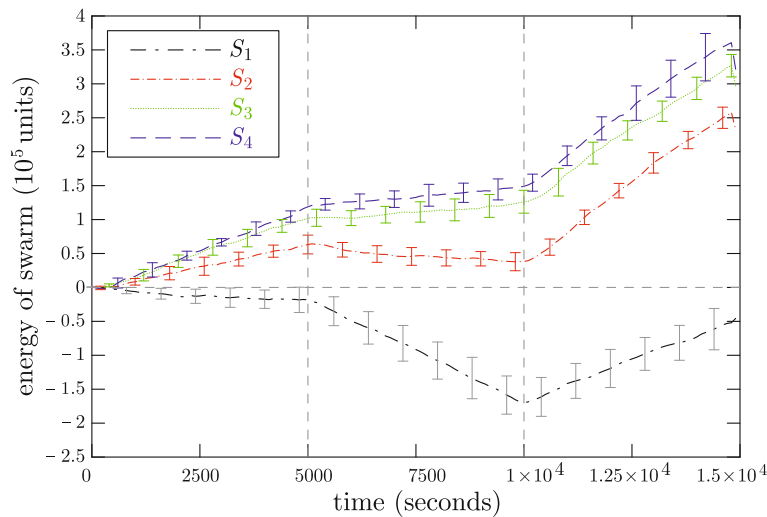
Foraging Robots, Figure 9
Instantaneous net energy E of a foraging swarm with self-organized division of labor. S_1 is the baseline (no cooperation strategy); S_2, S_3 and S_4 are three different cooperation strategies (see Table 4). Food density changes from 0.03 (medium) to 0.015 (poor) at t = 5000, then from 0.015 (poor) to 0.045 (rich) at t = 10000. Each plot is the average of 10 runs

in that state. The five basic states are S for *searching*, H for *homing*, G for *grabbing*, D for *depositing* and R for *resting*, and the average number of robots in each of these states is respectively N_S, N_H, N_G, N_D and N_R. τ_S, τ_H, τ_G, τ_D and τ_R represent the average times a robot will spend in each state before moving to the next state.

In each time step a robot in state S has probability γ_f of finding a food-item and moving to state G, in which it will

move towards the target food-item until it is close enough to grab it using the gripper. Once the robot successfully grabs the food-item it will move to state D, in which the robot moves back to the 'nest' carrying the food-item and deposits it. After the robot has unloaded the food-item it will rest in state R, for τ_R seconds and then move to S to resume *searching*. Meanwhile, if the robot in state S fails to find a food-item within time τ_S, it will move to state H, and
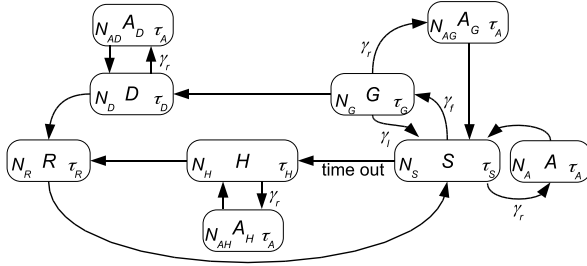
Foraging Robots, Figure 10
Probabilistic Finite State Machine (PFSM) for foraging with division of labor

Foraging Robots, Figure 11
Foraging environment showing 8 robots labeled *A — H*. The nest region is the *grey circle* with radius R_h at the center. Robot *A* is shown with its arc of vision in which it can sense food items; robots *C*, *E* and *F* have grabbed food items and are in the process of returning to the nest to deposit these. Food items, shown as *small squares*, 'grow' in order to maintain uniform density within the annular region between *circles* with radius R_{inner} and R_{outer}

return to the 'nest' to save energy or minimize interference with other robots. Because of competition among robots more than one robot may see the same food-item and thus move towards it at the same time; clearly only one of them can grab it, a robot in state G therefore has probability γ_l to lose sight of the food-item if it has already been grabbed by another robot, which in turn drives the robot back to state S to resume its search.

In foraging interference between robots because of overcrowding, competition for food-items or simply random collisions is a key aspect of the dynamics of foraging. Thus collision avoidance is modeled as follows. Robots in states S, G, D and H will move to *avoidance* states A, A_G, A_D and A_H respectively with probability γ_r, as shown in Fig. 10. The avoidance behavior then takes τ_A seconds to complete before the robot moves back to its previous state.

Constructing the mathematical model requires two further steps. Firstly, writing down a set of difference equations (DEs) describing the change in the average number of robots in each state from one time step to the next and, secondly, estimating the state transition probabilities. Expressing the PFSM as a set of DEs is relatively straightforward. For instance, the change in the average number of robots N_A in state A from time step k to $k+1$ is given as:

$$N_A(k+1) = N_A(k) + \gamma_r N_S(k) - \gamma_r N_S(k - T_A) \quad (8)$$

where $\gamma_r N_S(k)$ is the number of robots that move from the search to the avoidance state A and $\gamma_r N_S(k - T_A)$ is the number of robots that return to S from state A after time T_A (note T_A is τ_A discretized for time step duration Δt). The full set of DEs is given in [42]. Clearly, the total number of robots in the swarm remains constant from one time step to the next,

$$N = N_S(k) + N_R(k) + N_G(k) + N_D(k) + N_H(k) \\ + N_A(k) + N_{A_H}(k) + N_{A_G}(k) + N_{A_D}(k) \quad (9)$$

Estimating state transition probabilities can be challenging but if we simplify the environment by placing the 'nest' region at the center of a circular environment in which the food growing area is bounded by two concentric rings in a bounded arena, as shown in Fig. 11, then a purely geometrical approach can be used to estimate γ_f, γ_r and γ_l together with the average times for grabbing, depositing and homing τ_G, τ_D and τ_H. Clearly τ_R and τ_S are the design parameters we seek to optimize, while τ_A is determined by the physical design of the robot and its sensors.

Figure 12, from [42], plots the average number of robots, from both simulation and the mathematical model, in states *searching*, *resting* and *homing* for the swarm with $\tau_r = 80$. The average number of robots in each state predicted by the probabilistic model quickly settles to a constant value. In contrast, but as one would expect, the average number of robots from simulation oscillates over time but stays near the value predicted by the model.

Figure 13 compares the predicted value of net swarm energy from the mathematical model, with the measured value from simulation, for resting time parameter τ_r increasing from 0 to 200s. The two curves show, firstly a good match between measured and predicted curves
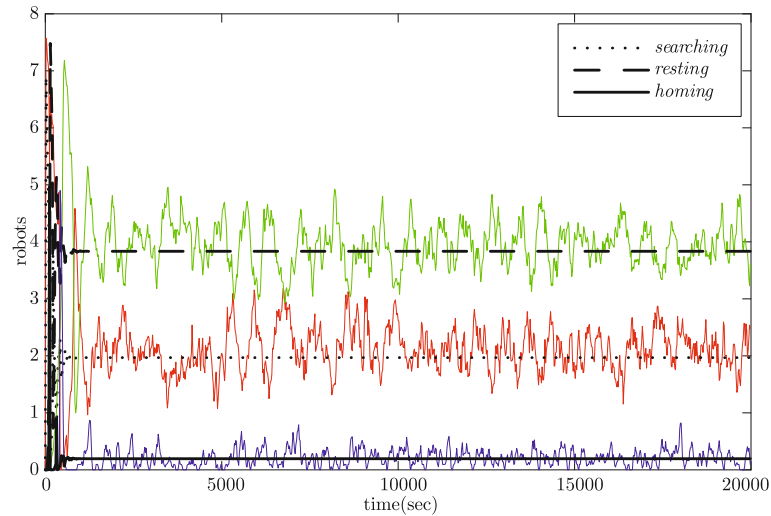
Foraging Robots, Figure 12
The number of robots in states *searching*, *resting* and *homing* for the swarm with $\tau_r = 80$ seconds. The horizontal black dashed lines are predicted by the mathematical model; colored graphs show the instantaneous number of robots measured from simulation
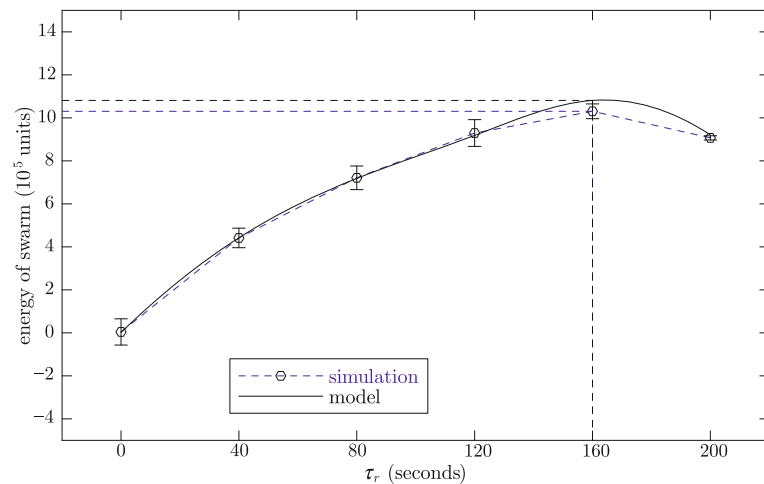
Foraging Robots, Figure 13
The net energy of the swarm for different values of the resting time parameter τ_r. The black curve is the prediction of the mathematical model; the *dashed curve* with error bars is measured from simulation

therefore validating the mathematical model and, secondly, that there is indeed an optimal value for τ_r (at about 160 seconds). We thus have confirmation that a mathematical model can be used to analyze the effect of individual parameters on the overall performance of collective foraging.

Future Directions

This article has defined robot foraging, set out a taxonomy and described both the development and state-of-

the-art in robot foraging. Although the principles of robot foraging are well understood, the engineering realization of those principles remains a research problem. Consider multi-robot cooperative robot foraging. Separate aspects have been thoroughly researched and demonstrated, and a number of exemplars have been described in this article. However, to date there has been no demonstration of autonomous multi-robot foraging which integrates self-organized cooperative search, object manipulation and transport in unknown or unstructured real-world environments. Such a demonstration would be a precursor to

a number of compelling real-world applications including search and rescue, toxic waste cleanup or foraging for recycling of materials.

The future directions for foraging robots lie along two separate axes. One axis is the continuing investigation and discovery of foraging algorithms – especially those which seek to mimic biologically inspired principles of self-organization. The other axis is the real-world application of foraging robots and it is here that many key challenges and future directions are to be found. Foraging robot teams are complex systems and the key challenges are in *systems integration and engineering*, which would need to address:

1. Principled design and test methodologies for self-organized multi-robot foraging robot systems.
2. Rigorous methodologies and tools for the specification, analysis and modeling of multi-robot foraging robot systems.
3. Agreed metrics and quantitative benchmarks to allow comparative evaluation of different approaches and systems.
4. Tools and methodologies for provable multi-robot foraging stability, safety and dependability [23,68].

Acknowledgments

Bibliography

Primary Literature

1. Arai T, Pagello E, Parker L (2002) Guest editorial: Advances in multirobot systems. IEEE Trans Robotics Autom 18:655–661
2. Arkin RC (1998) Behaviour-Based Robotics. MIT Press, Cambridge
3. Balch T (2002) Taxonomies of multirobot task and reward. In: Balch T, Parker LE (eds) Robot Teams. A K Peters, Wellesley, pp 23–35
4. Balch T, Arkin RC (1994) Communication in reactive multiagent robotic systems. Auton Robots 1:1–25
5. Balch T, Boone G, Collins T, Forbes H, MacKenzie D, Santamaria J (1995) Io, Ganymede and Callisto: A multiagent robot trash-collecting team. AI Magazine 16(2):39–53
6. Beckers R, Holland OE, Deneubourg JL (1994) From local actions to global tasks: Stigmergy and collective robotics. In: Artificial Life IV. MIT Press, Cambridge, pp 181–189
7. Beni G (2005) From swarm intelligence to swarm robotics. In: Şahin E, Spears W (eds) Swarm Robotics Workshop: State-of-the-art Survey, number 3342. Springer, Berlin, pp 1–9
8. Bonabeau E, Dorigo M, Theraulaz G (1999) Swarm Intelligence – From Natural to Artificial Systems. Oxford Univ Press, Oxford
9. Bonabeau E, Theraulaz C, Deneubourg JL (1996) Quantitative study of the fixed threshold model for the regulation of division of labor in insect societies. In: Proceedings of the Royal Society of London, Series B Biological Sciences 263:1565–1569
10. Bonabeau E, Theraulaz G, Deneubourg JL (1998) Fixed response thresholds and the regulation of division of labour in insect societies. Bull Math Biol 60:753–807
11. Braitenberg V (1984) Vehicles – Experiments in Synthetic Psychology. MIT Press, Cambridge
12. Brooks RA (1986) A robust layered control system for a mobile robot. J Robotics Autom 2:14–23
13. Bryson MT, Sukkarieh S (2007) Decentralised trajectory control for multi-UAV SLAM. In: 4th International Symposium on Mechatronics and its Applications (ISMA '07), Sharjah, United Arab Emirates, March 2007
14. Cao YU, Fukunaga AS, Kahng AB (1997) Cooperative mobile robotics: Antecedents and directions. Auton Robots 4:1–23
15. Connell JH (1990) Minimalist Mobile Robotics: A colony-style architecture for an artificial creature. Academic Press, San Diego
16. Şahin E (2005) Swarm robotics: From sources of inspiration to domains of application. In: Şahin E, Spears W (eds) Swarm Robotics Workshop: State-of-the-art Survey, number 3342. Lecture Notes in Computer Science. Springer, Berlin, pp 10–20
17. Deneubourg JL, Goss S, Pasteels JM, Fresneau D, Lachaud JP (1987) Self-organization mechanisms in ant societies (ii): learning in foraging and division of labour. Experientia Suppl 54:177–196
18. Dissanayake MWMG, Newman PM, Durrant-Whyte HF, Clark S, Csorba M (2001) A solution to the simultaneous localization and map building (SLAM) problem. IEEE Trans Robotics Autom 17(3):229–241
19. Dorigo M, Birattari M (2007) Swarm intelligence. Scholarpedia 2(9):1462
20. Dorigo M, Tuci E, Groß T, Trianni V, Labella TH, Nouyan S, Ampatzis C (2005) The SWARM-BOT project. In: Şahin E, Spears W (eds) Swarm Robotics Workshop: State-of-the-art Survey, number 3342. Lecture Notes in Computer Science. Springer, Berlin, pp 31–44
21. Dudek G, Jenkin M, Milios E, Wilkes D (1996) A taxonomy for multi-agent robotics. Auton Robots 3:375–397
22. Everett HR (1995) Sensors for mobile robots: Theory and applications. AK Peters, Wellesley
23. Gazi V, Passino KM (2004) Stability analysis of social foraging swarms. IEEE Trans Syst Man Cybernetics – Part B: Cybernetics 34(1):539–557
24. Gerkey BP, Matarić MJ (2004) A formal analysis and taxonomy of task allocation in multi-robot systems. Int J Robot Res 23(9):939–954
25. Goldberg D, Matarić MJ (1997) Interference as a tool for designing and evaluating multi-robot controllers. In: Proc 14th National conference on Artificial Intelligence (AAAI-97), Providence, July 1997. MIT Press, Cambridge, pp 637–642
26. Greenman J, Holland OE, Kelly I, Kendall K, McFarland D, Melhuish CR (2003) Towards robot autonomy in the natural world: A robot in predator's clothing. Mechatronics 13(3):195–228
27. Groß R, Tuci E, Dorigo M, Bonani M, Mondada F (2006) Object transport by modular robots that self-assemble. In: Proc IEEE International Conference on Robotics and Automation, Orlando, May 2006, pp 2558–2564
28. Hayes AT, Martinoli A, Goodman RMF (2002) Distributed odor

source localization. IEEE Sensors, Special Issue on Artificial Olfaction 2(3):260–271

29. Hicks RW, Hall EL (2000) A survey of robot lawn mowers. In: Casasent DP (ed) Proc SPIE Intelligent Robots and Computer Vision XIX: Algorithms, Techniques, and Active Vision, vol 4197. SPIE, Bellingham, pp 262–269

30. Hölldobler B, Wilson EO (1990) The Ants. Harvard University Press, Cambridge

31. Horn BKP (1986) Robot Vision. MIT Press, Cambridge

32. Ijspeert AJ, Martinoli A, Billard A, Gambardella LM (2001) Collaboration through the exploitation of local interactions in autonomous collective robotics: The stick pulling experiment. Auton Robots 11(2):149–171

33. Kelly I, Holland OE, Melhuish CR (2000) Slugbot: a robotic predator in the natural world. In: 5th Symposium on Artificial Life and Robotics (AROB2000), Oita, January 2000

34. Kondo N, Monta M, Shibano Y, Mohri K (1993) Basic mechanism of robot adapted to physical properties of tomato plant. In: Proc International Conference for Agricultural Machinery and Process Engineering, Seoul, October 1993, vol 3, pp 840–849. The Korean Society for Agricultural Machinery

35. Kondo N, Nakamura M, Monta M, Shibano Y, Mohri K, Arima S (1994) Visual sensor for cucumber harvesting robot. In: Proceedings of the Food Processing Automation Conference, Orlando, February 1994, pp 461–470

36. Krieger M, Billeter JB (2000) The call of duty: Self-organised task allocation in a population of up to twelve mobile robots. J Robotics Auton Syst 30:65–84

37. Labella TH, Dorigo M, Deneubourg JL (2006) Division of labour in a group of robots inspired by ants' foraging behaviour. ACM Trans Auton Adapt Syst 1(1):4–25

38. Lerman K (2002) Mathematical model of foraging in a group of robots: Effect of interference. Auton Robots 13(2):127–141

39. Lerman K, Galstyan A, Martinoli A, Ijspeert AJ (2002) A macroscopic analytical model of collaboration in distributed robotic systems. Artif Life 7:375–393

40. Lerman K, Jones C, Galstyan A, Matarić MJ (2006) Analysis of dynamic task allocation in multi-robot systems. Int J Robot Res 25(3):225–242

41. Lerman K, Martinoli A, Galstyan A (2005) A review of probabilistic macroscopic models for swarm robotic systems. In: Şahin E, Spears W (eds) Swarm Robotics Workshop: State-of-the-art Survey, number 3342. Springer, Berlin, pp 143–152

42. Liu W, Winfield AFT, Sa J (2007) Modelling swarm robotic systems: A case study in collective foraging. In: Towards Autonomous Robotic Systems (TAROS 07), pp 25–32, Aberystwyth, September 2007

43. Liu W, Winfield AFT, Sa J, Chen J, Dou L (2007) Towards energy optimisation: Emergent task allocation in a swarm of foraging robots. Adapt Behav 15(3):289–305

44. Martinoli A, Easton K, Agassounon W (2004) Modeling swarm robotic systems: A case study in collaborative distributed manipulation. Int J Robot Res, Special Issue on Experimental Robotics 23(4):415–436

45. Martinoli A, Ijspeert AJ, Gambardella LM (1999) A probabilistic model for understanding and comparing collective aggregation mechanisms. In: Proc Euro Conf on Artificial Life ECAL'99, Lausanne, September 1999, pp 575–584

46. Matarić MJ (1992) Designing emergent behaviours: From local interactions to collective intelligence. In: From Animals To Animats (SAB-92). MIT Press, Cambridge, pp 432–441

47. Matarić MJ, Marjanovic MJ (1993) Synthesizing complex behaviors by composing simple primitives. In: Proc Self Organization and Life, From Simple Rules to Global Complexity, European Conference on Artificial Life (ECAL-93), pp 698–707, Brussels, May (1993)

48. Melhuish C (1999) Employing secondary swarming with small scale robots: a biologically inspired collective approach. In: Proc of the 2nd Int Conf on Climbing & Walking Robots CLAWAR, Portsmouth, September 1999

49. Melhuish C, Ieropoulos I, Greenman J, Horsfield I (2006) Energetically autonomous robots: Food for thought. Auton Robots 21(3):187–198

50. Mondada F, Gambardella LM, Floreano D, Nolfi S, Deneubourg JL, Dorigo M (2005) The cooperation of Swarm-bots: Physical interactions in collective robotics. IEEE Robotics Autom Mag 12(2):21–28

51. De Nardi R, Holland OE (2007) Ultraswarm: A further step towards a flock of miniature helicopters. In: Second International Workshop on Swarm Robotics at SAB (2006), vol 4433. Springer, Heidelberg, pp 116–128

52. Nembrini J, Winfield AFT, Melhuish C (2002) Minimalist coherent swarming of wireless networked autonomous mobile robots. In: From Animals to Animats SAB'02. MIT Press, Cambridge, pp 373–382

53. Oster GF, Wilson EO (1978) Caste and Ecology in the Social Insects. Princeton University Press, Priceton

54. Østergaard EH, Sukhatme GS, Matarić MJ (2001) Emergent bucket brigading: A simple mechanism for improving performance in multi-robot constrained-space foraging tasks. In: Proc Int Conf on Autonomous Agents, Montreal, May 2001

55. Parker LE (1994) ALLIANCE: an architecture for fault tolerant, cooperative control of heterogeneous mobile robots. In: Proc IEEE/RSJ International Conference on Intelligent Robots and Systems, Munich, September 1994, pp 776–783

56. Pearson MJ, Pipe AG, Melhuish CR, Mitchinson B, Prescott TJ (2007) Whiskerbot: A robotic active touch system modeled on the rat whisker sensory system. Adapt Behav 15:223–240

57. Pilarski T, Happold M, Pangels H, Ollis M, Fitzpatrick K, Stentz A (1999) The Demeter system for automated harvesting. In: Proceedings of the 8th International Topical Meeting on Robotics and Remote Systems, Pittsberg, April 1999

58. Rooks B (2001) Robots reach the home floor. Ind Robot 28(1):27–28

59. Russell A (1993) Mobile robot guidance using a short-lived heat trail. Robotica 11:427–431

60. Russell A (1995) Laying and sensing odor markings as a strategy for assisting mobile robot navigation tasks. IEEE Robotics Autom Mag 2(3):3–9

61. Schenker PS, Huntsberger TL, Pirjanian P, Baumgartner ET, Tunstel E (2003) Planetary rover developments supporting mars exploration, sample return and future human-robotic colonization. Auton Robots 14(2–3):103–126

62. Schilling T (ed) (2000) Telerobotic Applications. Professional Engineering Publishing, London

63. Siegwart RY, Nourbakhsh IR (2004) Introduction to Autonomous Mobile Robots (Intelligent Robotics and Autonomous Agents). Bradford Books, Cambridge

64. Spero DJ, Jarvis RA (2002) Path planning for a mobile robot in a rough terrain environment. In: Third International Workshop on Robot Motion and Control, pp 417–422, Bukowy Dworek, November 2001

65. Støy K (2001) Using situated communication in distributed autonomous mobile robotics. In: 7th Scandinavian Conf on AI. Odense, February 2001, pp 44–52

66. Vertut J, Coiffet P (1986) Teleoperation and Robotics. Prentice Hall, Englewood Cliffs

67. Wettergreen D, Cabrol N, Baskaran V, Calderón F, Heys S, Jonak D, Lüders A, Pane D, Smith T, Teza J, Tompkins P, Villa D, Williams C, Wagner M (2005) Second experiments in the robotic investigation of life in the Atacama desert of Chile. In: Proc International Symposium on Artificial Intelligence, Robotics and Automation in Space, Munich, September 2005

68. Winfield AFT, Harper CJ, Nembrini J (2005) Towards dependable swarms and a new discipline of swarm engineering. In: Şahin E, Spears W (eds) Swarm Robotics Workshop: State-of-the-art Survey, number 3342. Springer, Berlin, pp 126–142

69. Winfield AFT, Holland OE (2000) The application of wireless local area network technology to the control of mobile robots. Microprocess Microsyst 23:597–607

70. Yuh J (1996) Underwater Robots. Kluwer Academic, Boston

Books and Reviews

Balch T, Parker LE (eds) (2002) Robot Teams: From diversity to polymorphism. AK Peters, Wellesley

Bekey GA (2005) Autonomous Robots: From Biological Inspiration to Implementation and Control. MIT Press, Cambridge

Brooks RA (1999) Cambrian Intelligence: The Early History of the New AI. MIT Press, Cambridge

Melhuish CR (2001) Strategies for Collective Minimalist Robotics. Professional Engineering Publishing, London

Nehmzow U (2003) Mobile Robotics: A Practical Introduction. Springer, New York

Fractal Geometry, A Brief Introduction to

Armin Bunde[1], Shlomo Havlin[2]
[1] Institut für Theoretische Physik, Gießen, Germany
[2] Institute of Theoretical Physics, Bar-Ilan-University, Ramat Gan, Israel

Article Outline

Definition of the Subject

In this chapter we present some definitions related to the fractal concept as well as several methods for calculating the fractal dimension and other relevant exponents. The purpose is to introduce the reader to the basic properties of fractals and self-affine structures so that this book will be self contained. We do not give references to most of the original works, but, we refer mostly to books and reviews on fractal geometry where the original references can be found.

Fractal geometry is a mathematical tool for dealing with complex systems that have no characteristic length scale. A well-known example is the shape of a coastline. When we see two pictures of a coastline on two different scales, with 1 cm corresponding for example to 0.1 km or 10 km, we cannot tell which scale belongs to which picture: both look the same, and this features characterizes also many other geographical patterns like rivers, cracks, mountains, and clouds. This means that the coastline is scale invariant or, equivalently, has no characteristic length scale. Another example are financial records. When looking at a daily, monthly or annual record, one cannot tell the difference. They all look the same.

Scale-invariant systems are usually characterized by noninteger ("fractal") dimensions. The notion of noninteger dimensions and several basic properties of fractal objects were studied as long ago as the last century by Georg Cantor, Giuseppe Peano, and David Hilbert, and in the beginning of this century by Helge von Koch, Waclaw Sierpinski, Gaston Julia, and Felix Hausdorff. Even earlier traces of this concept can be found in the study of arithmetic-geometric averages by Carl Friedrich Gauss about 200 years ago and in the artwork of Albrecht Dürer (see Fig. 1) about 500 years ago. Georg Friedrich Lichtenberg discovered, about 230 years ago, fractal discharge patterns. He was the first to describe the observed self-similarity of the patterns: A part looks like the whole. Benoit Mandelbrot [1] showed the relevance of fractal geometry to many systems in nature and presented many important features of fractals. For further books and reviews on fractals see [2,3,4,5,6,7,8,9,10,11,12,13,14,15,16].

Before introducing the concept of fractal dimension, we should like to remind the reader of the concept of dimension in regular systems. It is well known that in regular systems (with uniform density) such as long wires, large thin plates, or large filled cubes, the dimension d characterizes how the mass $M(L)$ changes with the linear size L
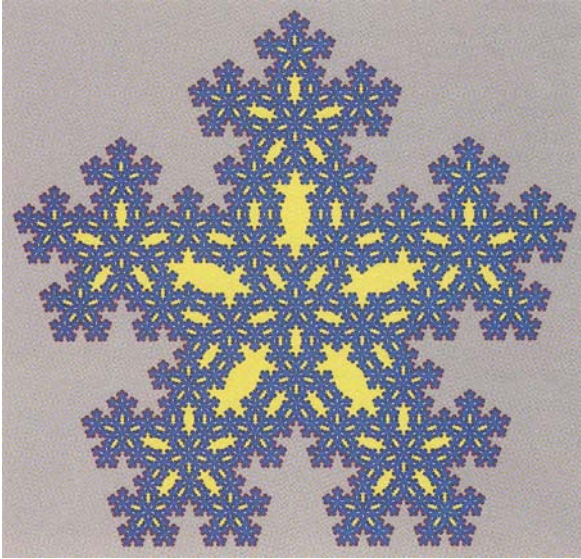
Fractal Geometry, A Brief Introduction to, Figure 1
The Dürer pentagon after five iterations. For the generating rule, see Fig. 8. The Dürer pentagon is in *blue*, its external perimeter is in *red*, Courtesy of M. Meyer

$$M(L/2) = \tfrac{1}{2} M(L) \qquad M(L/2) = \tfrac{1}{4} M(L) \qquad M(L/2) = \tfrac{1}{8} M(L)$$
$$d = 1 \qquad\qquad d = 2 \qquad\qquad d = 3$$

Fractal Geometry, A Brief Introduction to, Figure 2
Examples of regular systems with dimensions $d = 1$, $d = 2$, and $d = 3$

of the system. If we consider a smaller part of the system of linear size bL ($b < 1$), then $M(bL)$ is decreased by a factor of b^d, i. e.,

$$M(bL) = b^d M(L) . \qquad (1)$$

The solution of the functional equation (1) is simply $M(L) = AL^d$. For the long wire the mass changes linearly with b, i. e., $d = 1$. For the thin plates we obtain $d = 2$, and for the cubes $d = 3$; see Fig. 2.

Next we consider fractal objects. Here we distinguish between deterministic and random fractals. Deterministic fractals are generated iteratively in a deterministic way, while random fractals are generated using a stochastic process. Although fractal structures in nature are random, it is useful to study deterministic fractals where the fractal

properties can be determined exactly. By studying deterministic fractals one can gain also insight into the fractal properties of random fractals, which usually cannot be treated rigorously.

Deterministic Fractals

In this section, we describe several examples of deterministic fractals and use them to introduce useful fractal concepts such as fractal and chemical dimension, self similarity, ramification, and fractal substructures (minimum path, external perimeter, backbone, and red bonds).

The Koch Curve

One of the most common deterministic fractals is the Koch curve. Figure 3 shows the first $n = 4$ iterations of this fractal curve. By each iteration the length of the curve is increased by a factor of 4/3. The mathematical fractal is defined in the limit of infinite iterations, $n \to \infty$, where the total length of the curve approaches infinity.

The dimension of the curve can be obtained as for regular objects. From Fig. 3 we notice that, if we decrease the linear size by a factor of $b = 1/3$, the total length (mass) of the curve decreases by a factor of 1/4, i. e.,

$$M\left(\tfrac{1}{3} L\right) = \tfrac{1}{4} M(L) . \qquad (2)$$

Fractal Geometry, A Brief Introduction to, Figure 3
The first iterations of the Koch curve. The fractal dimension of the Koch curve is $d_f = \log 4 / \log 3$

This feature is very different from regular curves, where the length of the object decreases proportional to the linear scale. In order to satisfy Eqs. (1) and (2) we are led to introduce a *noninteger* dimension, satisfying $1/4 = (1/3)^d$, i. e., $d = \log 4/\log 3$. For such non-integer dimensions Mandelbrot coined the name "fractal dimension" and those objects described by a fractal dimension are called fractals. Thus, to include fractal structures, Eq. (1) is generalized by

$$M(bL) = b^{d_f} M(L), \qquad (3)$$

and

$$M(L) = AL^{d_f}, \qquad (4)$$

where d_f is the fractal dimension.

When generating the Koch curve and calculating d_f, we observe the striking property of fractals – the property of *self-similarity*. If we examine the Koch curve, we notice that there is a central object in the figure that is reminiscent of a snowman. To the right and left of this central snowman there are two other snowmen, each being an exact reproduction, only smaller by a factor of 1/3. Each of the smaller snowmen has again still smaller copies (by 1/3) of itself to the right and to the left, etc. Now, if we take any such triplet of snowmen (consisting of $1/3^m$ of the curve), for any m, and magnify it by 3^m, we will obtain exactly the original Koch curve. This property of self-similarity or scale invariance is the basic feature of all deterministic and random fractals: if we take a part of a fractal and magnify it by the same magnification factor in all directions, the magnified picture cannot be distinguished from the original.

For the Koch curve as well as for all deterministic fractals generated iteratively, Eqs. (3) and (4) are of course valid only for length scales L below the linear size L_0 of the whole curve (see Fig. 3). If the number of iterations n is finite, then Eqs. (3) and (4) are valid only above a lower cut off length L_{min}, $L_{min} = L_0/3^n$ for the Koch curve. Hence, for a finite number of iterations, there exist two cut-off length scales in the system, an upper cut-off $L_{max} = L_0$ representing the total linear size of the fractal, and a lower cut-off L_{min}. This feature of having two characteristic cut-off lengths is shared by all fractals in nature.

An interesting modification of the Koch curve is shown in Fig. 4, which demonstrates that the *chemical distance* is an important concept for describing structural properties of fractals (for a review see, for example, [16] and Chap. 2 in [13]). The chemical distance ℓ is defined as shortest path on the fractal between two sites of the fractal. In analogy to the fractal dimension d_f that characterizes how the mass of a fractal scales with (air) distance L, we
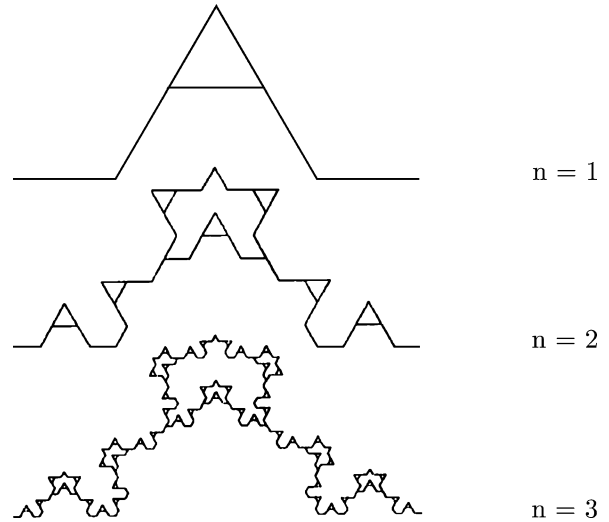
Fractal Geometry, A Brief Introduction to, Figure 4
The first iterations of a modified Koch curve, which has a nontrivial chemical distance metric

introduce the *chemical dimension* d_ℓ in order to characterize how the mass scales with the chemical distance ℓ,

$$M(b\ell) = b^{d_\ell} M(\ell), \quad \text{or} \quad M(\ell) = B\ell^{d_\ell}. \qquad (5)$$

From Fig. 4 we see that if we reduce ℓ by a factor of 5, the mass of the fractal within the reduced chemical distance is reduced by a factor of 7, i. e., $M(1/5\,\ell) = 1/7\,M(\ell)$, yielding $d_\ell = \log 7/\log 5 \cong 1.209$. Note that the chemical dimension is smaller than the fractal dimension $d_f = \log 7/\log 4 \cong 1.404$, which follows from $M(1/4\,L) = 1/7\,M(L)$.

The structure of the shortest path between two sites represents an interesting fractal by itself. By definition, the length of the path is the chemical distance ℓ, and the *fractal dimension of the shortest path*, d_{min}, characterizes how ℓ scales with (air) distance L. Using Eqs. (4) and (5), we obtain

$$\ell \sim L^{d_f/d_\ell} \equiv L^{d_{min}}, \qquad (6)$$

from which follows $d_{min} = d_f/d_\ell$. For our example we find that $d_{min} = \log 5/\log 4 \cong 1.161$. For the Koch curve, as well as for any linear fractal, one simply has $d_\ell = 1$ and hence $d_{min} = d_f$. Since, by definition, $d_{min} \geq 1$, it follows that $d_\ell \leq d_f$ for all fractals.

The Sierpinski Gasket, Carpet, and Sponge

Next we discuss the Sierpinski fractal family: the "gasket", the "carpet", and the "sponge".
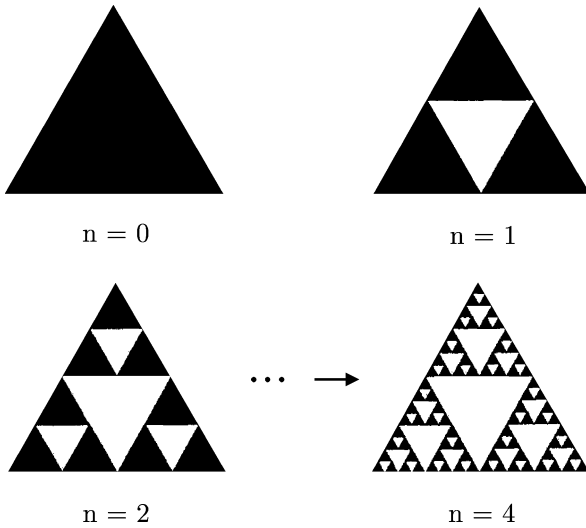
The Sierpinski gasket. The fractal dimension of the Sierpinski gasket is $d_f = \log 3/\log 2$

Fractal Geometry, A Brief Introduction to, Figure 6
A Sierpinski carpet with $n = 5$ and $k = 9$. The fractal dimension of this structure is $d_f = \log 16/\log 5$

The Sierpinski Gasket   The Sierpinski gasket is generated by dividing a full triangle into four smaller triangles and removing the central triangle (see Fig. 5). In the following iterations, this procedure is repeated by dividing each of the remaining triangles into four smaller triangles and removing the central triangles.

To obtain the fractal dimension, we consider the mass of the gasket within a linear size L and compare it with the mass within $1/2\, L$. Since $M(1/2\, L) = 1/3\, M(L)$, we have $d_f = \log 3/\log 2 \cong 1.585$. It is easy to see that $d_\ell = d_f$ and $d_{min} = 1$.

The Sierpinski Carpet   The Sierpinski carpet is generated in close analogy to the Sierpinski gasket.

Instead of starting with a full triangle, we start with a full square, which we divide into n^2 equal squares. Out of these squares we choose k squares and remove them. In the next iteration, we repeat this procedure by dividing each of the small squares left into n^2 smaller squares and removing those k squares that are located at the same positions as in the first iteration. This procedure is repeated again and again.

Figure 6 shows the Sierpinski carpet for $n = 5$ and the specific choice of $k = 9$. It is clear that the k squares can be chosen in many different ways, and the fractal structures will all look very different. However, since $M(1/n\, L) = 1/(n^2 - k)\, M(L)$ it follows that $d_f = \log(n^2 - k)/\log n$, irrespective of the way the k squares are chosen. Similarly to the gasket, we have $d_\ell = d_f$ and hence $d_{min} = 1$.

In contrast, the *external perimeter* ("hull", see also Fig. 1) of the carpet and its fractal dimension d_h depend strongly on the way the squares are chosen. The hull consists of those sites of the cluster, which are adjacent to empty sites and are connected with infinity via empty sites. In our example, see Fig. 6, the hull is a fractal with the fractal dimension $d_h = \log 9/\log 5 \cong 1.365$. On the other hand, if a Sierpinski gasket is constructed with the $k = 9$ squares chosen from the center, the external perimeter stays smooth and $d_h = 1$.

Although the rules for generating the Sierpinski gasket and carpet are quite similar, the resulting fractal structures belong to two different classes, to *finitely ramified* and *infinitely ramified* fractals. A fractal is called finitely ramified if any bounded subset of the fractal can be isolated by cutting a *finite* number of bonds or sites. The Sierpinski gasket and the Koch curve are finitely ramified, while the Sierpinski carpet is infinitely ramified. For finitely ramified fractals like the Sierpinski gasket many physical properties, such as conductivity and vibrational excitations, can be calculated exactly. These exact solutions help to provide insight onto the anomalous behavior of physical properties on fractals, as was shown in Chap. 3 in [13].

The Sierpinski Sponge   The Sierpinski sponge shown in Fig. 7 is constructed by starting from a cube, subdividing it into $3 \times 3 \times 3 = 27$ smaller cubes, and taking out the central small cube and its six nearest neighbor cubes. Each of the remaining 20 small cubes is processed in the same way, and the whole procedure is iterated ad infinitum. After each iteration, the volume of the sponge is reduced by a factor of 20/27, while the total surface area increases. In the limit of infinite iterations, the surface area is infinite,
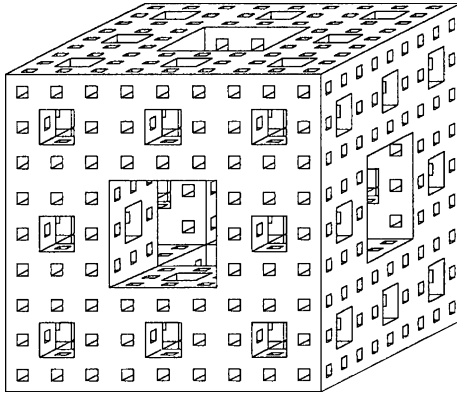
Fractal Geometry, A Brief Introduction to, Figure 7
The Sierpinski sponge (third iteration). The fractal dimension of the Sierpinski sponge is $d_f = \log 20/\log 3$
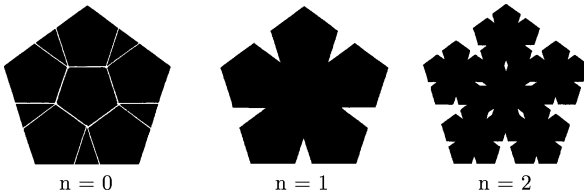
Fractal Geometry, A Brief Introduction to, Figure 8
The first iterations of the Dürer pentagon. The fractal dimension of the Dürer pentagon is $d_f = \log 6/\log(1 + g)$

Fractal Geometry, A Brief Introduction to, Figure 9
The first iterations of the David fractal. The fractal dimension of the David fractal is $d_f = \log 6/\log 3$

while the volume vanishes. Since $M(1/3\,L) = 1/20\,M(L)$, the fractal dimension is $d_f = \log 20/\log 3 \cong 2.727$. We leave it to the reader to prove that both the fractal dimension d_h of the external surface and the chemical dimension d_ℓ is the same as the fractal dimension d_f.

Modification of the Sierpinski sponge, in analogy to the modifications of the carpet can lead to fractals, where the fractal dimension of the hull, d_h, differs from d_f.

The Dürer Pentagon

Five-hundred years ago the artist Albrecht Dürer designed a fractal based on regular pentagons, where in each iteration each pentagon is divided into six smaller pentagons and five isosceles triangles, and the triangles are removed (see Fig. 8). In each triangle, the ratio of the larger side to the smaller side is the famous *proportio divina* or golden ratio, $g \equiv 1/(2\cos 72°) \equiv (1 + \sqrt{5})/2$. Hence, in each iteration the sides of the pentagons are reduced by $1 + g$. Since $M(L/(1 + g)) = 1/6\,M(L)$, the fractal dimension of the Dürer pentagon is $d_f = \log 6/\log(1 + g) \cong 1.862$. The external perimeter of the fractal (see Fig. 1) forms a fractal curve with $d_h = \log 4/\log(1 + g)$.

A nice modification of the Dürer pentagon is a fractal based on regular hexagons, where in each iteration one hexagon is divided into six smaller hexagons, six equilateral triangles, and a David-star in the center, and the triangles and the David-star are removed (see Fig. 9). We leave it as an exercise to the reader to show that $d_f = \log 6/\log 3$ and $d_h = \log 4/\log 3$.

The Cantor Set

Cantor sets are examples of disconnected fractals (*fractal dust*). The simplest set is the triadic Cantor set (see Fig. 10). We divide a unit interval $[0, 1]$ into three equal intervals and remove the central one. In each following iteration, each of the remaining intervals is treated in this way. In the limit of $n = \infty$ iterations one obtains a set of points. Since $M(1/3\,L) = 1/2\,M(L)$, we have $d_f = \log 2/\log 3 \cong 0.631$, which is smaller than one.

In chaotic systems, strange fractal attractors occur. The simplest strange attractor is the Cantor set. It occurs, for example, when considering the one-dimensional *logistic map*

$$x_{t+1} = \lambda x_t(1 - x_t). \tag{7}$$

The index $t = 0, 1, 2, \ldots$ represents a discrete time. For $0 \leq \lambda \leq 4$ and x_0 between 0 and 1, the trajectories x_t are bounded between 0 and 1. The dynamical behavior of x_t for $t \to \infty$ depends on the parameter λ. Below $\lambda_1 = 3$,
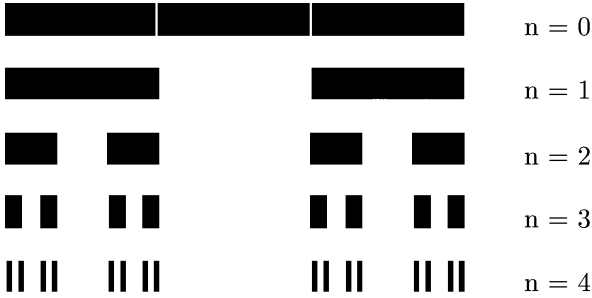
Fractal Geometry, A Brief Introduction to, Figure 10
The first iterations of the triadic Cantor set. The fractal dimension of this Cantor set is $d_f = \log 2/\log 3$

only one stable fixed-point exists to which x_t is attracted. At λ_1, this fixed-point becomes unstable and bifurcates into two new stable fixed-points. At large times, the trajectories move alternately between both fixed-points, and the motion is periodic with period 2. At $\lambda_2 = 1 + \sqrt{6} \cong 3.449$ each of the two fixed-points bifurcates into two new stable fix points and the motion becomes periodic with period 4. As λ is increased, further bifurcation points λ_n occur, with periods of 2^n between λ_n and λ_{n+1}.

For large n, the differences between λ_{n+1} and λ_n become smaller and smaller, according to the law $\lambda_{n+1} - \lambda_n = (\lambda_n - \lambda_{n-1})/\delta$, where $\delta \cong 4.6692$ is the so-called Feigenbaum constant. The Feigenbaum constant is "universal", since it applies to all nonlinear "single-hump" maps with a quadratic maximum [17].

At $\lambda_\infty \cong 3.569\,945\,6$, an infinite period occurs, where the trajectories x_t move in a "chaotic" way between infinite attractor points. These attractor points define the

strange attractor, which forms a Cantor set with a fractal dimension $d_f \cong 0.538$ [18]. For a further discussion of strange attractors and chaotic dynamics we refer to [3,8,9].

The Mandelbrot–Given Fractal

This fractal was suggested as a model for percolation clusters and its substructures (see Sect. 3.4 and Chap. 2 in [13]). Figure 11 shows the first three generations of the Mandelbrot–Given fractal [19]. At each generation, each segment of length a is replaced by 8 segments of length $a/3$. Accordingly, the fractal dimension is $d_f = \log 8/\log 3 \cong 1.893$, which is very close to $d_f = 91/46 \cong 1.896$ for percolation in two dimensions. It is easy to verify that $d_\ell = d_f$, and therefore $d_{min} = 1$. The structure contains loops, branches, and dangling ends of all length scales.

Imagine applying a voltage difference between two sites at opposite edges of a metallic Mandelbrot–Given fractal: the *backbone* of the fractal consists of those bonds which carry the electric current. The *dangling ends* are those parts of the cluster which carry no current and are connected to the backbone by a single bond only. The *red bonds* (or singly connected bonds) are those bonds that carry the total current; when they are cut the current flow stops. The *blobs*, finally, are those parts of the backbone that remain after the red bonds have been removed.

The backbone of this fractal can be obtained easily by eliminating the dangling ends when generating the fractal (see Fig. 12). It is easy to see that the fractal dimension of the backbone is $d_B = \log 6/\log 3 \cong 1.63$. The red bonds are all located along the x axis of the figure and form a Cantor set with the fractal dimension $d_{red} = \log 2/\log 3 \cong 0.63$.
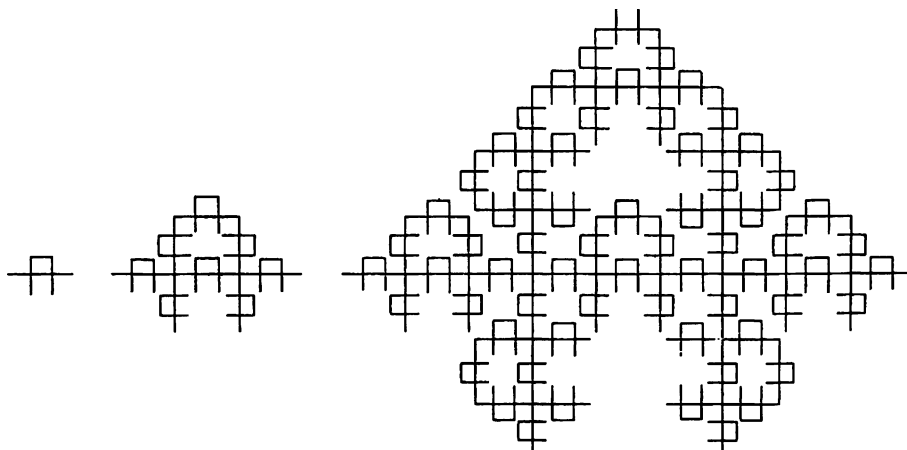
Fractal Geometry, A Brief Introduction to, Figure 11
Three generations of the Mandelbrot–Given fractal. The fractal dimension of the Mandelbrot–Given fractal is $d_f = \log 8/\log 3$
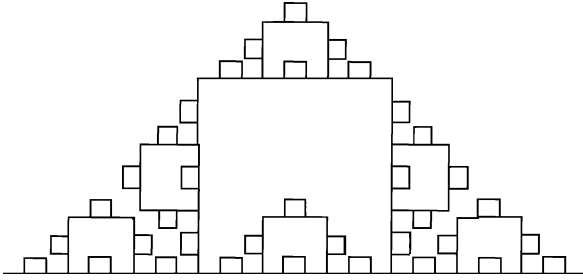
Fractal Geometry, A Brief Introduction to, Figure 12
The backbone of the Mandelbrot–Given fractal, with the red bonds shown in *bold*

Fractal Geometry, A Brief Introduction to, Figure 13
Julia sets for **a** $c = i$ and **b** $c = 0.11031 - i0.67037$. After [9]

Julia Sets and the Mandelbrot Set

A complex version of the logistic map (7) is

$$z_{t+1} = z_t^2 + c \,, \tag{8}$$

where both the trajectories z_t and the constant c are complex numbers. The question is: if a certain c-value is given, for example $c = -1.5652 - i1.03225$, for which initial values z_0 are the trajectories z_t bounded? The set of those values forms the *filled-in Julia set*, and the boundary points of them form the Julia set.

To clarify these definitions, consider the simple case $c = 0$. For $|z_0| > 1$, z_t tends to infinity, while for $|z_0| < 1$, z_t tends to zero. Accordingly, the filled-in Julia set is the set of all points $|z_0| \leq 1$, the Julia set is the set of all points $|z_0| = 1$.

In general, points on the Julia set form a chaotic motion on the set, while points outside the Julia set move away from the set. Accordingly, the Julia set can be regarded as a "repeller" with respect to Eq. (8). To generate the Julia set, it is thus practical to use the inverted transformation

$$z_t = \pm\sqrt{z_{t+1} - c} \,, \tag{9}$$

start with an arbitrarily large value for $t + 1$, and go backward in time. By going backward in time, even points far away from the Julia set are attracted by the Julia set.

For obtaining the Julia set for a given value of c, one starts with some arbitrary value for z_{t+1}, for example, $z_{t+1} = 2$. To obtain z_t, we use Eq. (9), and determine the sign randomly. This procedure is continued to obtain z_{t-1}, z_{t-2}, etc. By disregarding the initial points, e. g., the first 1000 points, one obtains a good approximation of the Julia set.

The Julia sets can be connected (Fig. 13a) or disconnected (Fig. 13b) like the Cantor sets. The self-similarity of the pictures is easy to see. The set of c values that yield connected Julia sets forms the famous Mandelbrot set. It has been shown by Douady and Hubbard [20] that the Mandelbrot set is identical to that set of c values for which z_t converges starting from the initial point $z_0 = 0$. For a detailed discussion with beautiful pictures see [10] and Chaps. 13 and 14 in [3].

Random Fractal Models

In this section we present several random fractal models that are widely used to mimic fractal systems in nature. We begin with perhaps the simplest fractal model, the random walk.

Random Walks

Imagine a random walker on a square lattice or a simple cubic lattice. In one unit of time, the random walker advances one step of length a to a randomly chosen nearest neighbor site. Let us assume that the walker is unwinding a wire, which he connects to each site along his way. The length (mass) M of the wire that connects the random walker with his starting point is proportional to the number of steps n (Fig. 14) performed by the walker.

Since for a random walk in any d-dimensional space the mean end-to-end distance R is proportional to $n^{1/2}$ (for
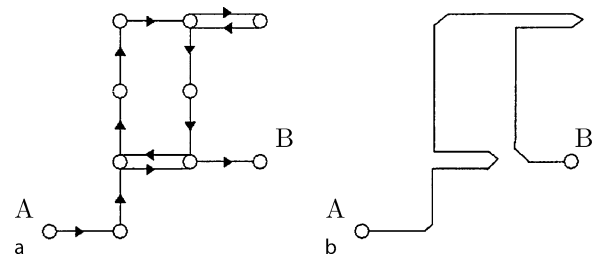
Fractal Geometry, A Brief Introduction to, Figure 14
a A normal random walk with loops. **b** A random walk without loops

a simple derivation see e. g., Chap. 3 in [13], it follows that $M \sim R^2$. Thus Eq. (4) implies that the fractal dimension of the structure formed by this wire is $d_f = 2$, for all lattices.

The resulting structure has loops, since the walker can return to the same site. We expect the chemical dimension d_ℓ to be 2 in $d = 2$ and to decrease with increasing d, since. Loops become less relevant. For $d \geq 4$ we have $d_\ell = 1$. If we assume, however, that there is no contact between sections of the wire connected to the same site (Fig. 14b), the structure is by definition linear, i. e., $d_\ell = 1$ for all d. For more details on random walks and its relation to Brownian motion, see Chap. 5 in [15] and [21].

Self-Avoiding Walks

Self-avoiding walks (SAWs) are defined as the subset of all nonintersecting random walk configurations. An example is shown in Fig. 15a. As was found by Flory in 1944 [22], the end-to-end distance of SAWs scales with the number of steps n as

$$R \sim n^\nu , \qquad (10)$$

with $\nu = 3/(d + 2)$ for $d \leq 4$ and $\nu = 1/2$ for $d > 4$. Since n is proportional to the mass of the chain, it follows from Eq. (4) that $d_f = 1/\nu$. Self-avoiding walks serve as models for polymers in solution, see [23].

Subsets of SAWs do not necessarily have the same fractal dimension. Examples are the kinetic growth walk (KGW) [24] and the smart growth walk (SGW) [25], sometimes also called the "true" or "intelligent" self-avoiding walk. In the KGW, a random walker can only step on those sites that have not been visited before. Asymptotically, after many steps n, the KGW has the same fractal dimension as SAWs. In $d = 2$, however, the asymptotic regime is difficult to reach numerically, since the random walker can be trapped with high probability (see Fig. 15b). A related structure is the hull of a random walk in $d = 2$. It has been conjectured by Mandelbrot [1] that the fractal dimension of the hull is $d_h = 4/3$, see also [26].

In the SGW, the random walker avoids traps by stepping only at those sites from which he can reach infinity. The structure formed by the SGW is more compact and characterized by $d_f = 7/4$ in $d = 2$ [25]. Related structures with the same fractal dimension are the hull of percolation clusters (see also Sect. "Percolation") and diffusion fronts (for a detailed discussion of both systems see also Chaps. 2 and 7 in [13]).

Kinetic Aggregation

The simplest model of a fractal generated by diffusion of particles is the diffusion-limited aggregation (DLA)
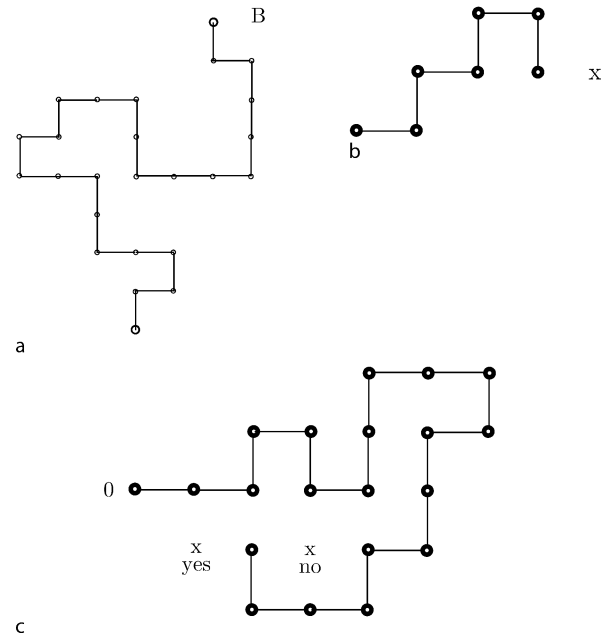
Fractal Geometry, A Brief Introduction to, Figure 15
a A typical self-avoiding walk. **b** A kinetic growth walk after 8 steps. The available sites are marked by *crosses*. **c** A smart growth walk after 19 steps. The only available site is marked by "yes"

model, which was introduced by Witten and Sander in 1981 [27]. In the lattice version of the model, a seed particle is fixed at the origin of a given lattice and a second particle is released from a circle around the origin. This particle performs a random walk on the lattice. When it comes to a nearest neighbor site of the seed, it sticks and a cluster (aggregate) of two particles is formed. Next, a third particle is released from the circle and performs a random walk. When it reaches a neighboring site of the aggregate, it sticks and becomes part of the cluster. This procedure is repeated many times until a cluster of the desired number of sites is generated. For saving computational time it is convenient to eliminate particles that have diffused too far away from the cluster (see Fig. 16).

In the continuum (off-lattice) version of the model, the particles have a certain radius a and are not restricted to diffusing on lattice sites. At each time step, the length ($\leq a$) and the direction of the step are chosen randomly. The diffusing particle sticks to the cluster, when its center comes within a distance a of the cluster perimeter. It was found numerically that for off-lattice DLA, $d_f = 1.71 \pm 0.01$ in $d = 2$ and $d_f = 2.5 \pm 0.1$ in $d = 3$ [28,29]. These results may be compared with the mean field result $d_f = (d^2 + 1)/(d + 1)$ [30]. For a renormalization group approach, see [31] and references
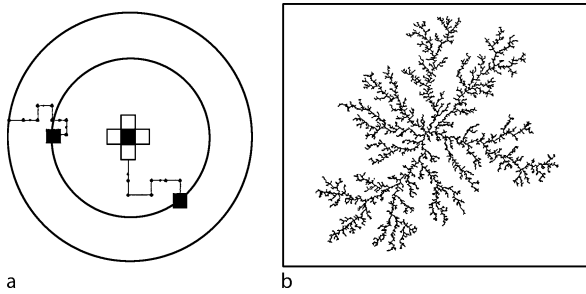
Fractal Geometry, A Brief Introduction to, Figure 16
a Generation of a DLA cluster. The inner release radius is usually a little larger than the maximum distance of a cluster site from the center, the outer absorbing radius is typically 10 times this distance. b A typical off-lattice DLA cluster of 10,000 particles

p = 0.2   p = 0.59   p = 0.8

Fractal Geometry, A Brief Introduction to, Figure 17
Square lattice of size 20 × 20. Sites have been randomly occupied with probability p ($p = 0.20, 0.59, 0.80$). Sites belonging to finite clusters are marked by *full circles*, while sites on the infinite cluster are marked by *open circles*

therein. The chemical dimension d_ℓ is found to be equal to d_f [32].

Diffusion-limited aggregation serves as an archetype for a large number of fractal realizations in nature, including viscous fingering, dielectric breakdown, chemical dissolution, electrodeposition, dendritic and snowflake growth, and the growth of bacterial colonies. For a detailed discussion of the applications of DLA we refer to [5,13], and [29]. Models for the complex structure of DLA have been developed by Mandelbrot [33] and Schwarzer et al. [34].

A somewhat related model for aggregation is the cluster-cluster aggregation (CCA) [35]. In CCA, one starts from a very low concentration of particles diffusing on a lattice. When two particles meet, they form a cluster of two, which can also diffuse. When the cluster meets another particle or another cluster, a larger cluster is formed. In this way, larger and larger aggregates are formed. The structures are less compact than DLA, with $d_f \cong 1.4$ in $d = 2$ and $d_f \cong 1.8$ in $d = 3$. CCA seems to be a good model for smoke aggregates in air and for gold colloids. For a discussion see Chap. 8 in [13].

Percolation

Consider a square lattice, where each site is occupied randomly with probability p or empty with probability $1 - p$. At low concentration p, the occupied sites are either isolated or form small clusters (Fig. 17a). Two occupied sites belong to the same cluster, if they are connected by a path of nearest neighbor occupied sites. When p is increased, the average size of the clusters increases. At a critical concentration p_c (also called the percolation threshold) a large cluster appears which connects opposite edges of the lattice (Fig. 17b). This cluster is called the *infinite* cluster, since its size diverges when the size of the lattice is in-

creased to infinity. When p is increased further, the density of the infinite cluster increases, since more and more sites become part of the infinite cluster, and the average size of the *finite* clusters decreases (Fig. 17c).

The percolation transition is characterized by the geometrical properties of the clusters near p_c. The probability P_∞ that a site belongs to the infinite cluster is zero below p_c and increases above p_c as

$$P_\infty \sim (p - p_c)^\beta \ . \tag{11}$$

The linear size of the *finite* clusters, below and above p_c, is characterized by the *correlation length* ξ. The correlation length is defined as the mean distance between two sites on the same finite cluster and represents the characteristic length scale in percolation. When p approaches p_c, ξ increases as

$$\xi \sim |p - p_c|^{-\nu} \ , \tag{12}$$

with the same exponent ν below and above the threshold. While p_c depends explicitly on the type of the lattice (e. g., $p_c \cong 0.593$ for the square lattice and $1/2$ for the triangular lattice), the *critical exponents* β and ν are universal and depend only on the dimension d of the lattice, but not on the type of the lattice.

Near p_c, on length scales smaller than ξ, both the infinite cluster and the finite clusters are self-similar. Above p_c, on length scales larger than ξ, the infinite cluster can be regarded as an homogeneous system which is composed of many unit cells of size ξ. Mathematically, this can be summarized as

$$M(r) \sim \begin{cases} r^{d_f} \ , & r \ll \xi \ , \\ r^d \ , & r \gg \xi \ . \end{cases} \tag{13}$$

The fractal dimension d_f can be related to β and ν:

$$d_f = d - \frac{\beta}{\nu} \ . \tag{14}$$

Since β and ν are universal exponents, d_f is also universal. One obtains $d_f = 91/48$ in $d = 2$ and $d_f \cong 2.5$ in $d = 3$.
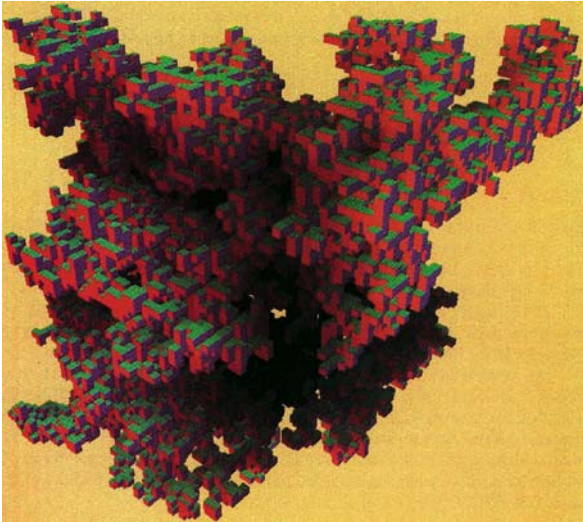
Fractal Geometry, A Brief Introduction to, Figure 18
A large percolation cluster in $d = 3$. The colors mark the topological distance from an arbitrary center of the cluster in the middle of the page. Courtesy of M. Meyer

The chemical dimension d_ℓ is smaller than d_f, $d_\ell \cong 1.15$ in $d = 2$ and $d_\ell \cong 1.33$ in $d = 3$. A large percolation cluster in $d = 3$ is shown in Fig. 18.

Interestingly, a percolation cluster is composed of several fractal sub-structures such as the backbone, dangling ends, blobs, External perimeter, and the red bonds, which are all described by different fractal dimensions.

The percolation model has found applications in physics, chemistry, and biology, where occupied and empty sites may represent very different physical, chemical, or biological properties. Examples are the physics of two component systems (the random resistor, magnetic or superconducting networks), the polymerization process in chemistry, and the spreading of epidemics and forest fires. For reviews with a comprehensive list of references, see Chaps. 2 and 3 of [13] and [36,37,38].

How to Measure the Fractal Dimension

One of the most important "practical" problems is to determine the fractal dimension d_f of either a computer generated fractal or a digitized fractal picture. Here we sketch the two most useful methods: the "sandbox" method and the "box counting" method.

The Sandbox Method

To determine d_f, we first choose one site (or one pixel) of the fractal as the origin for n circles of radii $R_1 < R_2 < \cdots < R_n$, where R_n is smaller than the radius R of the fractal, and count the number of points (pixels) $M_1(R_i)$ within each circle i. (Sometimes, it is more convenient to choose n squares of side length $L_1 \ldots L_n$ instead of the circles.) We repeat this procedure by choosing randomly many other (altogether m) pixels as origins for the n circles and determine the corresponding number of points $M_j(R_i)$, $j = 2, 3, \ldots, m$ within each circle (see Fig. 19a). We obtain the mean number of points $M(R_i)$ within each circle by averaging, $M(R_i) = 1/m \sum_{j=1}^{m} M_j(R_i)$, and plot $M(R_i)$ versus R_i in a double logarithmic plot. The slope of the curve, for large values of R_i, determines the fractal dimension.

In order to avoid boundary effects, the radii must be smaller than the radius of the fractal, and the centers of the circles must be chosen well inside the fractal, so that the largest circles will be well within the fractal. In order to obtain good statistics, one has either to take a very large fractal cluster with many centers of circles or many realizations of the same fractal.

The Box Counting Method

We draw a grid on the fractal that consists of N_1^2 squares, and determine the number of squares $S(N_1)$ that are needed to cover the fractal (see Fig. 19b). Next we choose finer and finer grids with $N_1^2 < N_2^2 < N_3^2 < \cdots < N_m^2$ squares and calculate the corresponding numbers of squares $S(N_1) \ldots S(N_m)$ needed to cover the fractal. Since $S(N)$ scales as

$$S(N) \sim N^{-d_f} , \tag{15}$$

we obtain the fractal dimension by plotting $S(N)$ versus $1/N$ in a double logarithmic plot. The asymptotic slope, for large N, gives d_f.

a   b
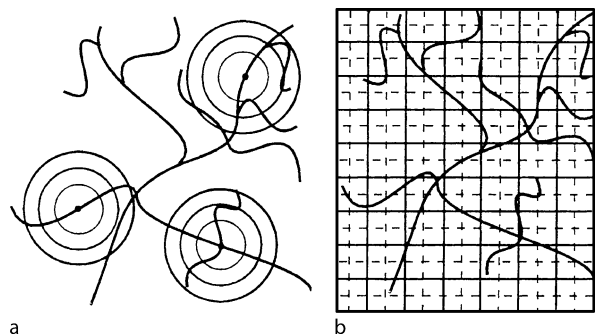
Fractal Geometry, A Brief Introduction to, Figure 19
Illustrations for determining the fractal dimension: **a** the sandbox method, **b** the box counting method

Of course, the finest grid size must be larger than the pixel size, so that many pixels can fall into the smallest square. To improve statistics, one should average $S(N)$ over many realizations of the fractal. For applying this method to identify self-similarity in real networks, see Song et al. [39].

Self-Affine Fractals

The fractal structures we have discussed in the previous sections are self-similar: if we cut a small piece out of a fractal and magnify it isotropically to the size of the original, both the original and the magnification look the same. By magnifying isotropically, we have rescaled the x, y, and z axis by the same factor.

There exist, however, systems that are invariant only under *anisotropic* magnifications. These systems are called *self-affine* [1]. A simple model for a self-affine fractal is shown in Fig. 20. The structure is invariant under the anisotropic magnification $x \to 4x$, $y \to 2y$. If we cut a small piece out of the original picture (in the limit of $n \to \infty$ iterations), and rescale the x axis by a factor of four and the y axis by a factor of two, we will obtain exactly the original structure. In other words, if we describe the form of the curve in Fig. 20 by the function $F(x)$, this function satisfies the equation $F(4x) = 2F(x) = 4^{1/2}F(x)$.

In general, if a self-affine curve is scale invariant under the transformation $x \to bx$, $y \to ay$, we have

$$F(bx) = aF(x) \equiv b^H F(x) \,, \tag{16}$$

where the exponent $H = \log a / \log b$ is called the Hurst exponent [1]. The solution of the functional equation (16) is simply $F(x) = Ax^H$. In the example of Fig. 20, $H = 1/2$.

Next we consider random self-affine structures, which are used as models for random surfaces. The simplest structure is generated by a one-dimensional random walk, where the abscissa is the time axis and the ordinate is the displacement $Z(t) = \sum_{i=1}^{t} e_i$ of the walker from its starting point. Here, $e_i = \pm 1$ is the unit step made by the random walker at time t. Since different steps of the random walker are uncorrelated, $\langle e_i e_j \rangle = \delta_{ij}$, it follows that the root mean square displacement $F(t) \equiv \langle Z^2(t) \rangle^{1/2} = t^{1/2}$, and the Hurst exponent of the structure is $H = 1/2$.

Next we assume that different steps i and j are correlated in such a way that $\langle e_i e_j \rangle = b|i - j|^{-\gamma}$, $1 > \gamma \geq 0$. To see how the Hurst exponent depends on γ, we have to evaluate again $\langle Z^2(t) \rangle = \sum_{i,j}^{t} \langle e_i e_j \rangle$. For calculating the double sum it is convenient to introduce the Fourier transform of e_i, $e_\omega = (1/\Omega)^{1/2} \sum_{l=1}^{\Omega} e_l \exp(-i\omega l)$, where Ω is the number of sites in the system. It is easy to verify that $\langle Z^2(t) \rangle$ can be expressed in terms of the power spectrum $\langle e_\omega e_{-\omega} \rangle$ [40]:

$$\langle Z^2(t) \rangle = \frac{1}{\Omega} \sum_{\omega} \langle e_\omega e_{-\omega} \rangle |f(\omega, t)|^2 \,, \tag{17a}$$

where

$$f(\omega, t) \equiv \frac{e^{-i\omega(t+1)} - 1}{e^{-i\omega} - 1} \,.$$
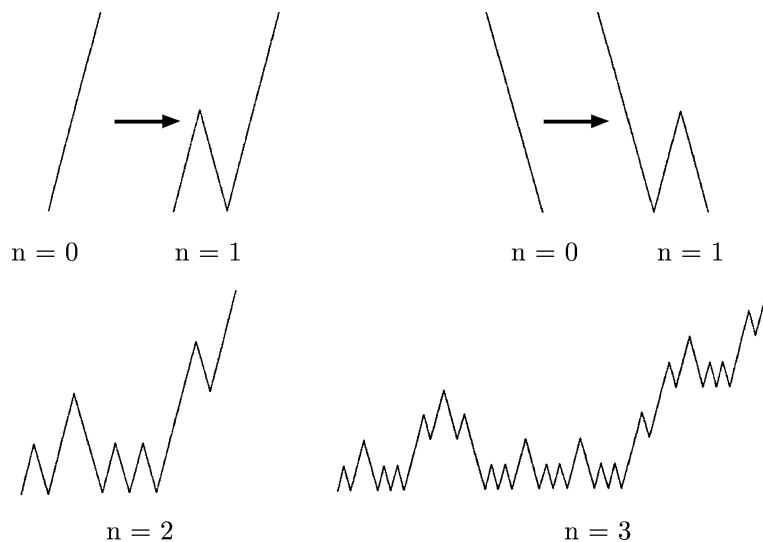
Fractal Geometry, A Brief Introduction to, Figure 20
A simple deterministic model of a self-affine fractal

$n=0$ $n=1$ $n=2$ $n=3$

$d=1$ o   o ⟶ o x o ⟶ o x o x o ⟶ oxoxoxoxo

scale $r_0=1$ $r_1=1/2$ $r_2=(1/2)^2$ $r_3=(1/2)^3$

$d=2$

scale $r_0=1$ $r_1=1/\sqrt{2}$ $r_2=1/(\sqrt{2})^2$ $r_3=1/(\sqrt{2})^3$

fluctuation $\sigma_0=1$ $\sigma_1=r_1^H$ $\sigma_2=r_2^H$ $\sigma_3=r_3^H$
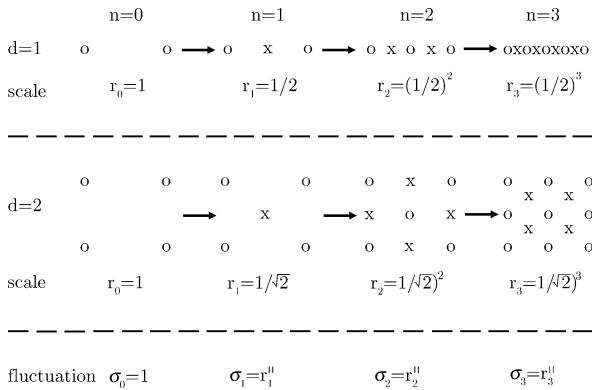
Fractal Geometry, A Brief Introduction to, Figure 21
Illustration of the successive random addition method in $d = 1$ and $d = 2$. The *circles* mark those points that have been considered already in the earlier iterations, the *crosses* mark the new midpoints added at the present iteration. At each iteration n, first the Z values of the midpoints are determined by linear interpolation from the neighboring points, and then random displacements of variance σ_n are added to all Z values

Since the power spectrum scales as

$$\langle e_\omega e_{-\omega} \rangle \sim \omega^{-(1-\gamma)} , \qquad (17b)$$

the integration of (17a) yields, for large t,

$$\langle Z^2(t) \rangle \sim t^{2-\gamma} . \qquad (17c)$$

Therefore, the Hurst exponent is $H = (2 - \gamma)/2$. According to Eq. (17c), for $0 < \gamma < 1$, $\langle Z^2(t) \rangle$ increases faster in time than the uncorrelated random walk. The long-range correlated random walks were called *fractional Brownian motion* by Mandelbrot [1].

There exist several methods to generate correlated random surfaces. We shall describe the *successive random additions* method [41], which iteratively generates the self-affine function $Z(x)$ in the unit interval $0 \leq x \leq 1$. An alternative method that is detailed in the chapter of Jan Kantelhardt is the Fourier-filtering technique and its variants.

In the $n = 0$ iteration, we start at the edges $x = 0$ and $x = 1$ of the unit interval and choose the values of $Z(0)$ and $Z(1)$ from a distribution with zero mean and variance $\sigma_0^2 = 1$ (see Fig. 21). In the $n = 1$ iteration, we choose the midpoint $x = 1/2$ and determine $Z(1/2)$ by linear interpolation, i. e., $Z(1/2) = (Z(0) + Z(1))/2$, and add to all so-far calculated Z values ($Z(0)$, $Z(1/2)$, and $Z(1)$) random displacements from the same distribution as before, but with a variance $\sigma_1 = (1/2)^H$ (see Fig. 21). In the $n = 2$ iteration we again first choose the midpoints ($x = 1/4$ and $x = 3/4$), determine their Z values by linear interpolation, and add to all so-far calculated Z values random displacements from the same distribution as before, but with
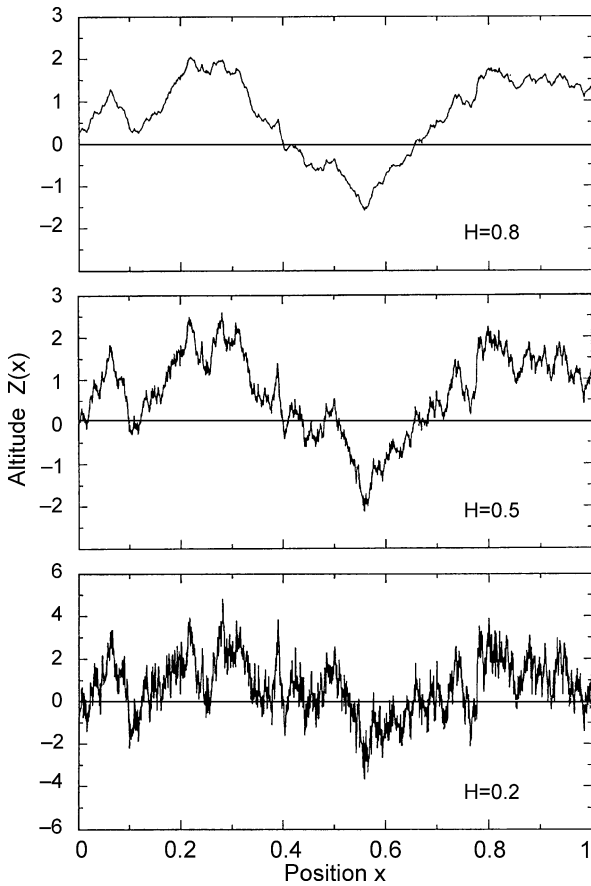
Fractal Geometry, A Brief Introduction to, Figure 22
Correlated random walks with $H = 0.2$, 0.5, and 0.8, generated by the successive random addition method in $d = 1$

a variance $\sigma_2 = (1/2)^{2H}$. In general, in the nth iteration, one first interpolates the Z values of the midpoints and then adds random displacements to all existing Z values, with variance $\sigma_n = (1/2)^{nH}$. The procedure is repeated until the required resolution of the surface is obtained. Figure 22 shows the graphs of three random surfaces generated this way, with $H = 0.2$, $H = 0.5$, and $H = 0.8$.

The generalization of the successive random addition method to two dimensions is straightforward (see Fig. 21). We consider a function $Z(x, y)$ on the unit square $0 \leq x, y \leq 1$. In the $n = 0$ iteration, we start with the four corners $(x, y) = (0, 0), (1, 0), (1, 1), (0, 1)$ of the unit square and choose their Z values from a distribution with zero mean and variance $\sigma_0^2 = 1$ (see Fig. 21). In the $n = 1$ iteration, we choose the midpoint at $(x, y) = (1/2, 1/2)$ and determine $Z(1/2, 1/2)$ by linear interpolation, i. e., $Z(1/2, 1/2) = (Z(0, 0) + Z(0, 1) + Z(1, 1) + Z(1, 0))/4$. Then we add to all so far calculated Z-values ($Z(0, 0)$, $Z(0, 1)$,

$Z(1, 0)$, $Z(1, 1)$ and $Z(1/2, 1/2)$) random displacements from the same distribution as before, but with a variance $\sigma_1 = (1/\sqrt{2})^H$ (see Fig. 21). In the $n = 2$ iteration we again choose the midpoints of the five sites $(0, 1/2)$, $(1/2, 0)$, $(1/2, 1)$ and $(1, 1/2)$, determine their Z value by linear interpolation, and add to all so far calculated Z values random displacements from the same distribution as before, but with a variance $\sigma_2 = (1/\sqrt{2})^{2H}$. This procedure is repeated again and again, until the required resolution of the surface is obtained.

At the end of this section we like to note that self-similar or self-affine fractal structures with features similar to those fractal models discussed above can be found in nature on all, astronomic as well as microscopic, length scales. Examples include clusters of galaxies (the fractal dimension of the mass distribution is about 1.2 [42]), the crater landscape of the moon, the distribution of earthquakes (see Chap. 2 in [15]), and the structure of coastlines, rivers, mountains, and clouds. Fractal cracks (see, for example, Chap. 5 in [13]) occur on length scales ranging from 10^3 km (like the San Andreas fault) to micrometers (like fractures in solid materials) [44].

Many naturally growing plants show fractal structures, examples range from trees and the roots of trees to cauliflower and broccoli. The patterns of blood vessels in the human body, the kidney, the lung, and some types of nerve cells have fractal features (see Chap. 3 in [15]). In materials sciences, fractals appear in polymers, gels, ionic glasses, aggregates, electro-deposition, rough interfaces and surfaces (see [13] and Chaps. 4 and 6 in [15]), as well as in fine particle systems [43]. In all these structures there is no characteristic length scale in the system besides the physical upper and lower cut-offs.

The occurrence of self-similar or self-affine fractals is not limited to structures in real space as we will discuss in the next section.

Long-Term Correlated Records

Long-range dependencies as described in the previous section do not only occur in surfaces. Of great interest is long-term memory in climate, physiology and financial markets, the examples range from river floods [45,46,47,48,49], temperatures [50,51,52,53,54], and wind fields [55] to market volatilities [56], heart-beat intervals [57,58] and internet traffic [59].

Consider a record x_i of discrete numbers, where the index i runs from 1 to N. x_i may be daily or annual temperatures, daily or annual river flows, or any other set of data consisting of N successive data points. We are interested in the fluctuations of the data around their (some-times seasonal) mean value. Without loss of generality, we assume that the mean of the data is zero and the variance equal to one. In analogy to the previous section, we call the data long-term correlated, when the corresponding auto-correlation function $C_x(s)$ decays by a power law,

$$C_x(s) = \langle x_i x_{i+s} \rangle \equiv \frac{1}{N-s} \sum_{i=1}^{N-s} x_1 x_{i+s} \sim s^{-\gamma}, \quad (18)$$

where γ denotes the correlation exponent, $0 < \gamma < 1$. Such correlations are named 'long-term' since the mean correlation time $T = \int_0^\infty C_x(s) \, ds$ diverges in the limit of infinitely long series where $N \to \infty$. If the x_i are uncorrelated, $C_x(s) = 0$ for $s > 0$. More generally, if correlations exist up to a certain correlation time s_x, then $C(s) > 0$ for $s < s_x$ and $C(s) = 0$ for $s > s_x$.

Figure 23 shows parts of an uncorrelated (*left*) and a long-term correlated (*right*) record, with $\gamma = 0.4$; both series have been generated by the computer. The red line is the moving average over 30 data points. For the uncorrelated data, the moving average is close to zero, while for the long-term correlated data set, the moving average can have large deviations from the mean, forming some kind of mountain valley structure. This structure is a consequence of the power-law persistence. The mountains and valleys in Fig. 23b look as if they had been generated by external trends, and one might be inclined to draw a trendline and to extrapolate the line into the near future for some kind of prognosis. But since the data are trend-free, only a short-term prognosis utilizing the persistence can be made, and not a longer-term prognosis, which often is the aim of such a regression analysis.

Alternatively, in analogy to what we described above for self-affine surfaces, one can divide the data set in K_s equidistant windows of length s and determine in each window ν the squared sum

$$F_\nu^2(s) = \left(\sum_{i=1}^s x_i \right)^2 \quad (19a)$$

and detect how the average of this quantity over all windows, $F^2(s) = 1/K_s \sum_{\nu=1}^{K_s} F_\nu^2(s)$, scales with the window size s. For long-term correlated data one can show that $F^2(s)$ scales as $\langle Z^2(t) \rangle$ in the previous section, i. e.

$$F^2(s) \sim s^{2\alpha}, \quad (19b)$$

where $\alpha = 1 - \gamma/2$. This relation represents an alternative way to determine the correlation exponent γ.

Since trends resemble long-term correlations and vice versa, there is a general problem to distinguish between
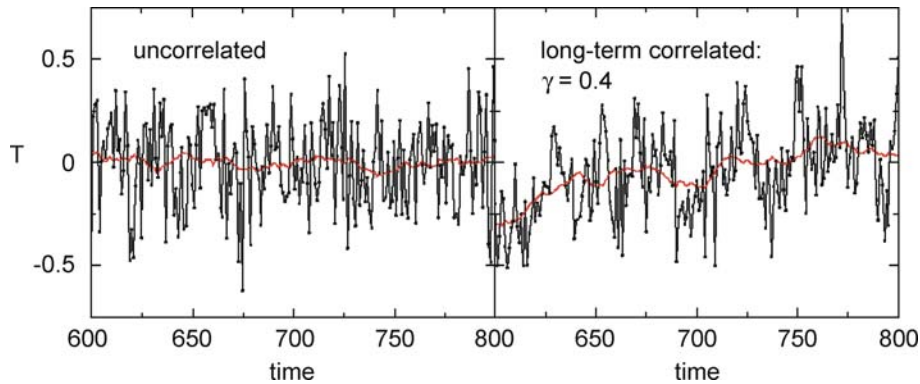
Fractal Geometry, A Brief Introduction to, Figure 23
Comparison of an uncorrelated and a long-term correlated record with $\gamma = 0.4$. The *full line* is the moving average over 30 data points

trends and long-term persistence. In recent years, several methods have been developed, mostly based on the hierarchical detrended fluctuation analysis (DFAn) where long-term correlations in the presence of smooth polynomial trends of order $n - 1$ can be detected [57,58,60] (see also ▶ Fractal and Multifractal Time Series). In DFAn, one considers the cumulated sum ("profile") of the x_i and divides the N data points of the profile into equidistant windows of fixed length s. Then one determines, in each window, the best fit of the profile by an nth order polynomial and determines the variance around the fit. Finally, one averages these variances to obtain the mean variance $F^2_{(n)}$ and the corresponding mean standard deviation (mean fluctuation) $F_{(n)}(s)$. One can show that for long-term correlated trend-free data $F_{(n)}(s)$ scales with the window size s as $F(s)$ in Eq. (19b), i. e., $F_{(n)}(s) \sim s^{\alpha}$, with $\alpha = 1 - \gamma/2$, irrespective of the order of the polynomial n. For short-term correlated records (including the case $\gamma \geq 1$), the exponent is $1/2$ for s above s_x. It is easy to verify that trends of order $k - 1$ in the original data are eliminated in $F_{(k)}(s)$ but contribute to $F_{(k-1)}, F_{(k-2)}$ etc., and this allows one to determine the correlation exponent γ in the presence of trends. For example, in the case of a linear trend, DFA0 and DFA1 (where $F_{(0)}(s)$ and $F_{(1)}(s)$ are determined) are affected by the trend and will exaggerate the asymptotic exponents α, while DFA2, DFA3 etc. (where $F_{(2)}(s)$ and $F_{(3)}(s)$ etc. is determined) are not affected by the trend and will show, in a double logarithmic plot, the same value of α, which then gives immediately the correlation exponent γ. When γ is known this way, one can try to detect the trend, but there is no unique treatment available. In recent papers [61,62,63], different kinds of analysis have been elaborated and applied to estimate trends in the temperature records of the Northern Hemisphere and Siberian locations.

Climate Records

Figure 24 shows representative results of the DFAn analysis, for temperature, precipitation and run-off data. For continental temperatures, the exponent α is around 0.65, while for island stations and sea surface temperatures the exponent is considerably higher. There is no crossover towards uncorrelated behavior at larger time scales. For the precipitation data, the exponent is close to 0.55, not being significantly larger than for uncorrelated records.

Figure 25 shows a summary of the exponent α for a large number of climate records. It is interesting to note that while the distribution of α-values is quite broad for run-off, sea-surface temperature, and precipitation records, the distribution is quite narrow, located around $\alpha = 0.65$ for continental atmospheric temperature records. For the island records, the exponent is larger. The quite universal exponent $\alpha = 0.65$ for continental stations can be used as an efficient test bed for climate models [62,64,65].

The time window accessible by DFAn is typically $1/4$ of the length of the record. For instrumental records, the time window is thus restricted to about 50 years. For extending this limit, one has to take reconstructed records or model data, which range up to 2000y. Both have, of course, large uncertainties, but it is remarkable that exactly the same kind of long-term correlations can be found in these data, thus extending the time scale where long-term memory exists to at least 500y [61,62].

Clustering of Extreme Events

Next we consider the consequences of long-term memory on the occurrence of rare events. Understanding (and predicting) the occurrence of extreme events is one of the
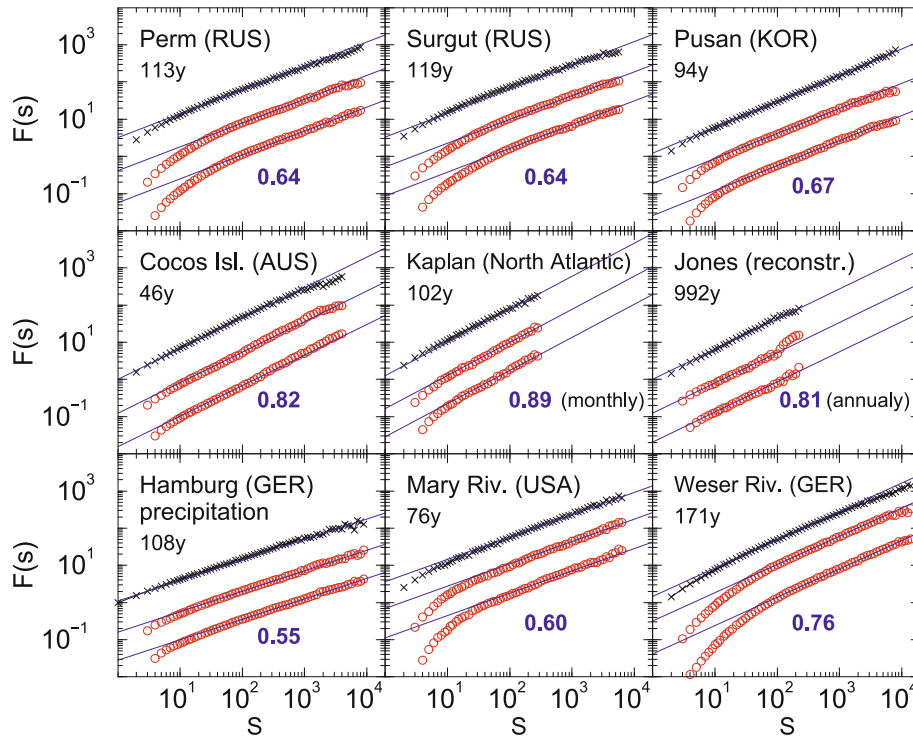
Fractal Geometry, A Brief Introduction to, Figure 24
DFA*n* analysis of six temperature records, one precipitation record and two run-off records. The *black curves* **are the DFA0 results, while the** *upper red curves* **refer to DFA1 and the** *lower red curves* **to DFA2. The** *blue numbers* **denote the asymptotic slopes of the curves**

major challenges in science (see, e. g., [68]). An important quantity here is the time interval between successive extreme events (see Fig. 26), and by understanding the statistics of these return intervals one aims to better understand the occurrence of extreme events.

Since extreme events are, by definition, very rare and the statistics of their return intervals poor, one usually studies also the return intervals between less extreme events, where the data are above some threshold q and where the statistics is better, and hopes to find some general "scaling" relations between the return intervals at low and high thresholds, which then allows one to extrapolate the results to very large, extreme thresholds (see Fig. 26).

For uncorrelated data, the return intervals are independent of each other and their probability density function (pdf) is a simple exponential, $P_q(r) = (1/R_q) \times \exp(-r/R_q)$. In this case, all relevant quantities can be derived from the knowledge of the mean return interval R_q. Since the return intervals are uncorrelated, a sequential ordering cannot occur. There are many cases, however, where some kind of ordering has been observed where the hazardous events cluster, for example in the floods

in Central Europe during the middle ages or in the historic water levels of the Nile river which are shown in Fig. 26 for 663y. Even by eye one can realize that the events are not distributed randomly but are arranged in clusters. A similar clustering was observed for extreme floods, winter storms, and avalanches in Central Europe (see, e. g., Figs. 4.4, 4.7, 4.10, and 4.13 in [69], Fig. 66 in [70], and Fig. 2 in [71]). The reason for this clustering is the long-term memory.

Figure 27 shows $P_q(r)$ for long-term correlated records with $\alpha = 0.4$ (corresponding to $\gamma = 0.8$), for three values of the mean return interval R_q (which is easily obtained from the threshold q and independent of the correlations). The pdf is plotted in a scaled way, i. e., $R_q P_q(r)$ as a function of r/R_q. The figure shows that all three curves collapse. Accordingly, when we know the functional form of the pdf for one value of R_q, we can easily deduce its functional form also for very large R_q values which due to its poor statistics cannot be obtained directly from the data. This scaling is a very important property, since it allows one to make predictions also for rare events which otherwise are not accessible with meaningful statistics. When the data
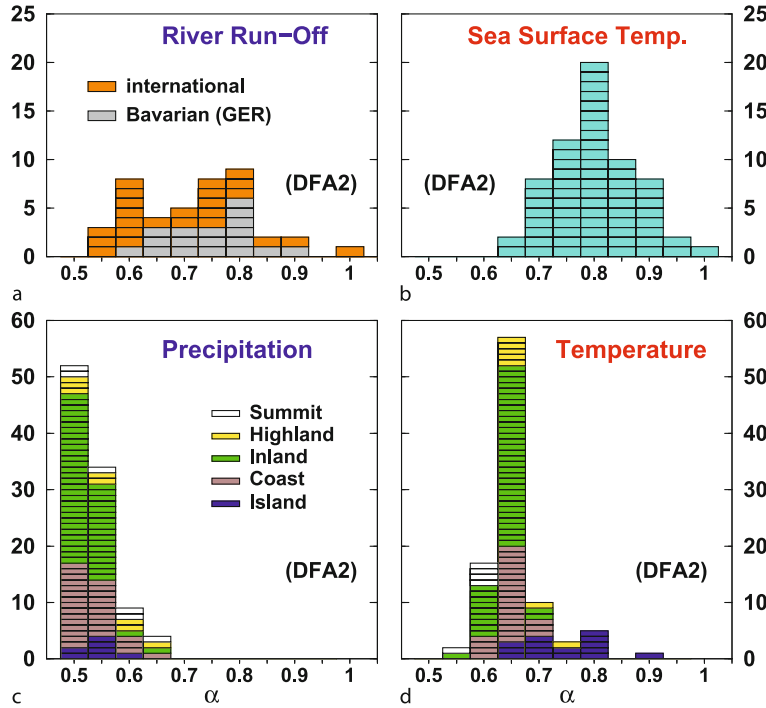
Fractal Geometry, A Brief Introduction to, Figure 25
Distribution of fluctuation exponents α for several kinds of climate records (from [53,66,67])
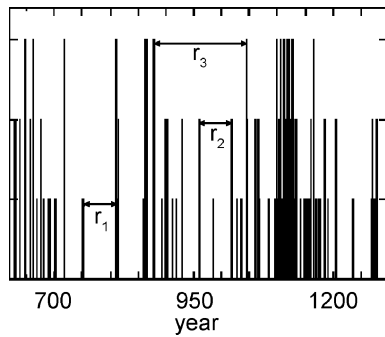
Fractal Geometry, A Brief Introduction to, Figure 26
Illustration of the return intervals for three equidistant thresh-old values q_1, q_2, q_3 for the water levels of the Nile at Roda (near Cairo, Egypt). One return interval for each threshold (quantile) q is indicated by *arrows*

Fractal Geometry, A Brief Introduction to, Figure 27
Probability density function of the return intervals in long-term correlated data, for three different return periods R_q, plotted in a scaled way. The *full line* is a stretched exponential, with expo-nent $\gamma = 0.4$ (after [73])

are shuffled, the long-term correlations are destroyed and the pdf becomes a simple exponential.

The functional form of the pdf is a quite natural extension of the uncorrelated case. The figure suggests that

$$\ln P_q(r) \sim -(r/R_q)^{\gamma} \tag{20}$$

i. e. simple stretched exponential behavior [72,73]. For γ approaching 1, the long-term correlations tend to vanish

and we obtain the simple exponential behavior characteristic for uncorrelated processes. For r well below R_q, however, there are deviations from the pure stretched exponential behavior. Closer inspection of the data shows that

for $r/R_q \ll$ the decay of the pdf is characterized by a power law, with the exponent $\gamma - 1$. This overall behavior does not depend crucially on the way the original data are distributed. In the cases shown here, the data had a Gaussian distribution, but similar results have been obtained also for exponential, power-law and log-normal distributions [74]. Indeed, the characteristic stretched exponential behavior of the pdf can also be seen in long historic and reconstructed records [73].

The form of the pdf indicates that return intervals both well below and well above their average value are considerably more frequent for long-term correlated data than for uncorrelated data. The distribution does not quantify, however, if the return intervals themselves are arranged in a correlated or in an uncorrelated fashion, and if clustering of rare events may be induced by long-term correlations.

To study this question, [73] and [74] have evaluated the autocorrelation function of the return intervals in synthetic long-term correlated records. They found that also the return intervals are arranged in a long-term correlated fashion, with the same exponent as the original data. Accordingly, a large return interval is more likely to be followed by a large one than by a short one, and a small return interval is more likely to be followed by a small one than by a large one, and this leads to clustering of events above some threshold q, including extreme events.
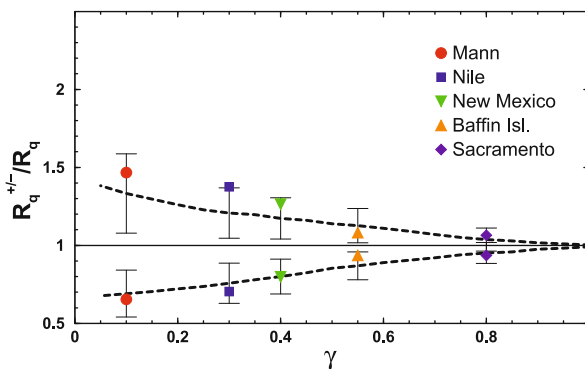
Fractal Geometry, A Brief Introduction to, Figure 28
Mean of the (conditional) return intervals that either follow a return interval below the median (*lower dashed line*) or above the median (*upper dashed line*), as a function of the correlation exponent γ, for five long reconstructed and natural climate records. The theoretical curves are compared with the corresponding values of the climate records (from *right* to *left*): The reconstructed run-offs of the Sacramento River, the reconstructed temperatures of Baffin Island, the reconstructed precipitation record of New Mexico, the historic water levels of the Nile and one of the reconstructed temperature records of the Northern hemisphere (Mann record) (after [73])

As a consequence of the long-term memory, the probability of finding a certain return interval depends on the preceding interval. This effect can be easily seen in synthetic data sets generated numerically, but not so well in climate records where the statistics is comparatively poor. To improve the statistics, we now only distinguish between two kinds of return intervals, "small" ones (below the median) and "large" ones (above the median), and determine the mean R_q^+ and R_q^- of those return intervals following a large $(+)$ or a small $(-)$ return interval. Due to scaling, R_q^+/R_q and R_q^-/R_q are independent of q. Figure 28 shows both quantities (calculated numerically for long-term correlated Gaussian data) as a function of the correlation exponent γ. The lower dashed line is R_q^-/R_q, the upper dashed line is R_q^+/R_q. In the limit of vanishing long-term memory, for $\gamma = 1$, both quantities coincide, as expected. Figure 28 also shows R_q^+/R_q and R_q^-/R_q for five climate records with different values of γ. One can see that the data agree very well, within the error bars, with the theoretical curves.

Long-Term Correlations in Financial Markets and Seismic Activity

The characteristic behavior of the return intervals, i. e. long-term correlations and stretched exponential decay, can also be observed in financial markets and seismic activity. It is well known (see, e. g. [56]) that the volatility of stocks and exchange rates is long-term correlated. Figure 29 shows that, as expected from the foregoing, also the return intervals between daily volatilities are long-term correlated, with roughly the same exponent γ as the original data [75]. It has further been shown [75] that also the pdfs show the characteristic behavior predicted above.

A further example where long-term correlations seem to play an important role, are earthquakes in certain bounded areas (e. g. California) in time regimes where the seismic activity is (quasi) stationary. It has been discovered recently by [76] that the magnitudes of earthquakes in Northern and Southern California, from 1995 until 1998, are long-term correlated with an exponent around $\gamma = 0.4$, and that also the return intervals between the earthquakes are long-term correlated with the same exponent. For the given exponential distribution of the earthquake magnitudes (following the Gutenberg–Richter law), the long-term correlations lead to a characteristic dependence on the scaled variable r/R_q which can explain, without any fit parameter, the previous results on the pdf of the return intervals by [77].
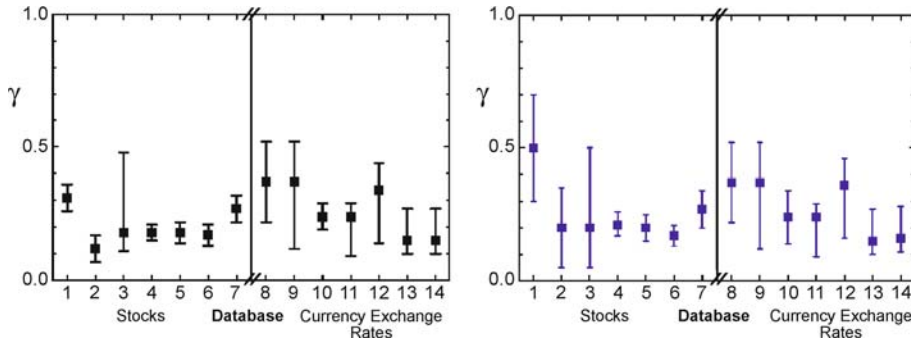
Fractal Geometry, A Brief Introduction to, Figure 29
Long-term correlation exponent γ for the daily volatility (*left*) and the corresponding return intervals (*right*). The studied commodities are (from *left* to *right*), the S&P 500 index, six stocks (IBM, DuPont, AT&T, Kodak, General Electric, Coca-Cola) and seven currency exchange rates (US\$ vs. Japanese Yen, British Pound vs. Swiss Franc, US\$ vs. Swedish Krona, Danish Krone vs. Australian \$, Danish Krone vs. Norwegian Krone, US\$ vs. Canadian \$ and US\$ vs. South African \$). Courtesy of Lev Muchnik

Multifractal Records

Many records do not exhibit a simple monofractal scaling behavior, which can be accounted for by a single scaling exponent. In some cases, there exist crossover (time-) scales s_x separating regimes with different scaling exponents, e. g. long-term correlations on small scales below s_x and another type of correlations or uncorrelated behavior on larger scales above s_x. In other cases, the scaling behavior is more complicated, and different scaling exponents are required for different parts of the series. In even more complicated cases, such different scaling behavior can be observed for many interwoven fractal subsets of the time series. In this case a multitude of scaling exponents is required for a full description of the scaling behavior, and a multifractal analysis must be applied (see, e. g., [78,79] and literature therein).

To see this, it is meaningful to extend Eqs. (19a) and (19b) by considering the more general average

$$F^q(s) = \frac{1}{K_s} \sum_{\nu=1}^{K_s} \left[F_\nu^2(s) \right]^{q/2} \tag{21}$$

with q between $-\infty$ and $+\infty$. For $q \ll -1$ the small fluctuations will dominate the sum, while for $q \gg 1$ the large fluctuations are dominant. It is reasonable to assume that the q-dependent average scales with s as

$$F^q(s) \sim s^{q\beta(q)} , \tag{22}$$

with $\beta(2) = \alpha$. Equation (22) generalizes Eq. (19b). If $\beta(q)$ is independent of q, we have $(F^q(s))^{1/q} \sim s^\alpha$, independent of q, and both large and small fluctuations scale the same. In this case, a single exponent is sufficient to characterize the record, which then is referred to as *monofractal*. If $\beta(q)$ is not identical to α, we have a *multifractal* [1,4,12]. In this case, the dependence of $\beta(q)$ on q characterizes the record. Instead of $\beta(q)$ one considers frequently the spectrum $f(\omega)$ that one obtains by Legendre transform of $q\beta(q)$, $\omega = \mathrm{d}(q\beta(q))/\mathrm{d}q$, $f(\omega) = q\omega - q\beta(q) + 1$. In the monofractal limit we have $f(\omega) = 1$.

For generating multifractal data sets, one considers mostly multiplicative random cascade processes, described, e. g., in [3,4]. In this process, the data set is obtained in an iterative way, where the length of the record doubles in each iteration. It is possible to generate random cascades with vanishing autocorrelation function ($C_x(s) = 0$ for $s \geq 1$) or with algebraically decaying autocorrelation functions ($C_x(s) \sim s^{-\gamma}$). Here we focus on a multiplicative random cascade with vanishing autocorrelation function, which is particularly interesting since it can be used as a model for the arithmetic returns $(P_i - P_{i-1})/P_i$ of daily stock closing prices P_i [80]. In the zeroth iteration $n = 0$, the data set (x_i) consists of one value, $x_1^{(n=0)} = 1$. In the nth iteration, the data $x_i^{(n)}$ consist of 2^n values that are obtained from

$$x_{2l-1}^{(n)} = x_l^{(n-1)} m_{2l-1}^{(n)}$$

and

$$x_{2l}^{(n)} = x_l^{(n-1)} m_{2l}^{(n)} , \tag{23}$$

where the multipliers m are independent and identically distributed (i.i.d.) random numbers with zero mean and unit variance. The resulting pdf is symmetric with log-normal tails, with vanishing correlation function $C_x(s)$ for $s \geq 1$.

It has been shown that in this case, the pdf of the return intervals decays by a power-law

$$P_q(r) \sim \left(\frac{r}{R_q}\right)^{-\delta(q)}, \tag{24}$$

where the exponent δ depends explicitly on R_q and seems to converge to a limiting curve for large data sets. Despite of the vanishing autocorrelation function of the original record, the autocorrelation function of the return intervals decays by a power law with a threshold-dependent exponent [80]. Obviously, these long-term correlations have been induced by the nonlinear correlations in the multifractal data set. Extracting the return interval sequence from a data set is a nonlinear operation, and thus the return intervals are influenced by the nonlinear correlations in the original data set. Accordingly, the return intervals in data sets without linear correlations are sensitive indicators for nonlinear correlations in the data records. The power-law dependence of $P_q(r)$ can be used for an improved risk estimation. Both power-law dependencies can be observed in economic and physiological records that are known to be multifractal [81].

Acknowledgments

Bibliography

1. Mandelbrot BB (1977) Fractals: Form, chance and dimension. Freeman, San Francisco; Mandelbrot BB (1982) The fractal geometry of nature. Freeman, San Francisco
2. Jones H (1991) Part 1: 7 chapters on fractal geometry including applications to growth, image synthesis, and neutral net. In: Crilly T, Earschaw RA, Jones H (eds) Fractals and chaos. Springer, New York
3. Peitgen H-O, Jürgens H, Saupe D (1992) Chaos and fractals. Springer, New York
4. Feder J (1988) Fractals. Plenum, New York
5. Vicsek T (1989) Fractal growth phenomena. World Scientific, Singapore
6. Avnir D (1992) The fractal approach to heterogeneous chemistry. Wiley, New York
7. Barnsley M (1988) Fractals everywhere. Academic Press, San Diego
8. Takayasu H (1990) Fractals in the physical sciences. Manchester University Press, Manchester
9. Schuster HG (1984) Deterministic chaos – An introduction. Physik Verlag, Weinheim
10. Peitgen H-O, Richter PH (1986) The beauty of fractals. Springer, Heidelberg
11. Stanley HE, Ostrowsky N (1990) Correlations and connectivity: Geometric aspects of physics, chemistry and biology. Kluwer, Dordrecht
12. Peitgen H-O, Jürgens H, Saupe D (1991) Chaos and fractals. Springer, Heidelberg
13. Bunde A, Havlin S (1996) Fractals and disordered systems. Springer, Heidelberg
14. Gouyet J-F (1992) Physique et structures fractales. Masson, Paris
15. Bunde A, Havlin S (1995) Fractals in science. Springer, Heidelberg
16. Havlin S, Ben-Avraham D (1987) Diffusion in disordered media. Adv Phys 36:695; Ben-Avraham D, Havlin S (2000) Diffusion and reactions in fractals and disordered systems. Cambridge University Press, Cambridge
17. Feigenbaum M (1978) Quantitative universality for a class of non-linear transformations. J Stat Phys 19:25
18. Grassberger P (1981) On the Hausdorff dimension of fractal attractors. J Stat Phys 26:173
19. Mandelbrot BB, Given J (1984) Physical properties of a new fractal model of percolation clusters. Phys Rev Lett 52:1853
20. Douady A, Hubbard JH (1982) Itération des polynômes quadratiques complex. CRAS Paris 294:123
21. Weiss GH (1994) Random walks. North Holland, Amsterdam
22. Flory PJ (1971) Principles of polymer chemistry. Cornell University Press, New York
23. De Gennes PG (1979) Scaling concepts in polymer physics. Cornell University Press, Ithaca
24. Majid I, Jan N, Coniglio A, Stanley HE (1984) Kinetic growth walk: A new model for linear polymers. Phys Rev Lett 52:1257; Havlin S, Trus B, Stanley HE (1984) Cluster-growth model for branched polymers that are "chemically linear". Phys Rev Lett 53:1288; Kremer K, Lyklema JW (1985) Kinetic growth models. Phys Rev Lett 55:2091
25. Ziff RM, Cummings PT, Stell G (1984) Generation of percolation cluster perimeters by a random walk. J Phys A 17:3009; Bunde A, Gouyet JF (1984) On scaling relations in growth models for percolation clusters and diffusion fronts. J Phys A 18:L285; Weinrib A, Trugman S (1985) A new kinetic walk and percolation perimeters. Phys Rev B 31:2993; Kremer K, Lyklema JW (1985) Monte Carlo series analysis of irreversible self-avoiding walks. Part I: The indefinitely-growing self-avoiding walk (IGSAW). J Phys A 18:1515; Saleur H, Duplantier B (1987) Exact determination of the percolation hull exponent in two dimensions. Phys Rev Lett 58:2325
26. Arapaki E, Argyrakis P, Bunde A (2004) Diffusion-driven spreading phenomena: The structure of the hull of the visited territory. Phys Rev E 69:031101
27. Witten TA, Sander LM (1981) Diffusion-limited aggregation, a kinetic critical phenomenon. Phys Rev Lett 47:1400
28. Meakin P (1983) Diffusion-controlled cluster formation in two, three, and four dimensions. Phys Rev A 27:604,1495
29. Meakin P (1988) In: Domb C, Lebowitz J (eds) Phase transitions and critical phenomena, vol 12. Academic Press, New York, p 335
30. Muthukumar M (1983) Mean-field theory for diffusion-limited cluster formation. Phys Rev Lett 50:839; Tokuyama M, Kawasaki K (1984) Fractal dimensions for diffusion-limited aggregation. Phys Lett A 100:337
31. Pietronero L (1992) Fractals in physics: Applications and theoretical developments. Physica A 191:85

32. Meakin P, Majid I, Havlin S, Stanley HE (1984) Topological properties of diffusion limited aggregation and cluster-cluster aggregation. Physica A 17:L975

33. Mandelbrot BB (1992) Plane DLA is not self-similar; is it a fractal that becomes increasingly compact as it grows? Physica A 191:95; see also: Mandelbrot BB, Vicsek T (1989) Directed recursive models for fractal growth. J Phys A 22:L377

34. Schwarzer S, Lee J, Bunde A, Havlin S, Roman HE, Stanley HE (1990) Minimum growth probability of diffusion-limited aggregates. Phys Rev Lett 65:603

35. Meakin P (1983) Formation of fractal clusters and networks by irreversible diffusion-limited aggregation. Phys Rev Lett 51:1119; Kolb M (1984) Unified description of static and dynamic scaling for kinetic cluster formation. Phys Rev Lett 53:1653

36. Stauffer D, Aharony A (1992) Introduction to percolation theory. Taylor and Francis, London

37. Kesten H (1982) Percolation theory for mathematicians. Birkhauser, Boston

38. Grimmet GR (1989) Percolation. Springer, New York

39. Song C, Havlin S, Makse H (2005) Self-similarity of complex networks. Nature 433:392

40. Havlin S, Blumberg-Selinger R, Schwartz M, Stanley HE, Bunde A (1988) Random multiplicative processes and transport in structures with correlated spatial disorder. Phys Rev Lett 61:1438

41. Voss RF (1985) In: Earshaw RA (ed) Fundamental algorithms in computer graphics. Springer, Berlin, p 805

42. Coleman PH, Pietronero L (1992) The fractal structure of the universe. Phys Rep 213:311

43. Kaye BH (1989) A random walk through fractal dimensions. Verlag Chemie, Weinheim

44. Turcotte DL (1997) Fractals and chaos in geology and geophysics. Cambridge University Press, Cambridge

45. Hurst HE, Black RP, Simaika YM (1965) Long-term storage: An experimental study. Constable, London

46. Mandelbrot BB, Wallis JR (1969) Some long-run properties of geophysical records. Wat Resour Res 5:321–340

47. Koscielny-Bunde E, Kantelhardt JW, Braun P, Bunde A, Havlin S (2006) Long-term persistence and multifractality of river runoff records: Detrended fluctuation studies. Hydrol J 322:120–137

48. Mudelsee M (2007) Long memory of rivers from spatial aggregation. Wat Resour Res 43:W01202

49. Livina VL, Ashkenazy Y, Braun P, Monetti A, Bunde A, Havlin S (2003) Nonlinear volatility of river flux fluctuations. Phys Rev E 67:042101

50. Koscielny-Bunde E, Bunde A, Havlin S, Roman HE, Goldreich Y, Schellnhuber H-J (1998) Indication of a universal persistence law governing athmospheric variability. Phys Rev Lett 81:729–732

51. Pelletier JD, Turcotte DL (1999) Self-affine time series: Application and models. Adv Geophys 40:91

52. Talkner P, Weber RO (2000) Power spectrum and detrended fluctuation analysis: Application to daily temperatures. Phys Rev E 62:150–160

53. Eichner JF, Koscielny-Bunde E, Bunde A, Havlin S, Schellnhuber H-J (2003) Power-law persistence and trends in the atmosphere: A detailed study of long temperature records. Phys Rev E 68:046133

54. Király A, Bartos I, Jánosi IM (2006) Correlation properties of daily temperature anormalies over land. Tellus 58A(5):593–600

55. Santhanam MS, Kantz H (2005) Long-range correlations and rare events in boundary layer wind fields. Physica A 345:713–721

56. Liu YH, Cizeau P, Meyer M, Peng C-K, Stanley HE (1997) Correlations in economic time series. Physica A 245:437; Liu YH, Gopikrishnan P, Cizeau P, Meyer M, Peng C-K, Stanley HE (1999) Statistical properties of the volatility of price fluctuations. Phys Rev E 60:1390

57. Peng C-K, Mietus J, Hausdorff JM, Havlin S, Stanley HE, Goldberger AL (1993) Long-range anticorrelations and non-gaussian behavior of the heartbeat. Phys Rev Lett 70:1343–1346

58. Bunde A, Havlin S, Kantelhardt JW, Penzel T, Peter J-H, Voigt K (2000) Correlated and uncorrelated regions in heart-rate fluctuations during sleep. Phys Rev Lett 85:3736

59. Leland WE, Taqqu MS, Willinger W, Wilson DV (1994) On the self-similar nature of Ethernet traffic. IEEE/Transactions ACM Netw 2:1–15

60. Kantelhardt JW, Koscielny-Bunde E, Rego HA, Bunde A, Havlin S (2001) Detecting long-range correlations with detrended fluctuation analysis. Physica A 295:441

61. Rybski D, Bunde A, Havlin S, Von Storch H (2006) Long-term persistence in climate and the detection problem. Geophys Res Lett 33(6):L06718

62. Rybski D, Bunde A (2008) On the detection of trends in long-term correlated records. Physica A

63. Giese E, Mossig I, Rybski D, Bunde A (2007) Long-term analysis of air temperature trends in Central Asia. Erdkunde 61(2):186–202

64. Govindan RB, Vjushin D, Brenner S, Bunde A, Havlin S, Schellnhuber H-J (2002) Global climate models violate scaling of the observed atmospheric variability. Phys Rev Lett 89:028501

65. Vjushin D, Zhidkov I, Brenner S, Havlin S, Bunde A (2004) Volcanic forcing improves atmosphere-ocean coupled general circulation model scaling performance. Geophys Res Lett 31:L10206

66. Monetti A, Havlin S, Bunde A (2003) Long-term persistence in the sea surface temperature fluctuations. Physica A 320:581–589

67. Kantelhardt JW, Koscielny-Bunde E, Rybski D, Braun P, Bunde A, Havlin S (2006) Long-term persistence and multifractality of precipitation and river runoff records. Geophys J Res Atmosph 111:1106

68. Bunde A, Kropp J, Schellnhuber H-J (2002) The science of disasters – climate disruptions, heart attacks, and market crashes. Springer, Berlin

69. Pfisterer C (1998) Wetternachhersage, 500 Jahre Klimavariationen und Naturkatastrophen 1496–1995. Verlag Paul Haupt, Bern

70. Glaser R (2001) Klimageschichte Mitteleuropas. Wissenschaftliche Buchgesellschaft, Darmstadt

71. Mudelsee M, Börngen M, Tetzlaff G, Grünwald U (2003) No upward trends in the occurrence of extreme floods in Central Europe. Nature 425:166

72. Bunde A, Eichner J, Havlin S, Kantelhardt JW (2003) The effect of long-term correlations on the return periods of rare events. Physica A 330:1

73. Bunde A, Eichner J, Havlin S, Kantelhardt JW (2005) Long-term memory: A natural mechanism for the clustering of extreme events and anomalous residual times in climate records. Phys Rev Lett 94:048701

74. Eichner J, Kantelhardt JW, Bunde A, Havlin S (2006) Extreme value statistics in records with long-term persistence. Phys Rev E 73:016130
75. Yamasaki K, Muchnik L, Havlin S, Bunde A, Stanley HE (2005) Scaling and memory in volatility return intervals in financial markets. PNAS 102:26 9424–9428
76. Lennartz S, Livina VN, Bunde A, Havlin S (2008) Long-term memory in earthquakes and the distribution of interoccurrence times. Europ Phys Lett 81:69001
77. Corral A (2004) Long-term clustering, scaling, and universality in the temporal occurrence of earthquakes. Phys Rev Lett 92:108501
78. Stanley HE, Meakin P (1988) Multifractal phenomena in physics and chemistry. Nature 355:405
79. Ivanov PC, Goldberger AL, Havlin S, Rosenblum MG, Struzik Z, Stanley HE (1999) Multifractality in human heartbeat dynamics. Nature 399:461
80. Bogachev MI, Eichner JF, Bunde A (2007) Effect of nonlinear correlations on the statistics of return intervals in multifractal data sets. Phys Rev Lett 99:240601
81. Bogachev MI, Bunde A (2008) Memory effects in the statistics of interoccurrence times between large returns in financial records. Phys Rev E 78:036114; Bogachev MI, Bunde A (2008) Improving risk extimation in multifractal records: Applications to physiology and financing. Preprint

Fractal Growth Processes

LEONARD M. SANDER
Physics Department and Michigan Center for Theoretical Physics, The University of Michigan, Ann Arbor, USA

Article Outline

Glossary

Fractal A fractal is a geometric object which is self-similar and characterized by an effective dimension which is not an integer.

Multifractal A multifractal measure is a non-negative real function) defined on a *support* (a geometric region) which has a spectrum of scaling exponents.

Diffusion-limited aggregation Diffusion-limited aggregation (DLA) is a discrete model for the irreversible growth of a cluster. The rules of the model involve a sequence of random walkers that are incorporated into a growing aggregate when they wander into contact with one of the previously aggregated walkers.

Dielectric breakdown model The dielectric breakdown model (DBM) is a generalization of DLA where the probability to grow is proportional to a power of the diffusive flux onto the aggregate. If the power is unity, the model is equivalent to DLA: in this version it is called Laplacian growth.

Harmonic measure If a geometric object is thought of as an isolated grounded conductor of fixed charge, the distribution of electric field on its surface is the harmonic measure. The harmonic measure of a DLA cluster is the distribution of growth probability on the surface.

Definition of the Subject

Fractal growth processes are a class of phenomena which produce self-similar, disordered objects in the course of development far from equilibrium. The most famous of these processes involve growth which can be modeled on the large scale by the diffusion-limited aggregation algorithm of Witten and Sander [1]. DLA is a paradigm for pattern formation modeling which has been very influential.

The algorithm describes growth limited by diffusion: many natural processes fall in this category, and the salient characteristics of clusters produced by the DLA algorithm are observed in a large number of systems such as electrodeposition clusters, viscous fingering patterns, colonies of bacteria, dielectric breakdown patterns, and patterns of blood vessels.

At the same time the DLA algorithm poses a very rich problem in mathematical physics. A full "solution" of the DLA problem, in the sense of a scaling theory that can predict the important features of computer simulations is still lacking. However, the problem shows many features that resemble thermal continuous phase transitions, and a number of partially successful approaches have been given. There are deep connections between DLA in two dimensions and theories such as Loewner evolution that use conformal maps.

Introduction

In the 1970s Benoit Mandelbrot [2] developed the idea of *fractal geometry* to unify a number of earlier studies of irregular shapes and natural processes. Mandelbrot focused on a particular set of such objects and shapes, those that are *self-similar*, i. e., where a part of the object is identical

(either in shape, or for an ensemble of shapes, in distribution) to a larger piece. He dubbed these *fractals*. He noted the surprising ubiquity of self-similar shapes in nature.

Of particular interest to Mandelbrot and his collaborators were incipient percolation clusters [3]. These are the shapes generated when, for example, a lattice is diluted by cutting bonds at random until a cluster of connected bonds just reaches across a large lattice. This model has obvious applications in physical processes such as transport in random media. The model has a non-trivial mapping to a statistical model [4] and can be treated by the methods of equilibrium statistical physics. It is likely that percolation processes account for quite a few observations of fractals in nature.

In 1981 Tom Witten and the present author observed that a completely different type of process surprisingly appears to make fractals [1,5]. These are unstable, irreversible, growth processes which we called *diffusion-limited aggregation* (DLA). DLA is a kinetic process which is not related to equilibrium statistical physics, but rather defined by growth rules. The rules idealize growth limited by diffusion: in the model there are random walking "particles" which attach irreversibly to a single cluster made up of previously aggregated particles. As we will see, quite a few natural processes can be described by DLA rules, and DLA-like clusters are reasonably common in nature. The Witten–Sander paper and subsequent developments unleashed a large amount of activity – the original work has been cited more than 2700 times as of this writing. The literature in this area up to 1998 was reviewed in a very comprehensive manner by T. Vicsek [6] and P. Meakin [7]. See also the chapter by the present author in [8]. For non-technical reviews see [9,10,11].

There are three major areas where self-similar shapes arising from non-equilibrium processes have been studied. The first is the related to the original DLA algorithm. The model may be seen as an idealization of solidification of an amorphous substance. The study of this simple-seeming model is quite unexpectedly rich, and quite difficult to treat theoretically. It will be our focus in this article. We will review the early work, but emphasize developments since [7].

Meakin [12] and Kolb, et al. [13] generalized DLA to consider cluster-cluster or colloid aggregation. In this process particles aggregate when they come into contact, but the clusters so formed are mobile, and themselves aggregate by attaching to each other. This is an idealization of colloid or coagulated aerosol formation. This model also produces fractals but this is not really a surprise: at each stage, clusters of similar size are aggregating, and the result is an approximately hierarchical object. This model is quite important in applications in colloid science. The interested reader should consult [6,14].

A third line of work arose from studies of ballistic aggregation [15] and the Eden model [16]. In the former case particles attach to an aggregate after moving in straight paths, and in the latter, particles are simply added to the surface of an aggregate at any available site, with equal probability. These models give rise to non-fractal clusters with random rough surfaces. The surfaces have scaling properties which are often characterized at the continuum level by a stochastic partial differential equation proposed by Kardar, Parisi, and Zhang [17]. For accounts of this area the reader can consult [18,19,20].

The remainder of this article is organized as follows: we first briefly review fractals and multifractals. Then we give details about DLA and related models, numerical methods, and applications of the models. The major development that has fueled a remarkable revival of interest in this area was the work of Hastings and Levitov [21] who related two-dimensional DLA to a certain class of conformal maps. Developments of this theme is the subject of the subsequent section. Then we discuss the question of the distribution of growth probabilities on the cluster surface, and finally turn to scaling theories of DLA.

Fractals and Multifractals

The kind of object that we deal with in this article is highly irregular; see Fig. 1. We think of the object as being made up of a large number, N, of units, and to be of overall linear dimension R. In growth processes the objects are formed by adding the units according to some dynamics. We will call the units "particles", and take their size to be a.

Such patterns need not be merely random, but can have well-defined scaling properties in the regime $a \ll R$. The picture is of a geometric object in two dimensions, but the concepts we introduce here are often applied to more abstract spaces. For example, a strange attractor in phase space often has the kind of scaling properties described here.

In order to characterize the geometry of such complex objects, we first cover the points in question with a set of $n(l)$ "boxes" of fixed size, l such that $a \ll l \ll R$. Clearly, for a smooth curve the product $ln(l)$ approaches a limit (the length of the curve) as $l \to 0$. This is a number of order R. For a planar region with smooth boundaries $l^2 n(l)$ approaches the area, of order R^2. The objects of ordinary geometry in d dimensions have measures given by the limit of $l^d n(l)$. For an object with many scales (a fractal), in general none of these relations hold. Rather, the product $l^D n(l)$ approaches a limit with D not necessarily an inte-
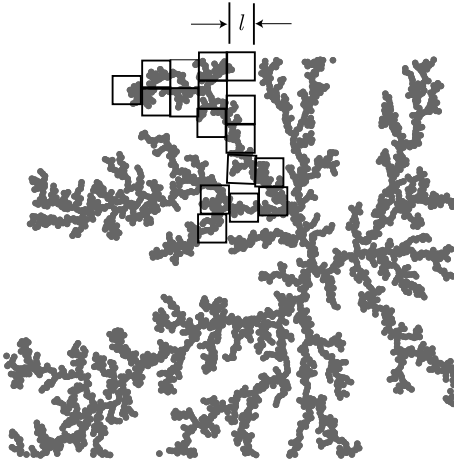
Fractal Growth Processes, Figure 1
A partial covering of a pattern with boxes. Smaller boxes reveal smaller scales of the pattern. For a pattern with many scales (like this one) there is a non-trivial scaling between the box size, *l* and the number of boxes

ger; D is called the (similarity) fractal dimension. Said another way, we define the fractal dimension by:

$$n(l) \propto (R/l)^D . \tag{1}$$

For many simple examples of mathematical objects with non-integer values of D see [2,6].

For an infinite fractal there are no characteristic lengths. For a finite size object there is a characteristic length, the overall scale, R. This can be taken to be any measure of the size such as the radius of gyration or the extremal radius – all such lengths must be proportional.

It is useful to generalize this definition in two ways. First we consider not only a geometric object, but also a *measure*, that is a non-negative function μ defined on the points of the object such that $\int d\mu = 1$. For the geometry, we take the measure to be uniform on the points. However, for the case of growing fractals, we could also consider the growth probability at a point. As we will see, this is very non-uniform for DLA. Second, we define a sequence of generalized dimensions. If we are interested in geometry, we denote the mass of the object covered by box i by p_i. For an arbitrary measure, we define:

$$p_i = \int d\mu , \tag{2}$$

where the integral is over the box labeled i. Then, following [22,23] we define a partition function for the p_i:

$$\chi(q) = \sum_{i=1}^{n} p_i^q , \tag{3}$$

where q is a real number. For an object with well-defined scaling properties we often find that χ scales with l in the following way as $l/R \to 0$:

$$\chi(q) \propto (R/l)^{-\tau(q)} \equiv (R/l)^{-(q-1)D_q} ; \tag{4}$$
$$\tau(q) = (q - 1)D_q .$$

Objects with this property are called fractals if all the D_q are the same. Otherwise they are *multifractals*.

Some of the D_q have special significance. The similarity (or box-counting) dimension mentioned above is D_0 since in this case $\chi = n$. If we take the limit $q \to 1$ we have the information dimension of dynamical systems theory:

$$D_1 = \left.\frac{d\tau}{dq}\right|_{q=1} = \sum_i p_i \frac{\ln p_i}{\ln l} . \tag{5}$$

D_2 is called the correlation dimension since p_i^2 measures the probability that two points are close together, i. e. the number of pairs within distance l. This interpretation gives rise to a popular way to measure D_2. If we suppose that the structure is homogeneous, then the number of pairs of points can be found by focusing on any point, and drawing a d-dimensional disk of radius r around it. The number of other points in the disk will scale as r^{D_2}. For DLA, it is common to simply count the number of points within radius r of the origin, or, alternatively, the dependence of some mean radius, R on the total number of aggregated particles, N, that is $N \propto R^{D_2}$ This method is closely related to the Grassberger–Procaccia correlation integral [24].

For a simple fractal all of the D_q are the same, and we use the symbol D. If the generalized dimensions differ, then we have a multifractal. Multifractals were introduced by Mandelbrot [25] in the context of turbulence. In the context of fractal growth processes, the clusters themselves are usually simple fractals. They are the support for a multifractal measure, the growth probability.

We need another bit of formalism. It is useful to look at the fractal measure and note how the p_i scale with l/R. Suppose we assume a power-law form, $p_i \propto (l/R)^\alpha$, where there are different values of α for different parts of the measure. Also, suppose that we make a histogram of the α, and look at the parts of the support on which we have the same scaling. It is natural to adopt a form like Eq. (1) for the size of these sets, $(l/R)^{-f(\alpha)}$. (It is natural to think of $f(\alpha)$ as the fractal dimension of the set on which the measure has exponent α, but this is not quite right because f can be negative due to ensemble averaging.) Then it is not hard to show [23] that $\alpha, f(\alpha)$ are related to $q, \tau(q)$ by a Legendre transform:

$$f(\alpha) = q\frac{d\tau}{dq} - \tau ; \quad \alpha = \frac{d\tau}{dq} . \tag{6}$$

Aggregation Models

In this section we review aggregation models of the DLA type, and their relationship to certain continuum processes. We look at practical methods of simulation, and exhibit a few applications to physical and biological processes.

DLA

The original DLA algorithm [1,5] was quite simple: on a lattice, declare that the point at the origin is the first member of the cluster. Then launch a random walker from a distant point and allow it to wander until it arrives at a neighboring site to the origin, and attach it to the cluster, i. e., freeze its position. Then launch another walker and let it attach to one of the two previous points, and so on. The name, diffusion-limited aggregation, refers to the fact that random walkers, i. e., diffusing particles, control the growth. DLA is a simplified view of a common physical process, growth limited by diffusion.

It became evident that for large clusters the overall shape is dependent on the lattice type [26,27], that is, DLA clusters are deformed by lattice anisotropy. This is an interesting subject [26,28,29,30,31] but most modern work is on DLA clusters without anisotropy, *off-lattice* clusters. The off-lattice algorithm is similar to the original one: instead of a random walk on a lattice the particle is considered to have radius a. For each step of the walk the particle moves its center from the current position to a random point on its perimeter. If it overlaps a particle of the current cluster, it is backed up until it just touches the cluster, and frozen at that point. Then another walker is launched. A reasonably large cluster grown in two dimensions is shown in Fig. 2, along with a smaller three-dimensional example. Most of the work on DLA has been for two dimensions, but dimensions up to 8 have been considered [32].

Patterns like the one in Fig. 2 have been analyzed for fractal properties. Careful measurements of both D_0 and D_2 (using the method, mentioned above, of fitting $n(r)$ to r^{D_2}) give the same fractal dimension, D=1.71 [32,33,34]; see Fig. 3. There is some disagreement about the next digit. There have been suggestions that DLA is a mass multifractal [35], but most authors now agree that all of the D_q are the same for the mass distribution. For three dimensions $D \approx 2.5$ [31,32], and for four dimensions $D \approx 3.4$ [32].

However, some authors [36,37] have claimed that plane DLA is not a self-similar fractal at all, and that the fractal dimension will drift towards 2 as the number of particles increases. More recent work based on conformal maps [38] has cast doubt on this. We will return to this point below.
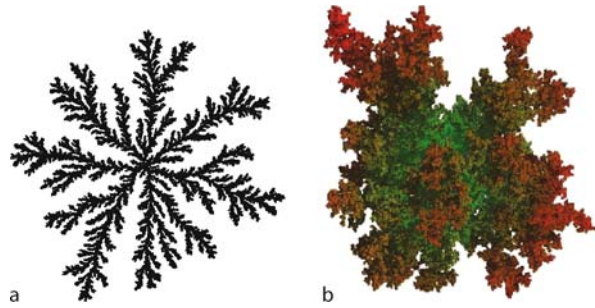
Fractal Growth Processes, Figure 2
a A DLA cluster of 50,000,000 particles produced with an off-lattice algorithm by E. Somfai. **b** A three-dimensional off-lattice cluster. Figure courtesy of R. Ball
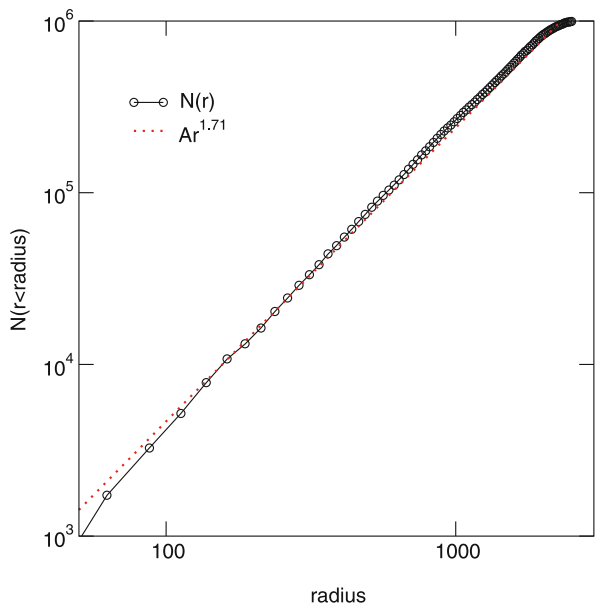
Fractal Growth Processes, Figure 3
The number of particles inside of radius *r* for a large DLA cluster. This plot gives an estimate of D_2

Plane DLA clusters can be grown in restricted geometries, see Fig. 4. The shape of such clusters is an interesting problem in pattern formation [39,40,41,42]. It was long thought that DLA grown in a channel had a different fractal dimension than in radial geometry [43,44,45]. However, more careful work has shown that the dimensions are the same in the two cases [34].

Laplacian Growth and DBM

Suppose we ask how the cluster of Fig. 2 gets to be rough. A simple answer is that if we already have a rough shape, it is quite difficult for a random walker to penetrate a nar-
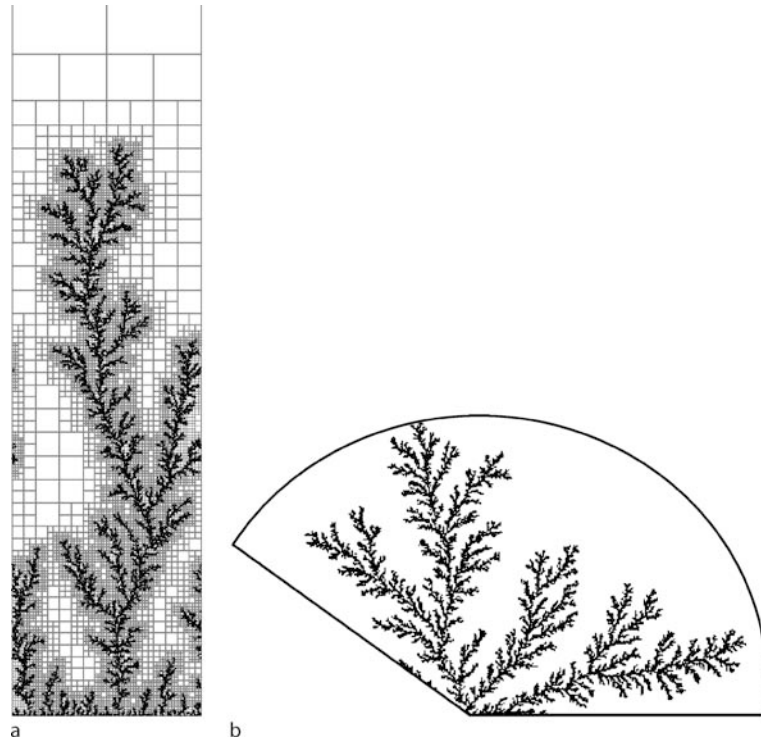
Fractal Growth Processes, Figure 4
DLA grown in a channel (**a**) and a wedge (**b**). The boxes in **a** are the hierarchical maps of the Sect. "Numerical Methods". Figures due to E. Somfai

row channel. (Just how difficult is treated in the Sect. "Harmonic Measure", below) The channels don't fill in, and the shape might be preserved. But we can also ask why a a smooth outline, e.g. a disk, does not continue to grow smoothly. In fact, it is easy to test that any initial condition is soon forgotten in the growth [5]. If we start with a smooth shape it roughens immediately because of a *growth instability* intrinsic to diffusion-limited growth. This instability was discovered by Mullins and Sekerka [46] who used a statement of the problem of diffusion-limited growth in continuum terms: this is known as the Stefan problem (see [47,48]), and is the standard way to idealize crystallization in the diffusion-limited case.

The Stefan problem goes as follows: suppose that we have a density $\phi(\mathbf{r}, t)$ of particles that diffuse until they reach the growing cluster where they deposit. Then we have:

$$\frac{\partial \phi}{\partial t} = \nu \nabla^2 \phi \tag{7}$$

$$\frac{\partial \phi}{\partial n} \propto \nu_n \tag{8}$$

That is, ϕ should obey the diffusion equation; ν is the diffusion constant. The normal growth velocity, ν_n, of the in-

terface is proportional to the flux onto the surface, $\partial \phi / \partial n$. However the term $\partial \phi / \partial t$ is of order $\nu \partial \phi / \partial x$, where ν is a typical growth velocity. Now $|\nabla^2 \phi| \approx (\nu/D)|\partial \phi / \partial n|$. In the DLA case we launch one particle at a time, so that the velocity goes to zero. Hence Eq. (7) reduces to the Laplace equation,

$$\nabla^2 \phi = 0 \tag{9}$$

Since the cluster absorbs the particles, we should think of it as having $\phi = 0$ on the surface. We are to solve an *electrostatics* problem: the cluster is a grounded conductor with fixed electric flux far away. We grow by an amount proportional to the electric field at each point on the surface. This is called the quasi-static or *Laplacian* growth regime for deterministic growth. A linear stability analysis of these equations gives the Mullins–Sekerka instability [46]. The qualitative reason for the instability is that near the tips of the cluster the contours of ϕ are compressed so that $\partial \phi / \partial n$, the growth rate, is large. Thus tips grow unstably. We expect DLA to have a growth instability.

However, we can turn the argument, and use these observations to give a restatement of the DLA algorithm in continuum terms: we calculate the electric field on the sur-

face of the aggregate, and interpret Eq. (8) as giving the distribution of the *growth probability*, p, at a point on the surface. We add a particle with this probability distribution, recalculate the potential using Eq. (9) and continue. This is called Laplacian growth. Simulations of Laplacian growth yield the same sort of clusters as the original discrete algorithm. (Some authors use the term Laplacian growth in a different way, to denote deterministic growth according to the Stefan model without surface tension [49].)

DLA is thus closely related to one of the classic problems of mathematical physics, dendritic crystal growth in the quasistatic regime. However, it is not quite the same for several reasons: DLA is dominated by noise, whereas the Stefan problem is deterministic. Also, the boundary conditions are different [48]: for a crystal, if we interpret u as $T - T_m$, where T is the temperature, and T_m the melting temperature, we have $\phi = 0$ only on a flat surface. On a curved surface we need $\phi \propto \gamma \kappa$ where γ is the surface stiffness, and κ is the curvature. The surface tension acts as a regularization which prevents the Mullins–Sekerka instability from producing sharp cusps [50]. In DLA the regularization is provided by the finite particle size. And, of course, crystals have anisotropy in the surface tension.

There is another classic problem very similar to this, that of viscous fingering in Hele–Shaw flow [48]. This is the description of the displacement of an incompressible viscous fluid by an inviscid one: the "bubble" of inviscid fluid plays the role of the cluster, ϕ is the difference of pressures in the viscous fluid and the bubble, and the Laplace equation is the direct result of incompressibility, $\nabla \cdot \mathbf{v} = 0$ and D'Arcy's law, $\mathbf{v} = k \nabla \phi$ where k is the permeability, and $\mathbf{v}$ the fluid velocity [51].

These considerations led Niemeyer, Pietronero, and Weismann [52] to a clever generalization of Laplacian growth. They were interested in dielectric breakdown with ϕ representing a real electrostatic potential. This is known to be a threshold process so that we expect that the breakdown probability is non-linear in $\partial\phi/\partial n$. To generalize they chose:

$$p \propto \left(\frac{\partial \phi}{\partial n} \right)^{\eta} , \qquad (10)$$

where η is a non-negative real number.

There are some interesting special cases for this model. For $\eta = 0$ each growth site is equally likely to be used. This is the Eden model [16]. For $\eta = 1$ we have the Laplacian growth version of DLA, and for larger η we get a higher probability to grow at the tips so that the aggregates are more spread out (as in a real dielectric breakdown pattern like atmospheric lightning). There is a remarkable fact which was suggested numerically in [53] and confirmed more recently [54,55]: for $\eta > 4$ the aggregate prefers growth at tips so much that it becomes essentially linear and non-fractal. DBM patterns for small numbers of particles are shown in Fig. 5.

A large number of variants of the basic model have been proposed such as having the random walkers perform Lévy flights, having a variable particle size, imposing a drift on the random walkers, imposing anisotropy in attachment, etc. For references on these and other variants, see [7]. There are two qualitative results from these studies that are worth mentioning here: In the presence of drift, DLA clusters cross over to (non-fractal) Eden-like clusters [56,57]. And, anisotropy deforms the shape of the clusters on the large scale, as we mentioned above.

Numerical Methods

In the foregoing we have talked about the algorithm for DLA, but we have not described what is actually done to compute the positions of 50,000,000 particles. This is a daunting computational task, and considerable ingenu-
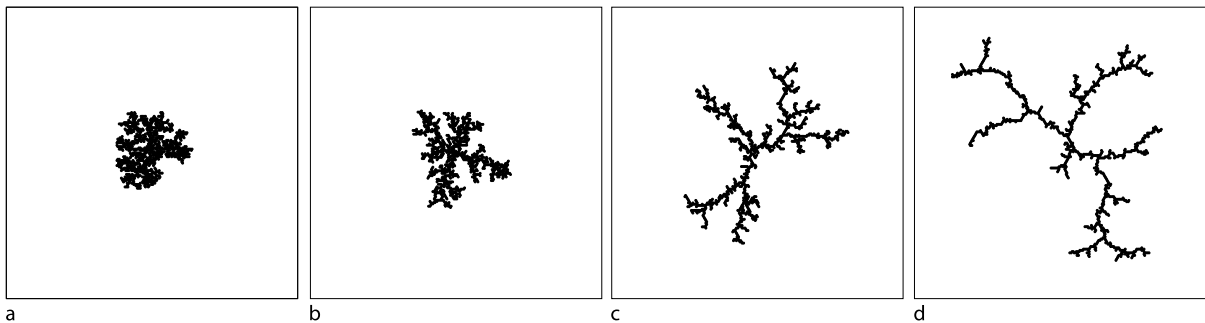
Fractal Growth Processes, Figure 5
DBM patterns for 1000 particles on a triangular lattice. **a** $\eta = 0.5$ **b** $\eta = 1$. This is a small DLA cluster. **c** $\eta = 2$. **d** $\eta = 3$
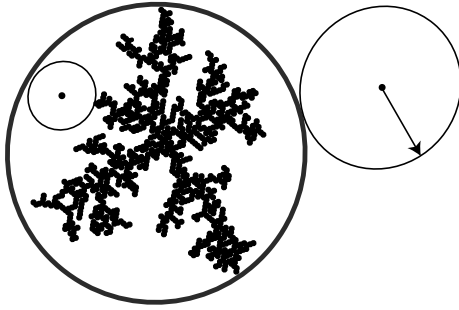
Fractal Growth Processes, Figure 6
A cluster is surrounded by a starting circle (*gray*). A random walker outside the circle can safely take steps of length of the *black circle*. If the walker is at any point, it can walk on a circle equal in size to the distance to the nearest point of the cluster. However, finding that circle appears to require a time-consuming search

ity has been devoted to making efficient algorithms. The techniques are quite interesting in their own right, and have been applied to other types of simulation.

The first observation to make is that the algorithm is an idealization of growth due to particles wandering into contact with the aggregate from far away. However, it is not necessary to start the particles far away: they arrive at the aggregate with uniform probability on a circle which just circumscribes all of the presently aggregated particles; thus we can start the particles there. As the cluster grows, the starting circle grows. This was already done in the original simulations [1,5].

However, particles can wander in and then out of the starting circle without attaching. These walkers must be followed, and could take a long time to find the cluster again. However, since there is no matter outside, it is possible to speed up the algorithm by noting that if the random walker takes *large steps* it will still have the same probability distribution, provided that it cannot encounter any matter. For example, if it walks onto the circumference of a circle that just reaches the starting circle, it will have the correct distribution. The radius of this circle is easy to find. This observation, due to P. Meakin, is the key to what follows: see Fig. 6.

We should note that the most efficient way to deal with particles that wander away is to return the particle to the starting circle in one step using the Green's function for a point charge outside an absorbing circle. A useful algorithm to do this is given in [58]. However, the idea of taking big steps is still a good one because there is a good deal of empty space *inside* the starting circle. If we could take steps in this empty space (see Fig. 6) we could again speed up the algorithm. The trick is to efficiently find the largest

circle centered on the random walker that has no point of the aggregate within it.

One could imagine simply doing a spiral search from the current walker position. This technique has actually been used in a completely different setting, that of Kinetic Monte Carlo simulations of surface growth in materials science [59]. For the case of DLA an even more efficient method, the method of hierarchical maps, was devised [26]. It was extended and applied to higher dimensions and off-lattice in [32], and is now the standard method.

One version of the idea is illustrated in Fig. 4a for the case of growth in a channel. What is shown is an adaptively refined square mesh. The cluster is covered with a square – a map. The square is subdivided into four smaller squares, and each is further divided, but only if the cluster is closer to it than half of the side of the square. The subdivision continues only up to a predefined maximum depth so that the smallest maps are a few particle diameters. All particles of the cluster will be in one of the smallest maps: a list of the particles is attached to these maps.

As the cluster grows the maps are updated. Each time a particle is added to a previously empty smallest map, the neighboring maps (on all levels) are checked to see whether they satisfy the rule. If not, they are subdivided until they do. When a walker lands, we find the smallest map containing the point. If this map is not at the maximum depth, then the particle is far away from any matter, and half the side of the map is a lower estimate of the walker's distance from the cluster. If, on the other hand, the particle lands in a map of maximum depth, then it is close to the cluster. The particle lists of the map and of the neighboring smallest size maps can be checked to calculate the exact distance from the cluster. Either way, the particle is enclosed in an empty circle of known radius, and can be brought to the perimeter of the circle in one step. Note that updating the map means that there is only a search for the cluster if we are in the smallest map. Empirically, the computational time, T for an N particle cluster obeys $T \sim N^{1.1}$, and the memory is linear in N. A more recent version of the algorithm for three-dimensional growth uses a covering with balls rather than cubes [31].

For simulations of the Laplacian growth version of the situation is quite different, and simulations are much slower. The reason is that a literal interpretation of the algorithm requires that the Laplace equation be solved each time a particle is added, for example, by the relaxation or boundary-integral method. This is how corresponding simulations for viscous fingering are done [60]. For DBM clusters the growth step requires taking a power of the electric field at the surface.

It is possible to make DBM clusters using random walkers [61], but the method is rather subtle. It involves estimating the growth probability at a point by figuring out the "age" of a site, i. e., the time between attachments. This can be converted into an estimate of the growth probability. This algorithm is quite fast.

There is another class of methods involving conformal maps. These are slow methods: the practical limit of the size of clusters that can be grown this way is about 50,000. However, this method calculates the growth probability as it goes along, and thus is of great interest. Conformal mapping is discussed in the Sect. "Loewner Evolution and the Hastings–Levitov Scheme", below.

Selected Applications

Probably the most important impact of the DLA model has been the realization that diffusion-limited growth naturally gives rise to tenuous, branched objects. Of course, in the context of crystallization, this was understood in the centuries-long quest to understand the shape of snowflakes. However, applications of DLA gave rise to a unification of this with the study of many other physical applications. Broad surveys are given in [6,7,10]. Here we will concentrate on a few illustrative examples.

Rapid crystallization in a random environment is the most direct application of DLA scheme. One particularly accessible example is the growth of islands on surfaces in molecular beam epitaxy experiments. In the proper growth conditions it is easy to see patterns dominated by the Mullins–Sekerka instability. These are often re-ferred to with the unlovely phrase "fractal-like". An example is given in Fig. 7a. There are many examples of this type, e. g. [62]. A related example is the electrodeposition of metal ions from solution. For overpotential situations DLA patterns are often observed [44,63,64,65].

There are also examples of Laplacian growth. A case of this type was discovered by Matsushita and collaborators [66], and exploited in a series of very interesting studies by the group of Ben-Jacob [67] and others. This is the growth of colonies of bacteria on hard agar plates in conditions of low nutrient supply. In this case, bacteria movement is suppressed (by the hard agar) and the limiting step in colony growth is the diffusion of nutrients to the colony. Thus we have an almost literal realization of Laplacian growth, and, indeed, colonies do look like DLA clusters in these conditions; see Fig. 7b. The detailed study of this system has led to very interesting insights into the biophysics of bacteria: these are far from our subject here, and the reader should consult [67].

We remarked above that noisy viscous fingering patterns are similar to DLA clusters, but not the same in detail: in viscous fingering the surface tension boundary condition is different from that of DLA which involves discrete particles. We should note that for viscous fingering patterns in a channel, the asymptotic state is not a disorderly cluster, but rather a single finger that fills half the channel [48] because surface tension smooths the finger. In radial growth this is not true: the Mullins–Sekerka instability gives rise to a disorderly pattern [51] which looks very much like a DLA cluster, see Fig. 8. Even for growth in a channel, if there is sufficient noise, patterns look rather
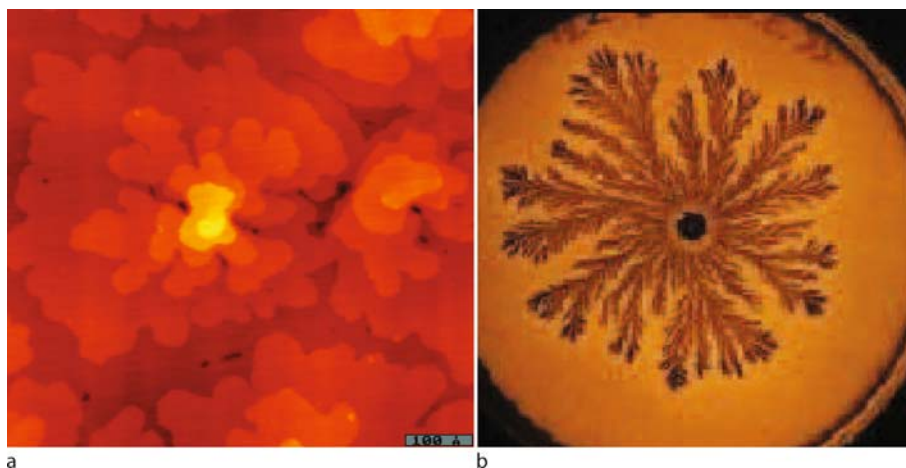
Fractal Growth Processes, Figure 7
a A scanning tunneling microscope picture of Rh metal islands. The *color scale* indicates height, and the figure is about 500 Å across. Figure courtesy of R. Clarke. **b** A bacteria colony on a Petri dish. The figure is a few centimeters across. Figure courtesy of E. Ben-Jacob

Fractal Growth Processes, Figure 8
A radial viscous fingering pattern. Courtesy of M. Moore and E. Sharon

like those in Fig. 4a. In fact, Tang [68] used a version of DLA which allowed him to reduce the noise (see below) and introduce surface tension to do simulations of viscous fingering.

We should ask if this resemblance is more than a mere coincidence. It is the case that the measured fractal dimension of patterns like the one in Fig. 8 is close to 1.71, as for DLA. On this basis there are several claims in the literature that the large-scale structure of the patterns is identical [9,51,69]. Many authors have disagreed and given ingenious arguments about how to verify this. For example, some have claimed that viscous patterns become two-dimensional at large scales [70], or that viscous fingering patterns are most like DBM patterns with $\eta \approx 1.2$ [71]. Most recently a measurement of the growth probability of a large viscous fingering pattern was found to agree with that of DLA [72]. On this basis, these authors claim that DLA and viscous fingering are in the same universality class. In our opinion, this subject is still open.

Conformal Mapping

For pattern formation in two dimensions the use of analytic function theory and conformal mapping methods allows a new look at growth processes. The idea is to think of a pattern in the z plane as the image of a simple reference shape, e. g. the unit circle in the w plane, under a time-dependent analytic function, $z = F_t(w)$. More precisely, we think of the region outside of the pattern as the image of the region outside of the reference shape. By the Riemann mapping theorem the map exists and is unique if we set

a boundary condition such as $F(w) \to r_0 w$ as $w \to \infty$. We will also use the inverse map, $w = G(z) = F^{-1}(z)$.

For Laplacian growth processes this idea is particularly interesting since the growth rules depend on solving the Laplace equation outside the cluster. We recall that the Laplace equation is conformally invariant; that is, it retains its form under a conformal transformation. Thus we can solve in the w plane, and transform the solution: if $\nabla^2 \phi(w) = 0$, and $\phi = 0$ on the unit circle, and we take ϕ to be the real part of a complex potential Φ in the w plane, then Re $\Phi(G(z))$ solves the Laplace equation in the z plane with Re $\Phi = 0$ on the cluster boundary. Thus we can solve for the potential outside the unit circle in w-space (which is easy): $\Phi = \ln(w)$. Then if we map to z space we have the solution outside of the cluster:

$$\Phi(z) = \ln G(z) \,. \tag{11}$$

Note that the constant r_0 has the interpretation of a mean radius since $\Phi \to \ln(z/r_0)$ as $z \to \infty$. In fact, r_0 is the radius of the disk that gives the same potential as the cluster far away.

This has another consequence: the electric field (i. e. the growth probability) is uniform around the unit circle in the w plane. This means that equal intervals on the unit circle in w space map into equal regions of growth probability in the z plane. The map contains not only the shape of the cluster (the image of $|w| = 1$) but also information about the growth probability: using Eq. (11) we have:

$$|\nabla \Phi| = |G'| = \frac{1}{|F'|} \,. \tag{12}$$

The problem remains to construct the maps G or F for a given cluster. Somfai, et al. gave a direct method [38]: for a given cluster release a large number, M, of random walkers and record where they hit, say at points z_j. We know from the previous paragraph that the images of these points are spaced roughly equally on the unit circle in the w plane. That is, if we start somewhere on the cluster and number the landing positions sequentially around the cluster surface, the images in the w-plane, $w_j = r_j e^{i\theta_j}$, are given by $\theta_j = 2\pi j/M$, $r_j = 1$. Thus we have the boundary values of the map and by analytic continuation we can construct the entire map. In fact, if we represent F by a Laurent series:

$$F(w) = r_0 w + A_0 + \sum_{j=1}^{\infty} \frac{A_j}{w^j} \,, \tag{13}$$

it is easy to see that the Fourier coefficients of the function $F(\theta_j)$ are the A_j.

Unfortunately, for DLA this method only gives a map of the highly probable points on the surface of the cluster. The inner, frozen, regions are very badly sampled by random walkers, so the generated map will not represent these regions.

Loewner Evolution and the Hastings–Levitov Scheme

A more useful approach to finding F is to impose a dynamics on the map which gives rise to the growth of the cluster. This is very closely related to Loewner evolution, where a curve in the z plane is generated by a map that obeys an equation of motion:

$$\frac{\mathrm{d}G_t(z)}{\mathrm{d}t} = \frac{2}{G_t(z) - \xi(t)} \,. \tag{14}$$

The map is to the upper half plane from the upper half plane minus the set of singularities, $G = \xi$. (For an elementary discussion and references see [73].) If $\xi(t)$ is a stochastic process then many interesting statistical objects such as percolation clusters can be generated.

For DLA a similar approach was presented by Hastings and Levitov [21,30]. In this case the evolution is discrete, and is designed to represent the addition of particles to a cluster. The process is iterative: suppose we know the map for N particles, F_N. Then we want to add a "bump" corresponding to a new particle. This is accomplished by adding the bump of area λ in the w-plane on the surface of the unit circle at angle θ. There are various explicit functions that generate bumps; for the most popular example see [21]. This is a function that depends on a parameter which gives the aspect ratio of the bump.

Let us call the resulting transformation $f_{\lambda,\theta}$. If we use F_N to transform the unit circle *with a bump* we get a cluster with an extra bump in the z-plane: that is, we have added a particle to the cluster and $F_{N+1} = F_N \circ f$. The scheme is represented in Fig. 9.

There are two matters that need to be dealt with. First, we need to pick θ. Since the probability to grow is uniform on the circle in w-space, we take θ to be a random variable uniformly distributed between 0 and 2π. Also, we need the bump in z-space to have a fixed area, λ_0. That means that the bump in w-space needs to be adjusted because conformal transformations stretch lengths by $|F'|$. A first guess for the area λ is:

$$\lambda_{N+1} = \frac{\lambda_0}{|F'_N(e^{i\theta_{N+1}})|^2} \,. \tag{15}$$

However, this is just a first guess. The stretching of lengths varies over the cluster, and there can be some regions where the approximation of Eq. (15) is not adequate. In
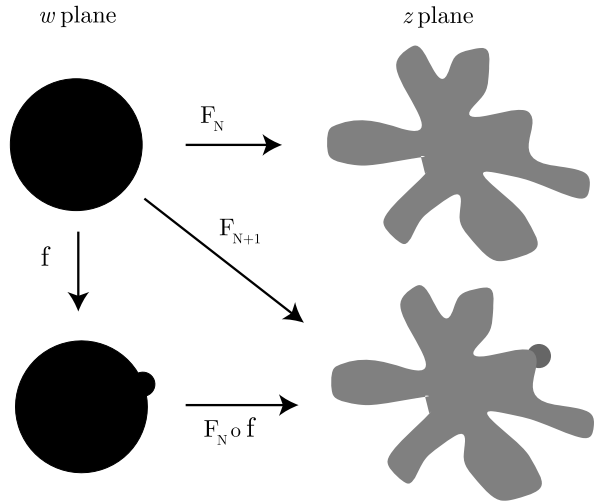
Fractal Growth Processes, Figure 9
The Hastings–Levitov scheme for fractal growth. At each stage a "bump" of the proper size is added to the unit circle. The composition of the bump map, ϕ, with the map at stage N gives the map at stage $N + 1$

this case an iterative procedure is necessary to get the area right [30,34]. The transformation itself is given by:

$$F_N = f_{\lambda_1,\theta_1} \circ f_{\lambda_2,\theta_2} \circ \cdots \circ f_{\lambda_N,\theta_N} \,. \tag{16}$$

All of the information needed to specify the mapping is contained in the list λ_j, θ_j, $1 \le j \le N$. An example of a cluster made this way is shown in Fig. 10.

If we choose θ uniformly in w-space, we have chosen points with the harmonic measure, $|F'|^{-1}$ and we make DLA clusters. To simulate DBM clusters with $\eta \ne 1$ we must choose the angles non-uniformly. Hastings [54] has shown how to do this: since a uniform distribution gives a distribution according to $|F'|^{-1}$ (see Eq. (12)) we have to pick angles with probability $|F'|^{1-\eta}$ in order to grow with probability $|\nabla\phi|^\eta$. This can be done with a Metropolis algorithm with $p(\theta_{N+1}) = |F'_N(e^{i\theta})|^{1-\eta}$ playing the role of a Boltzmann factor.

Applications of the Hastings–Levitov Method: Scaling and Crossovers

The Hastings–Levitov method is a new numerical algorithm for DLA, but not a very efficient one. Constructing the composed map takes of order N steps so that the algorithm is of order N^2. One may wonder what has been gained. The essential point is that new understanding arises from considering the objects that are produced in the course of the computation. For example, Davidovitch and collaborators [74] showed that averages of λ over the
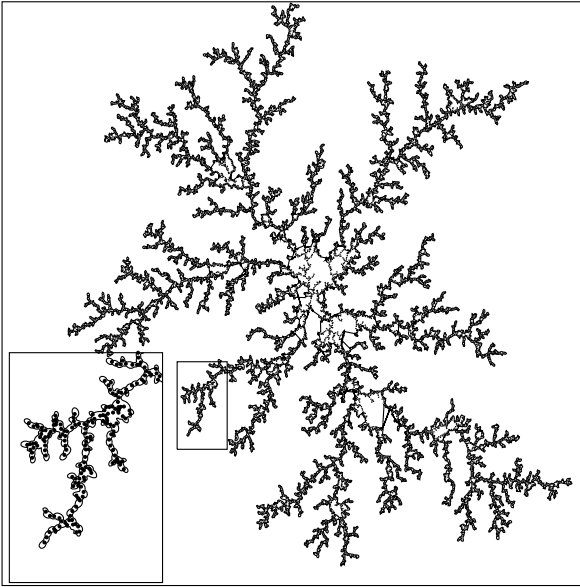
Fractal Growth Processes, Figure 10
A DLA cluster made by iterated conformal maps. Courtesy of E. Somfai

cluster boundary are related to the generalized dimensions of the growth probability measure, cf. Eq. (4). Further, the Laurent coefficients of the map, Eq. (13), are meaningful in themselves and have interesting scaling properties. We have seen above that r_0 is the radius of a disk with the same capacitance as the cluster. It scales as $N^{1/D}$, as we expect. The constant, A_0 gives the wandering of the center of the cluster from its original position. Its scaling can be worked out in terms of D and the generalized dimension, D_2.

The other coefficients, A_j, are related to the moments of the charge density on the cluster surface. We expect them to scale in the same way as r_0 for the following reason: there is an elementary theorem in complex analysis [75] for any univalent map that says, in our notation:

$$\pi r_0^2 = S_N + \pi \sum_{k=1} k |A_k|^2 . \tag{17}$$

Here S_N is the area of the cluster. However, this is just the area of N particles, and is linear in N. Therefore, in leading order, the sum must cancel the $N^{2/D}$ dependence of r_0^2. The simplest way this can occur is if every term in the sum goes as $N^{2/D}$. This seemed to be incorrect according to the data in [74]: they found that for for the first few k the scaling exponents of the $\langle |A_k|^2 \rangle$ were smaller than $2/D$.

However, later work [38] showed that the asymptotic scaling of *all* of the coefficients is the same. The apparent difference in the exponents is due to a slow *crossover*. This

effect also seems to resolve a long-standing controversy about the asymptotic scaling behavior of DLA clusters, namely the anomalous behavior of the penetration depth of random walkers as the clusters grow.

The anomaly was pointed out numerically by Plischke and Racz [76] soon after the DLA algorithm was introduced. These authors showed numerically that the width of the region where deposition occurred, ξ, (a measure of penetration of random walkers into the cluster) seemed to grow more slowly with N than the mean radius of deposition, R_{dep}. However for a simple fractal all of the characteristic lengths must scale in the same way, $R \propto N^{1/D}$. Mandelbrot and collaborators [36,37,77] used this and other numerical evidence to suggest that DLA would not be a simple fractal for large N. However, Meakin and Sander [78] gave numerical evidence that the anomaly in the scaling of ξ is not due to a different exponent, but is a crossover.

The controversy is resolved [38] by noting that the penetration depth can be estimated from the Laurent coefficients of f:

$$\xi^2 = \frac{1}{2} \sum_{k=1} |A_k|^2 . \tag{18}$$

It is easy to see [38] that this version of the penetration depth can be interpreted as the rms deviation of the position of the end of a field line from those for a disk of radius r_0. Thus an anomaly in the scaling of the A_k is related to the effect discovered in [76]. Further, a slow crossover in A_k for small k is reasonable geometrically since this is a slow crossover in low moments of the charge. These low moments might be expected to have intrinsically slow dynamics.

In [38,79,80] strong numerical evidence was given for the crossover and for universal asymptotic scaling of the A_k. Indeed, many quantities with the interpretation of a length fit an expression of the form:

$$N^{1/D} \left(1 + \frac{C}{N^{\nu}} \right) , \tag{19}$$

where the subdominant correction to scaling is characterized by an exponent $\nu = 0.33 \pm 0.06$. This observation verifies the crossover of ξ. The asymptotic value of the penetration depth is measured to be $\xi/R_{\text{dep}} \to 0.12$. The same value is found in three dimensions [81].

In fact, the crossover probably accounts for the anomalies reported in [36,37,77]. Further, a careful analysis of numerical data shows that the reported *multiscaling* [82] (an apparent dependence of D on the distance from the center of the cluster) is also a crossover effect. These in-
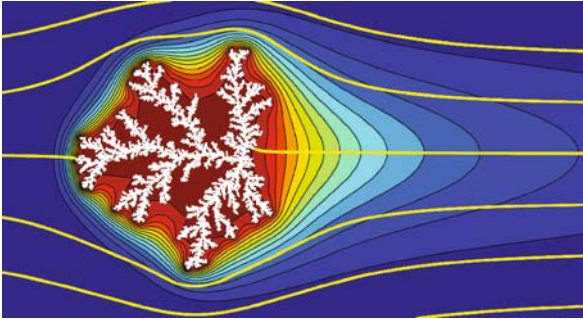
Fractal Growth Processes, Figure 11
An aggregate grown by the advection-diffusion mechanism, Eq. (20) for Pe=10. Courtesy of M. Bazant

sights lead to the view that DLA clusters are simple fractals up to a slow crossover. The origin of the crossover exponent, $\nu \approx 1/3$, is not understood. In three dimensions a similar crossover exponent was found by different techniques [81] with a value of 0.22 ± 0.03.

Conformal Mapping for Other Discrete Growth Processes

The Hastings–Levitov scheme depends on the conformal invariance of the Laplace equation. Bazant and collaborators [83,84] pointed out that the same techniques can be used for more general growth processes. For example, consider growth of an aggregate in a flowing fluid. Now there are two relevant fields, the particle concentration, c, and the velocity of the fluid, $\mathbf{v} = \nabla \psi$ (for potential flow). The current associated with c can be written $\mathbf{j} = \mu c \mathbf{v} - \nu \nabla c$, where μ is a mobility and ν a diffusion coefficient. For steady incompressible flow we have, after rescaling:

$$\mathrm{Pe} \nabla \psi \cdot \nabla c = \nabla^2 c \, ; \quad \nabla^2 \psi = 0 \, . \qquad (20)$$

Here Pe is the Péclet number, UL/ν, where U is the value of the velocity far from the aggregate and L its initial size. These equations are conformally invariant, and the problem of flow past a disk in the w plane is easily solved. In [83] growth was accomplished by choosing bumps with probability distribution $\partial c / \partial n$ using the method described above. This scheme solves the flow equation past the complex growing aggregate and adds particles according to their flux at the surface. An example of a pattern of this type is given in Fig. 11.

It has long been known that it is possible to treat quasi-static fracture as a growth process similar to DLA [85,86,87]. This is due to the fact that the Lamé equation of elasticity is similar in form to the Laplace equation and the condition for breaking a region of the or-

der of a process zone is related to the stress at the current crack, i. e., boundary values of derivatives of the displacement field. Recent work has exploited this similarity to give yet another application of conformal mapping. For example, for Mode III fracture the quasi-static elastic equation reduces to the Laplace equation, and it is only necessary to replace the growth probability and the boundary conditions in order to use the method of iterated conformal maps [88]. For Mode I and Mode II fracture it is necessary to solve for two analytic functions, but this can also be done [89].

Harmonic Measure

The distribution of the boundary values of the normal derivatives of a potential on a electrode of complex shape is called the problem of the harmonic measure. For DLA it is equivalent to the distribution of growth probabilities on the surface of the cluster, or, in other terms, the penetration of random walkers into the cluster. For other complex shapes the problem is still interesting. Its practical significance is that of the study of electrodes [90] or catalytic surfaces. In some cases the harmonic measure has deep relationships to conformal field theory.

For the case of the harmonic measure we can interpret the variable α of Eq. (6) as the singularity strength of the electric field near a sharp tip. This is seen as follows: it is well known [91] that near the apex of a wedge-shaped conductor the electric field diverges as $r^{\pi/\beta-1}$ where r is the distance from the tip and the exterior angle of the wedge is β. For example, near a square corner with $\beta = 3\pi/2$ there is a $r^{-1/3}$ divergence. Now the quantity p_i is the integral over a box of size l. Thus a sequence of boxes centered on the tip will give a power-law $l^{\pi/\beta} = l^\alpha$. Smaller α means stronger divergence.

For fractals that can be treated with the methods of conformal field theory, a good deal is known about the harmonic measure. For example, for a percolation cluster at p_c the harmonic measure is completely understood [92]. The D_q are given by a formula:

$$D_q = \frac{1}{2} + \frac{5}{\sqrt{24q + 1} + 5} \, , \quad q \geq -\frac{1}{24} \, . \qquad (21)$$

The $f(\alpha)$ spectrum is easy to compute from this. This formula in good accord with the numerical results of Meakin and collaborators [93] who sampled the measure by firing many random walkers at a cluster. There is an interesting feature of this formula: $D_0 = 4/3$ is the dimension of the support of the measure. This is *less* than the dimension of a percolation hull, 7/4. There is a large part of the surface of a percolation cluster which is inaccessible to random walk-

ers – the interior surfaces are cut off from the exterior by narrow necks whose widths vanish in the scaling regime.

For DLA the harmonic measure is much less well understood. Some results are known. For example, Makarov [94] proved that $D_1 = 1$ under very general circumstances for a two-dimensional harmonic measure. Halsey [95] used Green's functions for electrostatics to prove the following for DBM models with parameter η:

$$\tau(\eta + 2) - \tau(\eta) = D, \tag{22}$$

Here D is the fractal dimension of the cluster. For DLA $\eta = 1$, and $\tau(1) = 0$. Thus $\tau(3) = D = 1.71$.

We introduced the function $f(\alpha)$ as the Legendre transform of $D(q)$. There is an argument due to Turke-vich and Sher [96] which allows us to see a feature of $f(\alpha)$ directly by giving an estimate of the singularity associated with the most active tip of the growing cluster. Note that the growth rate of the extremal radius of the cluster is related to the fractal dimension because $R_{ext} \propto N^{1/D}$. Suppose we imagine adding one particle per unit time and recall that $p_{tip} \propto (l/R_{ext})^{\alpha_{tip}}$. Then:

$$\frac{dR_{ext}}{dt} = \frac{dR_{ext}}{dN} \frac{dN}{dt} \propto R_{ext}^{-\alpha_{tip}}$$

$$D = 1 + \alpha_{tip} \approx 1 + \alpha_{min}. \tag{23}$$

Since the singularity at the tip is close to being the most active one, we have an estimate of the minimum value of α.

There have been a very large number of numerical investigations of $D(q)$, $f(\alpha)$ for two dimensional DLA; see [7] for a comprehensive list. They proceed either by launching many random walkers, e. g. [93] or by solving the Laplace equation [97], or, most recently, by using the Hastings–Levitov method [98]. The general features of the results are shown in Fig. 12. There is fairly good agreement about the left hand side of the curve which corresponds to large probabilities. The intercept for small α is close to the Turkevich–Sher relation above. The maximum of the curve corresponds to $df/d\alpha = q = 0$; thus it is the dimension of the support of the measure. This seems to be quite close to D so that the whole surface is accessible to random walkers.

There is very little agreement about the right hand side of the curve. It arises from regions with small probabilities which are very hard to estimate. The most reliable current method is that of [98] where conformal maps are manipulated to get at probabilities as small as 10^{-70}, quite beyond the reach of other techniques. Unfortunately, these computations are for rather small clusters ($N \approx 50,000$) which are the largest ones that can be made by the Hast-
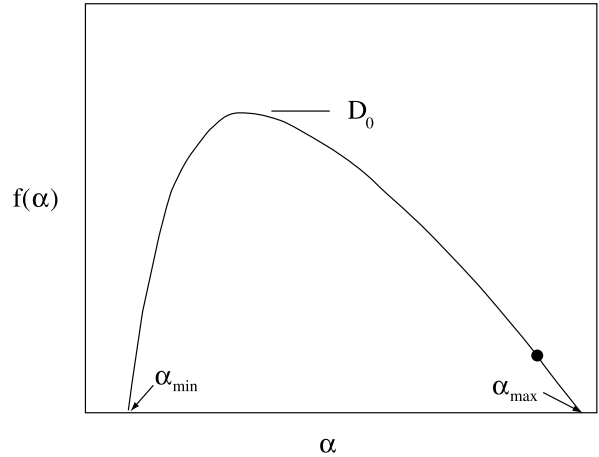
Fractal Growth Processes, Figure 12
A sketch of the *f(α)* curve. Some authors find a maximum slope at a position like that marked by *the dot* so that the curve ends there. The curve is extended to the real axis with a *straight line*. This is referred to as a phase transition

ings–Levitov method. A particularly active controversy relates to the value of α_{max}, if it exists at all; that is, the question is whether there is a maximum value of the slope $d\tau/dq$. The authors of [98] find $\alpha_{max} \approx 20$.

Scaling Theories

Our understanding of non-equilibrium fractal growth processes is not very satisfactory compared to that of equilibrium processes. A long-term goal of many groups has been to find a "theory of DLA" which has the nice features of the renormalization theory of critical phenomena. The result of a good deal of work is that we have theories that give the general features of DLA, but they do not explain things in satisfactory detail.

We should note that a mere estimate of the fractal dimension, D, is not what we have in mind. There are several ad hoc estimates in the literature that give reasonable values of D [41,99]. We seek, rather, a theory that allows a good understanding of the fixed point of fractal growth with a description of relevant and irrelevant operators, crossovers, etc. There have been an number of such attempts. In the first section below we describe a semi-numerical scheme which sheds light on the fixed point. Then we look at several attempts at ab initio theory.

Scaling of Noise

The first question one might ask about DLA growth is the role of noise. It might seem that for a very large cluster the noise of individual particle arrivals should average

out. However, this is not the case. Clusters fluctuate on all scales for large N.

Further, the noise-free case seems to be unstable. We can see this by asking about the related question of the growth of viscous fingers in two dimensions. As was remarked long ago [51] this sort of growth is always unstable. Numerical simulations of viscous fingering [69] show that any asymmetric initial condition develops into a pattern with a fractal dimension close to that of DLA. DLA organizes its noise to a fixed point exactly as turbulence does.

Several authors [79,81,100] have looked at the idea that the noise (measured in some way) flows to a fixed value as clusters grow large. For example, we could characterize the shape fluctuations by measuring the scaled variance of the number of particles necessary to grow to a fixed extremal radius:

$$\frac{\delta N}{N}\bigg|_{R_{\text{ext}}} = \sqrt{A^*}\,. \tag{24}$$

This is easy to measure for large clusters. In two dimensions $A^* = .0036$ [79]. In [100] it was argued that DLA will be at its fixed point if if one unit of growth acts as a coarse graining of DLA on finer length scales. This amounts to a kind of real-space renormalization.

For the original DLA model the scaled noise to grow one unit of length is of order unity. We can reduce it to the fixed point value by *noise reduction*. The slow approach to the fixed point governed by the exponent ν can be interpreted as the drift of the noise to its proper value as N grows.

Noise reduction was introduced into lattice models by Tang [68]. He kept a counter on each site and only grew there if a certain number of random walkers, m, hit that point. For off-lattice clusters a similar reduction is obtained if shallow bumps of height A are added to the cluster by letting the particles overlap as they stick. We must add $m = 1/A$ particles to advance the growth by one particle diameter [79]. For the Hastings–Levitov method it is equivalent to use a bump map with a small aspect ratio [30].

In either case if we examine a number of sites whose mean advance is one unit, we will find $\delta N/N = \sqrt{A}$. We should expect that if we tune the input value of A to A^* we will get to asymptotic behavior quickly. This is what is observed in two dimensions [79] and in three dimensions [81]. The amplitude of the crossover, the parameter C in Eq. (19), is smallest for A near the fixed point. In Fig. 13 we show two clusters, a small one grown with noise reduction, and a much larger one with $A = 1$. They both
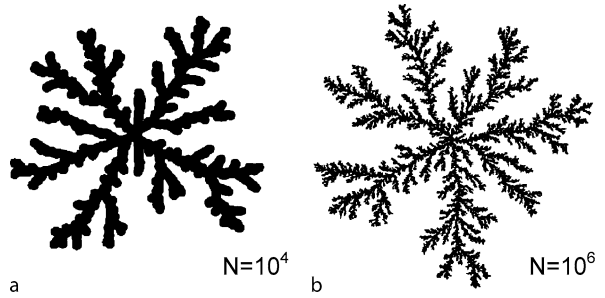
Fractal Growth Processes, Figure 13
A small noise-reduced cluster and a larger one with no noise reduction. Noise reduction of the proper size accelerates the approach to the fixed point. From [79]

have the same value of ξ/R_{dep}, near the asymptotic value of 0.12.

Attempts at Theoretical Description

The last section described a semi-empirical approach to DLA. The reader may wonder why the techniques of phase transition theory are not simply applicable here. One problem is that one of the favorite methods in equilibrium theory, the ε-expansion, cannot be used because DLA has no upper critical dimension [5,101].

To be more precise, in other cluster theories such as percolation there is a dimension, d_c such that if $d > d_c$ the fractal dimension of the cluster does not change. For example, percolation clusters are 4 dimensional for all dimensions above 6. For DLA this is not true because if d is much bigger than D then random walkers will *penetrate and fill up the cluster* so that its dimension would increase. This results [5] from the fact that for a random walker the number of sites visited in radius R is $N_w \propto R^2$. In the same region there are R^D sites of the cluster, so the density of cluster sites goes as R^{D-d}. The mean number of intersections is $R^{D-d}R^2$. Therefore if $D + 2 < d$ there are a vanishing number of intersections, and the cluster will fill up. Thus we must have $D \geq d - 2$. A related argument [101] sharpens the bound to $D \geq d - 1$: the fractal dimension increases without limit as d increases.

Halsey and collaborators have given a theoretical description based on branch competition [102,103]. This method has been used to give an estimate of D_q for positive q [104]. The idea is to think of two branches that are born from a single site. Then the probability to stick to the first or second is called p_1, p_2 where $p_1 + p_2 = p_b$ is the total probability to stick to that branch. Similarly there are numbers of particles, $n_1 + n_2 = n_b$. Define $x = p_1/p_b, y = n_1/n_b$. The two variables x, y regulate the

branch competition. On the average $p_1 = dn_1/dn_b$ so we have:

$$n_b \frac{dy}{dn_b} = x - y. \qquad (25)$$

The equation of motion for x is assumed to be of similar form:

$$n_b \frac{dx}{dn_b} = g(x, y). \qquad (26)$$

Thus the competition is reduced to a two-dimensional dynamical system with an unstable fixed point at $x = y = 1/2$ corresponding to the point when the two branches start with equal numbers. The fixed point is unstable because one branch will eventually screen the other. If g is known, then the unstable manifold of this fixed point describes the growth of the dominant branch, and it turns out, by counting the number of particles that attach to the dominant branch, that the eigenvalue of the unstable manifold is $1/D$. The starting conditions for the growth are taken to result from microscopic processes that distribute points randomly near the fixed point.

The problem remains to find g. This was done several ways: numerically, by doing a large number of simulations of branches that start equally, or in terms of a complicated self-consistent equation [103]. The result is a fractal dimension of 1.66, and a multifractal spectrum that agrees pretty well with direct simulations [104].

Another approach is due to Pietronero and collaborators [105]. It is called the method of fixed scale transformations. It is a real-space method where a small system at one scale is solved essentially exactly, and the behavior at the next coarse-grained scale estimated by assuming that there is a scale-invariant dynamics and estimating the parameters from the fixed-scale solution. The method is much more general than a theory of DLA: in [105] it is applied to directed percolation, the Eden model, sandpile models, and DBM. For DLA the fractal dimension calculated is about 1.6. The rescaled noise (cf. the previous section) comes out to be of order unity rather than the small value, 0.0036, quoted above [106].

The most recent attempt at a fundamental theory is due to Ball and Somfai [71,107]. The idea depends on a mapping from DLA to an instance of the DBM which has different boundary conditions on the growing tip. The scaling of the noise and the multifractal spectrum (for small α) are successfully predicted.

Future Directions

The DLA model is 27 years old as of this writing. Every year (including last year) there have been about 100 ref-

erences to the paper. Needless to say, this author has only read a small fraction of them. Space and time prevented presenting here even the interesting ones that I am familiar with.

For example, there is a remarkable literature associated with the viscous-fingering problem without surface tension which seems, on the one hand, to describe some facets of experiments [108] and on the other to have deep relationships with the theory of 2d quantum gravity [109,110]. Where this line of work will lead is a fascinating question. There are other examples: I hope that those whose work I have not covered will not feel slighted. There is simply too much going on.

A direction that should be pursued is to use the ingenious techniques that have been developed for the DLA problem for problems in different areas; [59,83] are examples of this.

It is clear that this field is as lively as ever after 27 years, and will certainly hold more surprises.

Bibliography

Primary Literature

1. Witten TA, Sander LM (1981) Diffusion-limited aggregation, a kinetic critical phenomenon. Phys Rev Lett 47:1400
2. Mandelbrot BB (1982) The Fractal Geometry of Nature. Freeman, San Francisco
3. Stauffer D, Aharony A (1994) Introduction to percolation theory. Taylor & Francis, London
4. Fortuin CM, Kasteleyn PW (1972) On the random-cluster model: I introduction and relation to other models. Physica 57(4):536–564
5. Witten TA, Sander LM (1983) Diffusion-limited aggregation. Phys Rev B 27:5686
6. Vicsek T (1992) Fractal Growth Phenomena, 2nd edn. World Scientific, Singapore
7. Meakin P (1998) Fractals, scaling, and growth far from equilibrium. Cambridge University Press, Cambridge
8. Godreche G (1991) Solids far from equilibrium. Cambridge, Cambridge, New York
9. Sander LM (1986) Fractal growth-processes. Nature 322(6082):789–793
10. Sander LM (2000) Diffusion limited aggregation, a kinetic critical phenomenon? Contemporary Physics 41:203–218
11. Halsey TC (2000) Diffusion-limited aggregation: A model for pattern formation. Physics Today 53(4):36–41
12. Meakin P (1983) Formation of fractal clusters and networks by irreversible diffusion-limited aggregation. Phys Rev Lett 51(13):1119–1122
13. Kolb M, Botet R, Jullien R (1983) Scaling of kinetically growing clusters. Phys Rev Lett 51(13):1123–1126
14. Meakin P (1988) Fractal aggregates. Adv Colloid Interface Sci 28(4):249–331
15. Meakin P, Ramanlal P, Sander LM, Ball RC (1986) Ballistic deposition on surfaces. Phys Rev A 34(6):5091–5103

16. Eden M (1961) A two-dimensional growth model. In: Neyman J (ed) Proceedings of the 4th Berkeley symposium on mathematics, statistics, and probability. University of California Press, Berkeley

17. Kardar M, Parisi G, Zhang Y (1986) Dynamic scaling of growing interfaces. Phys Rev Lett 56:889

18. Barabasi A, Stanley HE (1995) Fractal Concepts in Surface Growth. Cambridge, Cambridge, New York

19. Family F, Vicsek T (1992) The Dynamics of Growing Interfaces. World Scientific, Singapore

20. Halpin-Healy T, Zhang Y-C (1995) Kinetic roughening phenomena, stochastic growth, directed polymers and all that. Aspects of multidisciplinary statistical mechanics. Phys Rep 254(4–6):215–414

21. Hastings MB, Levitov LS (1998) Laplacian growth as one-dimensional turbulence. Physica D 116:244–252

22. Hentschel HGE, Procaccia I (1983) The infinite number of generalized dimensions of fractals and strange attractors. Physica D 8(3):435–444

23. Halsey TC, Jensen MH, Kadanoff LP, Procaccia I, Shraiman BI (1986) Fractal measures and their singularities – the characterization of strange sets. Phys Rev A 33(2):1141–1151

24. Grassberger P, Procaccia I (1983) Measuring the strangeness of strange attractors. Physica D 9:189–208

25. Mandelbrot BB (1974) Intermittant turbulence in self-similar cascades; divergence of high moments and dimension of the carrier. Fluid J Mech 62:331–358

26. Ball RC, Brady RM (1985) Large-scale lattice effect in diffusion-limited aggregation. Phys J A-Math Gen 18(13):L809

27. Ball RC, Brady RM, Rossi G, Thompson BR (1985) Anisotropy and cluster growth by diffusion-limited aggregation. Phys Rev Lett 55(13):1406–1409

28. Meakin P, Ball RC, Ramanlal P, Sander LM (1987) Structure of large two-dimensional square-lattice diffusion-limited aggregates – approach to asymptotic-behavior. Phys Rev A 35(12):5233–5239

29. Eckmann JP, Meakin P, Procaccia I, Zeitak R (1990) Asymptotic shape of diffusion-limited aggregates with anisotropy. Phys Rev Lett 65(1):52–55

30. Stepanov MG, Levitov LS (2001) Laplacian growth with separately controlled noise and anisotropy. Phys Rev E 63:061102

31. Goold NR, Somfai E, Ball RC (2005) Anisotropic diffusion limited aggregation in three dimensions: Universality and nonuniversality. Phys Rev E 72(3):031403

32. Tolman S, Meakin P (1989) Off-lattice and hypercubic-lattice models for diffusion-limited aggregation in dimensionalities 2–8. Phys Rev A 40:428–437

33. Ossadnik P (1991) Multiscaling analysis of large-scale off-lattice DLA. Physica A 176:454–462

34. Somfai E, Ball RC, DeVita JP, Sander LM (2003) Diffusion-limited aggregation in channel geometry. Phys Rev E 68: 020401(R)

35. Vicsek T, Family F, Meakin P (1990) Multifractal geometry of diffusion-limited aggregates. Europhys Lett 12(3): 217–222

36. Mandelbrot BB, Kaufman H, Vespignani A, Yekutieli I, Lam CH (1995) Deviations from self-similarity in plane DLA and the infinite drift scenario. Europhys Lett 29(8):599–604

37. Mandelbrot BB, Vespignani A, Kaufman H (1995) Crosscut analysis of large radial DLA – departures from self-similarity and lacunarity effects. Europhys Lett 32(3):199–204

38. Somfai E, Sander LM, Ball RC (1999) Scaling and crossovers in diffusion limited aggregation. Phys Rev Lett 83:5523–5526

39. Arneodo A, Elezgaray J, Tabard M, Tallet F (1996) Statistical analysis of off-lattice diffusion-limited aggregates in channel and sector geometries. Phys Rev E 53(6):6200–6223(B)

40. Tu YH, Levine H (1995) Mean-field theory of the morphology transition in stochastic diffusion-limited growth. Phys Rev E 52(5):5134–5141

41. Kessler DA, Olami Z, Oz J, Procaccia I, Somfai E, Sander LM (1998) Diffusion-limited aggregation and viscous fingering in a wedge: Evidence for a critical angle. Phys Rev E 57(6): 6913–6916

42. Sander LM, Somfai E (2005) Random walks, diffusion limited aggregation in a wedge, and average conformal maps. Chaos 15:026109

43. Meakin P, Family F (1986) Diverging length scales in diffusion-limited aggregation. Phys Rev A 34(3):2558–2560

44. Argoul F, Arneodo A, Grasseau G, Swinney HL (1988) Self-similarity of diffusion-limited aggregates and electrodeposition clusters. Phys Rev Lett 61(22):2558–2561

45. Kol B, Aharony A (2001) Diffusion-limited aggregation as markovian process: Site-sticking conditions. Phys Rev E 63(4): 046117

46. Mullins WW, Sekerka RF (1963) Morphological stability of a particle growing by diffusion or heat flow. J Appl Phys 34:323

47. Langer JS (1980) Instabilities and pattern formation in crystal growth. Rev Mod Phys 52:1

48. Pelcé P (2004) New visions on form and growth: fingered growth, dendrites, and flames. In: Thèorie des formes de croissance. Oxford University Press, Oxford

49. Mineev-Weinstein MB, Dawson SP (1994) Class of nonsingular exact-solutions for laplacian pattern-formation. Phys Rev E 50(1):R24–R27

50. Shraiman B, Bensimon D (1984) Singularities in nonlocal interface dynamics. Phys Rev A 30:2840–2842

51. Paterson L (1984) Diffusion-limited aggregation and 2-fluid displacements in porous-media. Phys Rev Lett 52(18): 1621–1624

52. Niemeyer L, Pietronero L, Wiesmann HJ (1984) Fractal dimension of dielectric breakdown. Phys Rev Lett 52:1033–1036

53. Sanchez A, Guinea F, Sander LM, Hakim V, Louis E (1993) Growth and forms of Laplacian aggregates. Phys Rev E 48:1296–1304

54. Hastings MB (2001) Fractal to nonfractal phase transition in the dielectric breakdown model. Phys Rev Lett 87:175502

55. Hastings MB (2001) Growth exponents with 3.99 walkers. Phys Rev E 64:046104

56. Meakin P (1983) Effects of particle drift on diffusion-limited aggregation. Phys Rev B 28(9):5221–5224

57. Nauenberg M, Richter R, Sander LM (1983) Crossover in diffusion-limited aggregation. Phys Rev B 28(3):1649–1651

58. Sander E, Sander LM, Ziff R (1994) Fractals and fractal correlations. Comput Phys 8:420

59. DeVita JP, Sander LM, Smereka P (2005) Multiscale kinetic monte carlo algorithm for simulating epitaxial growth. Phys Rev B 72(20):205421

60. Hou TY, Lowengrub JS, Shelley MJ (1994) Removing the stiffness from interfacial flow with surface-tension. J Comput Phys 114(2):312–338

61. Somfai E, Goold NR, Ball RC, DeVita JP, Sander LM (2004)

Growth by random walker sampling and scaling of the dielectric breakdown model. Phys Rev E 70:051403

62. Radnoczi G, Vicsek T, Sander LM, Grier D (1987) Growth of fractal crystals in amorphous GeSe$_2$ films. Phys Rev A 35(9):4012–4015

63. Brady RM, Ball RC (1984) Fractal growth of copper electrodeposits. Nature 309(5965):225–229

64. Grier D, Ben-Jacob E, Clarke R, Sander LM (1986) Morphology and microstructure in electrochemical deposition of zinc. Phys Rev Lett 56(12):1264–1267

65. Sawada Y, Dougherty A, Gollub JP (1986) Dendritic and fractal patterns in electrolytic metal deposits. Phys Rev Lett 56(12):1260–1263

66. Fujikawa H, Matsushita M (1989) Fractal growth of bacillussubtilis on agar plates. Phys J Soc Jpn 58(11):3875–3878

67. Ben-Jacob E, Cohen I, Levine H (2000) Cooperative self-organization of microorganisms. Adv Phys 49(4):395–554

68. Tang C (1985) Diffusion-limited aggregation and the Saffman–Taylor problem. Phys Rev A 31(3):1977–1979

69. Sander LM, Ramanlal P, Ben-Jacob E (1985) Diffusion-limited aggregation as a deterministic growth process. Phys Rev A 32:3160–3163

70. Barra F, Davidovitch B, Levermann A, Procaccia I (2001) Laplacian growth and diffusion limited aggregation: Different universality classes. Phys Rev Lett 87:134501

71. Ball RC, Somfai E (2002) Theory of diffusion controlled growth. Phys Rev Lett 89:135503

72. Mathiesen J, Procaccia I, Swinney HL, Thrasher M (2006) The universality class of diffusion-limited aggregation and viscous-limited aggregation. Europhys Lett 76(2):257–263

73. Gruzberg IA, Kadanoff LP (2004) The Loewner equation: Maps and shapes. J Stat Phys 114(5–6):1183–1198

74. Davidovitch B, Hentschel HGE, Olami Z, Procaccia I, Sander LM, Somfai E (1999) Diffusion limited aggregation and iterated conformal maps. Phys Rev E 59:1368–1378

75. Duren PL (1983) Univalent Functions. Springer, New York

76. Plischke M, Rácz Z (1984) Active zone of growing clusters: Diffusion-limited aggregation and the Eden model. Phys Rev Lett 53:415–418

77. Mandelbrot BB, Kol B, Aharony A (2002) Angular gaps in radial diffusion-limited aggregation: Two fractal dimensions and nontransient deviations from linear self-similarity. Phys Rev Lett 88:055501

78. Meakin P, Sander LM (1985) Comment on "Active zone of growing clusters: Diffusion-limited aggregation and the Eden model". Phys Rev Lett 54:2053–2053

79. Ball RC, Bowler NE, Sander LM, Somfai E (2002) Off-lattice noise reduction and the ultimate scaling of diffusion-limited aggregation in two dimensions. Phys Rev E 66:026109

80. Somfai E, Ball RC, Bowler NE, Sander LM (2003) Correction to scaling analysis of diffusion-limited aggregation. Physica A 325(1–2):19–25

81. Bowler NE, Ball RC (2005) Off-lattice noise reduced diffusion-limited aggregation in three dimensions. Phys Rev E 71(1):011403

82. Amitrano C, Coniglio A, Meakin P, Zannetti A (1991) Multiscaling in diffusion-limited aggregation. Phys Rev B 44:4974–4977

83. Bazant MZ, Choi J, Davidovitch B (2003) Dynamics of conformal maps for a class of non-laplacian growth phenomena. Phys Rev Lett 91(4):045503

84. Bazant MZ (2004) Conformal mapping of some non-harmonic functions in transport theory. Proceedings of the Royal Society of London Series A – Mathematical Physical and Engineering Sciences 460(2045):1433–1452

85. Louis E, Guinea F (1987) The fractal nature of fracture. Europhys Lett 3(8):871–877

86. Pla O, Guinea F, Louis E, Li G, Sander LM, Yan H, Meakin P (1990) Crossover between different growth regimes in crack formation. Phys Rev A 42(6):3670–3673

87. Yan H, Li G, Sander LM (1989) Fracture growth in 2d elastic networks with Born model. Europhys Lett 10(1):7–13

88. Barra F, Hentschel HGE, Levermann A, Procaccia I (2002) Quasistatic fractures in brittle media and iterated conformal maps. Phys Rev E 65(4)

89. Barra F, Levermann A, Procaccia I (2002) Quasistatic brittle fracture in inhomogeneous media and iterated conformal maps: Modes I, II, and III. Phys Rev E 66(6):066122

90. Halsey TC, Leibig M (1992) The double-layer impedance at a rough-surface – theoretical results. Ann Phys 219(1):109–147

91. Jackson JD (1999) Classical electrodynamics, 3rd edn. Wiley, New York

92. Duplantier B (1999) Harmonic measure exponents for two-dimensional percolation. Phys Rev Lett 82(20):3940–3943

93. Meakin P, Coniglio A, Stanley HE, Witten TA (1986) Scaling properties for the surfaces of fractal and nonfractal objects – an infinite hierarchy of critical exponents. Phys Rev A 34(4):3325–3340

94. Makarov NG (1985) On the distortion of boundary sets under conformal-mappings. P Lond Math Soc 51:369–384

95. Halsey TC (1987) Some consequences of an equation of motion for diffusive growth. Phys Rev Lett 59:2067–2070

96. Turkevich LA, Scher H (1985) Occupancy-probability scaling in diffusion-limited aggregation. Phys Rev Lett 55(9):1026–1029

97. Ball RC, Spivack OR (1990) The interpretation and measurement of the f(alpha) spectrum of a multifractal measure. J Phys A 23:5295–5307

98. Jensen MH, Levermann A, Mathiesen J, Procaccia I (2002) Multifractal structure of the harmonic measure of diffusion-limited aggregates. Phys Rev E 65:046109

99. Ball RC (1986) Diffusion limited aggregation and its response to anisotropy. Physica A: Stat Theor Phys 140(1–2):62–69

100. Barker PW, Ball RC (1990) Real-space renormalization of diffusion-limited aggregation. Phys Rev A 42(10):6289–6292

101. Ball RC, Witten TA (1984) Causality bound on the density of aggregates. Phys Rev A 29(5):2966–2967

102. Halsey TC, Leibig M (1992) Theory of branched growth. Phys Rev A 46:7793–7809

103. Halsey TC (1994) Diffusion-limited aggregation as branched growth. Phys Rev Lett 72(8):1228–1231

104. Halsey TC, Duplantier B, Honda K (1997) Multifractal dimensions and their fluctuations in diffusion-limited aggregation. Phys Rev Lett 78(9):1719–1722

105. Erzan A, Pietronero L, Vespignani A (1995) The fixed-scale transformation approach to fractal growth. Rev Mod Phys 67(3):545–604

106. Cafiero R, Pietronero L, Vespignani A (1993) Persistence of screening and self-criticality in the scale-invariant dynamics of diffusion-limited aggregation. Phys Rev Lett 70(25):3939–3942

107. Ball RC, Somfai E (2003) Diffusion-controlled growth: Theory and closure approximations. Phys Rev E 67(2):021401, part 1
108. Ristroph L, Thrasher M, Mineev-Weinstein MB, Swinney HL (2006) Fjords in viscous fingering: Selection of width and opening angle. Phys Rev E 74(1):015201, part 2
109. Mineev-Weinstein MB, Wiegmann PB, Zabrodin A (2000) Integrable structure of interface dynamics. Phys Rev Lett 84(22):5106–5109
110. Abanov A, Mineev-Weinstein MB, Zabrodin A (2007) Self-similarity in Laplacian growth. Physica D – Nonlinear Phenomena 235(1–2):62–71

Books and Reviews

Vicsek T (1992) Fractal Growth Phenomena, 2nd edn. World Scientific, Singapore
Meakin P (1998) Fractals, scaling, and growth far from equilibrium. Cambridge University Press, Cambridge
Godreche G (1991) Solids far from equilibrium. Cambridge, Cambridge, New York
Sander LM (2000) Diffusion limited aggregation, a kinetic critical phenomenon? Contemporary Physics 41:203–218
Halsey TC (2000) Diffusion-limited aggregation: A model for pattern formation. Physics Today 53(4)

Fractal and Multifractal Scaling of Electrical Conduction in Random Resistor Networks

SIDNEY REDNER
Center for Polymer Studies and Department of Physics, Boston University, Boston, USA

Article Outline

Glossary

Conductance (G) The relation between the current I in an electrical network and the applied voltage V: $I = GV$.

Conductance exponent (t) The relation between the conductance G and the resistor (or conductor) concentration p near the percolation threshold: $G \sim (p - p_c)^t$.

Effective medium theory (EMT) A theory to calculate the conductance of a heterogeneous system that is based on a homogenization procedure.

Fractal A geometrical object that is invariant at any scale of magnification or reduction.

Multifractal A generalization of a fractal in which different subsets of an object have different scaling behaviors.

Percolation Connectivity of a random porous network.

Percolation threshold p_c The transition between a connected and disconnected network as the density of links is varied.

Random resistor network A percolation network in which the connections consist of electrical resistors that are present with probability p and absent with probability $1 - p$.

Definition of the Subject

Consider an arbitrary network of nodes connected by links, each of which is a resistor with a specified electrical resistance. Suppose that this network is connected to the leads of a battery. Two natural scenarios are: (a) the "bus-bar geometry" (Fig. 1), in which the network is connected to two parallel lines (in two dimensions), plates (in three dimensions), etc., and the battery is connected across the two plates, and (b) the "two-point geometry", in which a battery is connected to two distinct nodes, so that a current I injected at a one node and the same current withdrawn from the other node. In both cases, a basic question is: what is the nature of the current flow through the network?

There are many reasons why current flows in resistor networks have been the focus of more than a century of research. First, understanding currents in networks is one of the earliest subjects in electrical engineering. Second, the development of this topic has been characterized by beautiful mathematical advancements, such as Kirchhoff's formal solution for current flows in networks in terms of tree matrices [52], symmetry arguments to determine the electrical conductance of continuous two-component media [10,29,48,69,74], clever geometrical methods to simplify networks [33,35,61,62], and the use of integral transform methods to solve node voltages on regular networks [6,16,94,95]. Third, the nodes voltages of a network through which a steady electrical current flows are *harmonic* [26]; that is, the voltage at a given node is a suitably-weighted average of the voltages at neighboring nodes. This same harmonicity also occurs in the probability distribution of random walks. Consequently, there are deep connections between the probability distribution of
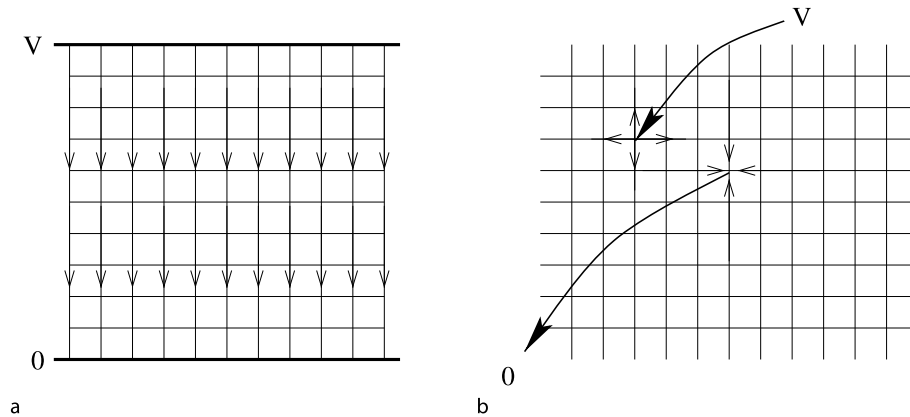
Fractal and Multifractal Scaling of Electrical Conduction in Random Resistor Networks, Figure 1
Resistor networks in a the bus-bar geometry, and b the two-point geometry

random walks on a given network and the node voltages on the same network [26].

Another important theme in the subject of resistor networks is the essential role played by *randomness* on current-carrying properties. When the randomness is weak, *effective medium theory* [10,53,54,55,57,74] is appropriate to characterize how the randomness affects the conductance. When the randomness is strong, as embodied by a network consisting of a random mixture of resistors and insulators, this *random resistor network* undergoes a transition between a conducting phase and an insulating phase when the resistor concentration passes through a percolation threshold [54]. The feature underlying this phase change is that for a small density of resistors, the network consists of disconnected clusters. However, when the resistor density passes through the percolation threshold, a macroscopic cluster of resistors spans the system through which current can flow. Percolation phenomenology has motivated theoretical developments, such as scaling, critical point exponents, and multifractals that have advanced our understanding of electrical conduction in random resistor networks.

This article begins with an introduction to electrical current flows in networks. Next, we briefly discuss analytical methods to solve the conductance of an arbitrary resistor network. We then turn to basic results related to percolation: namely, the conduction properties of a large random resistor network as the fraction of resistors is varied. We will focus on how the conductance of such a network vanishes as the percolation threshold is approached from above. Next, we investigate the more microscopic current *distribution* within each resistor of a large network. At the percolation threshold, this distribution is *multifractal* in that all moments of this distribution have indepen-

dent scaling properties. We will discuss the meaning of multifractal scaling and its implications for current flows in networks, especially the largest current in the network. Finally, we discuss the relation between resistor networks and random walks and show how the classic phenomena of recurrence and transience of random walks are simply related to the conductance of a corresponding electrical network.

The subject of current flows on resistor networks is a vast subject, with extensive literature in physics, mathematics, and engineering journals. This review has the modest goal of providing an overview, from my own myopic perspective, on some of the basic properties of random resistor networks near the percolation threshold. Thus many important topics are simply not mentioned and the reference list is incomplete because of space limitations. The reader is encouraged to consult the review articles listed in the reference list to obtain a more complete perspective.

Introduction

In an elementary electromagnetism course, the following classic problem has been assigned to many generations of physics and engineering students: consider an infinite square lattice in which each bond is a 1 ohm resistor; equivalently, the conductance of each resistor (the inverse resistance) also equals 1. There are perfect electrical connections at all vertices where four resistors meet. A current I is injected at one point and the same current I is extracted at a nearest-neighbor lattice point. What is the electrical resistance between the input and output? A more challenging question is: what is the resistance between two diagonal points, or between two arbitrary points? As we

shall discuss, the latter questions can be solved elegantly using Fourier transform methods.

For the resistance between neighboring points, superposition provides a simple solution. Decompose the current source and sink into its two constituents. For a current source I, symmetry tells us that a current $I/4$ flows from the source along each resistor joined to this input. Similarly, for a current sink $-I$, a current $I/4$ flows into the sink along each adjoining resistor. For the source/sink combination, superposition tells us that a current $I/2$ flows along the resistor directly between the source and sink. Since the total current is I, a current of $I/2$ flows indirectly from source to sink via the rest of the lattice. Because the direct and indirect currents between the input and output points are the same, the resistance of the direct resistor and the resistance of rest of the lattice are the same, and thus both equal to 1. Finally, since these two elements are connected in parallel, the resistance of the infinite lattice between the source and the sink equals 1/2 (conductance 2). As we shall see in Sect. "Effective Medium Theory", this argument is the basis for constructing an effective medium theory for the conductance of a random network.

More generally, suppose that currents I_i are injected at each node of a lattice network (normally many of these currents are zero and there would be both positive and negative currents in the steady state). Let V_i denote the voltage at node i. Then by Kirchhoff's law, the currents and voltages are related by

$$I_i = \sum_j g_{ij}(V_i - V_j)\,, \tag{1}$$

where g_{ij} is the conductance of link ij, and the sum runs over all links ij. This equation simply states that the current flowing into a node by an external current source equals the current flowing out of the node along the adjoining resistors. The right-hand side of Eq. (1) is a *discrete Laplacian* operator. Partly for this reason, Kirchhoff's law has a natural connection to random walks. At nodes where the external current is zero, the node voltages in Eq. (1) satisfy

$$V_i = \frac{\sum_j g_{ij} V_j}{\sum_j g_{ij}} \rightarrow \frac{1}{z}\sum_j V_j\,. \tag{2}$$

The last step applies if all the conductances are identical; here z is the coordination number of the network. Thus for steady current flow, the voltage at each unforced node equals the weighted average of the voltages at the neighboring sites. This condition defines V_i as a *harmonic function* with respect to the weight function g_{ij}.

An important general question is the role of spatial disorder on current flows in networks. One important exam-

ple is the *random resistor network*, where the resistors of a lattice are either present with probability p or absent with probability $1 - p$ [54]. Here the analysis tools for regular lattice networks are no longer applicable, and one must turn to qualitative and numerical approaches to understand the current-carrying properties of the system. A major goal of this article is to outline the essential role that spatial disorder has on the current-carrying properties of a resistor network by such approaches.

A final issue that we will discuss is the deep relation between resistor networks and random walks [26,63]. Consider a resistor network in which the positive terminal of a battery (voltage $V = 1$) is connected to a set of boundary nodes, defined to be $\mathcal{B}_+$), and where a disjoint set of boundary nodes $\mathcal{B}_-$ are at $V = 0$. Now suppose that a random walk hops between nodes of the same geometrical network in which the probability of hopping from node i to node j in a single step is $g_{ij}/\sum_k g_{ik}$, where k is one of the neighbors of i and the boundary sets are absorbing. For this random walk, we can ask: what is the probability F_i for a walk to eventually be absorbed on $\mathcal{B}_+$ when it starts at node i? We shall show in Sect. "Random Walks and Resistor Networks" that F_i satisfies Eq. (2): $F_i = \sum_j g_{ij} F_j/\sum_j g_{ij}$! We then exploit this connection to provide insights about random walks in terms of known results about resistor networks and vice versa.

Solving Resistor Networks

Fourier Transform

The translational invariance of an infinite lattice resistor network with identical bond conductances $g_{ij} = 1$ cries out for applying Fourier transform methods to determine node voltages. Let's study the problem mentioned previously: what is the voltage at any node of the network when a unit current enters at some point? Our discussion is specifically for the square lattice; the extension to other lattices is straightforward.

For the square lattice, we label each site i by its x, y coordinates. When a unit current is injected at $\mathbf{r_0} = (x_0, y_0)$, Eq. (1) becomes

$$-\delta_{x,x_0}\,\delta_{y,y_0} = V(x+1,y) + V(x-1,y)$$
$$+ V(x,y+1) + V(x,y-1) - 4V(x,y)\,, \tag{3}$$

which clearly exposes the second difference operator of the discrete Laplacian. To find the node voltages, we define $V(\mathbf{k}) = \sum_{\mathbf{r}} V(\mathbf{r})\,e^{i\mathbf{k}\cdot\mathbf{r}}$ and then we Fourier transform Eq. (3) to convert this infinite set of difference equations

into the single algebraic equation

$$V(\mathbf{k}) = \frac{e^{i\mathbf{k}\cdot\mathbf{r_0}}}{4 - 2(\cos k_x + \cos k_y)} \ . \tag{4}$$

Now we calculate $V(\mathbf{r})$ by inverting the Fourier transform

$$V(\mathbf{r}) = \frac{1}{(2\pi)^2} \int\limits_{-\pi}^{\pi} \int\limits_{-\pi}^{\pi} \frac{e^{-i\mathbf{k}\cdot(\mathbf{r}-\mathbf{r_0})}}{4 - 2(\cos k_x + \cos k_y)} \ d\mathbf{k} \ . \tag{5}$$

Formally, at least, the solution is trivial. However, the integral in the inverse Fourier transform, known as a Watson integral [96], is non-trivial, but considerable understanding has gradually been developed for evaluating this type of integral [6,16,94,95,96].

For a unit input current at the origin and a unit sink of current at $\mathbf{r_0}$, the resistance between these two points is $V(0) - V(\mathbf{r_0})$, and Eq. (5) gives

$$R = V(0) - V(\mathbf{r_0})$$
$$= \frac{1}{(2\pi)^2} \int\limits_{-\pi}^{\pi} \int\limits_{-\pi}^{\pi} \frac{(1 - e^{i\mathbf{k}\cdot\mathbf{r_0}})}{4 - 2(\cos k_x + \cos k_y)} \ d\mathbf{k} \ . \tag{6}$$

Tables for the values of R for a set of closely-separated input and output points are given in [6,94]. As some specific examples, for $\mathbf{r_0} = (1, 0)$, $R = 1/2$, thus reproducing the symmetry argument result. For two points separated by a diagonal, $\mathbf{r_0} = (1, 1)$, $R = 2/\pi$. For $\mathbf{r_0} = (2, 0)$, $R = 2 - 4/\pi$. Finally, for two points separated by a knight's move, $\mathbf{r_0} = (2, 1)$, $R = 4/\pi - 1/2$.

Direct Matrix Solution

Another way to solve Eq. (1), is to recast Kirchhoff's law as the matrix equation

$$I_i = \sum_{j=1}^{N} G_{ij} V_j \ , \quad i = 1, 2, \dots, N \tag{7}$$

where the elements of the *conductance matrix* are:

$$G_{ij} = \begin{cases} \sum_{k \neq i} g_{ik} \ , & i = j \\ -g_{ij} \ , & i \neq j \ . \end{cases}$$

The conductance matrix is an example of a *tree matrix*, as **G** has the property that the sum of any row or any column equals zero. An important consequence of this tree property is that all cofactors of G are identical and are equal to the *spanning tree polynomial* [43]. This polynomial is obtained by enumerating all possible tree graphs

(graphs with no closed loops) on the original electrical network that includes each node of the network. The weight of each spanning tree is simply the product of the conductances for each bond in the tree.

Inverting Eq. (7), one obtains the voltage V_i at each node i in terms of the external currents I_j ($j = 1, 2, \dots, N$) and the conductances g_{ij}. Thus the two-point resistance R_{ij} between two arbitrary (not necessarily connected) nodes i and j is then given by $R_{ij} = (V_i - V_j)/I$, where the network is subject to a specified external current; for example, for the two-point geometry, $I_i = 1$, $I_j = -1$, and $I_k = 0$ for $k \neq i, j$. Formally, the two-point resistance can be written as [99]

$$R_{ij} = \frac{|G^{(ij)}|}{|G^{(j)}|} \ , \tag{8}$$

where $|G^{(j)}|$ is the determinant of the conductance matrix with the jth row and column removed and $|G^{(ij)}|$ is the determinant with the ith and jth rows and columns removed. There is a simple geometric interpretation for this conductance matrix inversion. The numerator is just the spanning tree polynomial for the original network, while the denominator is the spanning tree polynomial for the network with the additional constraint that nodes i and j are identified as a single point. This result provides a concrete prescription to compute the conductance of an arbitrary network. While useful for small networks, this method is prohibitively inefficient for larger networks because the number of spanning trees grows exponentially with network size.

Potts Model Connection

The matrix solution of the resistance has an alternative and elegant formulation in terms of the spin correlation function of the q-state Potts model of ferromagnetism in the $q \to 0$ limit [88,99]. This connection between a statistical mechanical model in a seemingly unphysical limit and an enumerative geometrical problem is one of the unexpected charms of statistical physics. Another such example is the n-vector model, in which ferromagnetically interacting spins "live" in an n-dimensional spin space. In the limit $n \to 0$ [20], the spin correlation functions of this model are directly related to all self-avoiding walk configurations.

In the q-state Potts model, each site i of a lattice is occupied by a spin s_i that can assume one of q discrete values. The Hamiltonian of the system is

$$\mathcal{H} = -\sum_{i,j} J \delta_{s_i, s_j},$$

where the sum is over all nearest-neighbor interacting spin pairs, and δ_{s_i,s_j} is the Kronecker delta function ($\delta_{s_i,s_j} = 1$ if $s_i = s_j$ and $\delta_{s_i,s_j} = 0$ otherwise). Neighboring aligned spin pairs have energy $-J$, while spin pairs in different states have energy zero. One can view the spins as pointing from the center to a vertex of a q-simplex, and the interaction energy is proportional to the dot product of two interacting spins.

The partition function of a system of N spins is

$$Z_N = \sum_{\{s\}} e^{\beta \sum_{i,j} J \delta_{s_i,s_j}} , \qquad (9)$$

where the sum is over all 2^N spin states $\{s\}$. To make the connection to resistor networks, notice that: (i) the exponential factor associated with each link ij in the partition function takes the values 1 or $e^{\beta J}$, and (ii) the exponential of the sum can be written as the product

$$Z_N = \sum_{\{s_i\}} \prod_{i,j} (1 + v \delta_{s_i,s_j}) , \qquad (10)$$

with $v = \tanh \beta J$. We now make a high-temperature (small-v) expansion by multiplying out the product in (10) to generate all possible graphs on the lattice, in which each bond carries a weight $v \delta_{s_i,s_j}$. Summing over all states, the spins in each disjoint cluster must be in the same state, and the last sum over the common state of all spins leads to each cluster being weighted by a factor of q. The partition function then becomes

$$Z_N = \sum_{\text{graphs}} q^{N_c} v^{N_b} , \qquad (11)$$

where N_c is the number of distinct clusters and N_b is the total number of bonds in the graph.

It was shown by Kasteleyn and Fortuin [47] that the limit $q = 1$ corresponds to the percolation problem when one chooses $v = p / (1 - p)$, where p is the bond occupation probability in percolation. Even more striking [34], if one chooses $v = \alpha q^{1/2}$, where α is a constant, then $\lim_{q \to 0} Z_N / q^{(N+1)/2} = \alpha^{N-1} T_N$, where T_N is again the spanning tree polynomial; in the case where all interactions between neighboring spins have the same strength, then the polynomial reduces to the number of spanning trees on the lattice. It is because of this connection to spanning trees that the resistor network and Potts model are intimately connected [99]. In a similar vein, one can show that the correlation function between two spins at nodes i and j in the Potts model is simply related to the conductance between these same two nodes when the interactions J_{ij} between the spins at nodes i and j are equal to the conductances g_{ij} between these same two nodes in the corresponding resistor network [99].
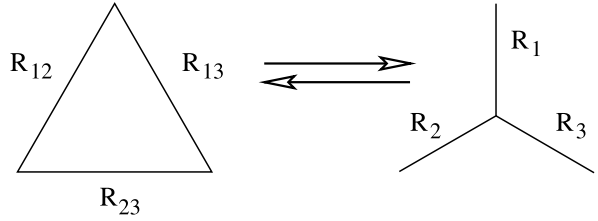
Fractal and Multifractal Scaling of Electrical Conduction in Random Resistor Networks, Figure 2
Illustration of the Δ-Y and Y-Δ transforms

Δ-Y and Y-Δ Transforms

In elementary courses on circuit theory, one learns how to combine resistors in series and parallel to reduce the complexity of an electrical circuit. For two resistors with resistances R_1 and R_2 in series, the net resistance is $R = R_1 + R_2$, while for resistors in parallel, the net resistance is $R = \left(R_1^{-1} + R_2^{-2}\right)^{-1}$. These rules provide the resistance of a network that contains only series and parallel connections. What happens if the network is more complicated? One useful way to simplify such a network is by the Δ-Y and Y-Δ transforms that was apparently first discovered by Kennelly in 1899 [50] and applied extensively since then [33,35,61,62,81].

The basic idea of the Δ-Y transform is illustrated in Fig. 2. Any triangular arrangement of resistors R_{12}, R_{13}, and R_{23} within a larger circuit can be replaced by a star, with resistances R_1, R_2, and R_3, such that all resistances between any two points among the three vertices in the triangle and the star are the same. The conditions that all two-point resistances are the same are:

$$(R_1 + R_2) = \left[R_{12}^{-1} + (R_{13} + R_{23})^{-1}\right]^{-1}$$
$$\equiv a_{12} \quad + \text{cyclic permutations} .$$

Solving for R_1, R_2, and R_3 gives $R_1 = \frac{1}{2}(a_{12} - a_{23} + a_{13})$ + cyclic permutations; the explicit result in terms of the R_{ij} is:

$$R_1 = \frac{R_{12} R_{13}}{R_{12} + R_{13} + R_{23}} \quad + \text{cyclic permutations}, \quad (12)$$

as well as the companion result for the conductances $G_i = R_i^{-1}$:

$$G_1 = \frac{G_{12} G_{13} + G_{12} G_{23} + G_{13} G_{23}}{G_{23}} + \text{cyclic permutations}.$$

These relations allow one to replace any triangle by a star to reduce an electrical network.

However, sometimes we need to replace a star by a triangle to simplify a network. To construct the inverse Y-Δ

transform, notice that the Δ-Y transform gives the resistance ratios $R_1/R_2 = R_{13}/R_{23}$ + cyclic permutations, from which $R_{13} = R_{12}(R_3/R_2)$ and $R_{23} = R_{12}(R_3/R_1)$. Substituting these last two results in Eq. (12), we eliminate R_{13} and R_{23} and thus solve for R_{12} in terms of the R_i:

$$R_{12} = \frac{R_1 R_2 + R_1 R_3 + R_2 R_3}{R_3} + \text{cyclic permutations}, \quad (13)$$

and similarly for $G_{ij} = R_{ij}^{-1}$. To appreciate the utility of the Δ-Y and Y-Δ transforms, the reader is invited to apply them on the Wheatstone bridge.

When employed judiciously and repeatedly, these transforms systematically reduce planar lattice circuits to a single bond, and thus provide a powerful approach to calculate the conductance of large networks near the percolation threshold. We will return to this aspect of the problem in Sect. "Conductance Exponent".

Effective Medium Theory

Effective medium theory (EMT) determines the macroscopic conductance of a random resistor network by a homogenization procedure [10,53,54,55,57,74] that is reminiscent of the Curie–Weiss effective field theory of magnetism. The basic idea in EMT is to replace the random network by an effective homogeneous medium in which the conductance of each resistor is determined self-consistently to optimally match the conductances of the original and homogenized systems. EMT is quite versatile and has been applied, for example, to estimate the dielectric constant of dielectric composites and the conductance of conducting composites. Here we focus on the conductance of random resistor networks, in which each resistor (with conductance g_0) is present with probability p and absent with probability $1 - p$. The goal is to determine the conductance as a function of p.

To implement EMT, we first replace the random network by an effective homogeneous medium in which each bond has the same conductance g_m (Fig. 3). If a voltage is applied across this effective medium, there will be a potential drop V_m and a current $I_m = g_m V_m$ across each bond. The next step in EMT is to assign one bond in the effective medium a conductance g and adjust the external voltage to maintain a fixed total current I passing through the network. Now an additional current δi passes through the conductor g. Consequently, a current $-\delta i$ must flow through one terminal of g to the other terminal via the remainder of the network (Fig. 3). This current perturbation leads to an additional voltage drop δV across g. Thus the current-voltage relations for the marked bond and the re-
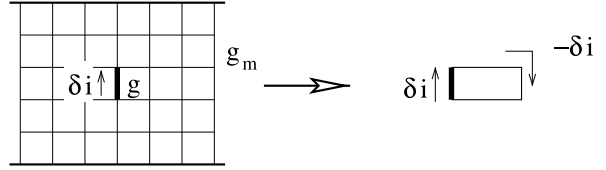
Fractal and Multifractal Scaling of Electrical Conduction in Random Resistor Networks, Figure 3
Illustration of EMT. (*left*) The homogenized network with conductances g_m and one bond with conductance g. (*right*) The equivalent circuit to the lattice

mainder of the network are

$$I_m + \delta i = g(V_m + \delta V)$$
$$-\delta i = G_{ab}\delta V, \quad (14)$$

where G_{ab} is the conductance of the rest of the lattice between the terminals of the conductor g.

The last step in EMT is to require that the mean value δV averaged over the probability distribution of individual bond conductances is zero. Thus the effective medium "matches" the current-carrying properties of the original network. Solving Eq. (14) for δV, and using the probability distribution $P(g) = p\delta(g - g_0) + (1 - p)\delta(g)$ appropriate for the random resistor network, we obtain

$$\langle \delta V \rangle = V_m \left[\frac{(g_m - g_0)p}{(G_{ab} + g_0)} + \frac{g_m(1 - p)}{G_{ab}} \right] = 0. \quad (15)$$

It is now convenient to write $G_{ab} = \alpha g_m$, where α is a lattice-dependent constant of the order of one. With this definition, Eq. (15) simplifies to

$$g_m = g_o \frac{p(1 + \alpha) - 1}{\alpha}. \quad (16)$$

The value of α – the proportionality constant for the conductance of the initial lattice with a single bond removed – can usually be determined by a symmetry argument of the type presented in Sect. "Introduction to Current Flows". For example, for the triangular lattice (coordination number 6), the conductance $G_{ab} = 2g_m$ and $\alpha = 2$. For the hypercubic lattice in d dimensions (coordination number $z = 2^d$), $G_{ab} = ((z - 2)/2)g_m$.

The main features of the effective conductance g_m that arises from EMT are: (i) the conductance vanishes at a lattice-dependent percolation threshold $p_c = 1/(1 + \alpha)$; for the hypercubic lattice $\alpha = (z - 2)/2$ and the percolation threshold $p_c = 2/z = 2^{1-d}$ (fortuitously reproducing the exact percolation threshold in two dimensions); (ii) the conductance varies linearly with p and vanishes linearly in $p - p_c$ as p approaches p_c from above. The linearity

of the effective conductance away from the percolation threshold accords with numerical and experimental results. However, EMT fails near the percolation threshold, where large fluctuations arise that invalidate the underlying assumptions of EMT. In this regime, alternative methods are needed to estimate the conductance.

Conduction Near the Percolation Threshold

Scaling Behavior

EMT provides a qualitative but crude picture of the current-carrying properties of a random resistor network. While EMT accounts for the existence of a percolation transition, it also predicts a linear dependence of the conductance on p. However, near the percolation threshold it is well known that the conductance varies non-linearly in $p - p_c$ near p_c [85]. This non-linearity defines the conductance exponent t by

$$G \sim (p - p_c)^t \qquad p \downarrow p_c \,, \tag{17}$$

and much research on random resistor networks [85] has been performed to determine this exponent. The conductance exponent generically depends only on the spatial dimension of the network and not on any other details (a notable exception, however, is when link resistances are broadly distributed, see [40,93]). This *universality* is one of the central tenets of the theory of critical phenomena [64,83]. For percolation, the mechanism underlying universality is the absence of a characteristic length scale; as illustrated in Fig. 4, clusters on all length scales exist when a network is close to the percolation threshold.

The scale of the largest cluster defines the correlation length ξ by $\xi \sim (p_c - p)^{-\nu}$ as $p \to p_c$. The divergence in ξ also applies for $p > p_c$ by defining the correlation length as the typical size of finite clusters only (Fig. 4), thus eliminating the infinite percolating cluster from consideration. At the percolation threshold, clusters on all length scales exist, and the absence of a characteristic length implies that the singularity in the conductance should not depend on microscopic variables. The only parameter remaining upon which the conductance exponent t can depend upon is the spatial dimension d [64,83]. As typifies critical phenomena, the conductance exponent has a constant value in all spatial dimensions $d > d_c$, where d_c is the upper critical dimension which equals 6 for percolation [22]. Above this critical dimension, mean-field theory (not to be confused with EMT) gives the correct values of critical exponents.

While there does not yet exist a complete theory for the dimension dependence of the conductance exponent below the critical dimension, a crude but useful *nodes, links, and blobs* picture of the infinite cluster [21,82,84] provides partial information. The basic idea of this picture is that
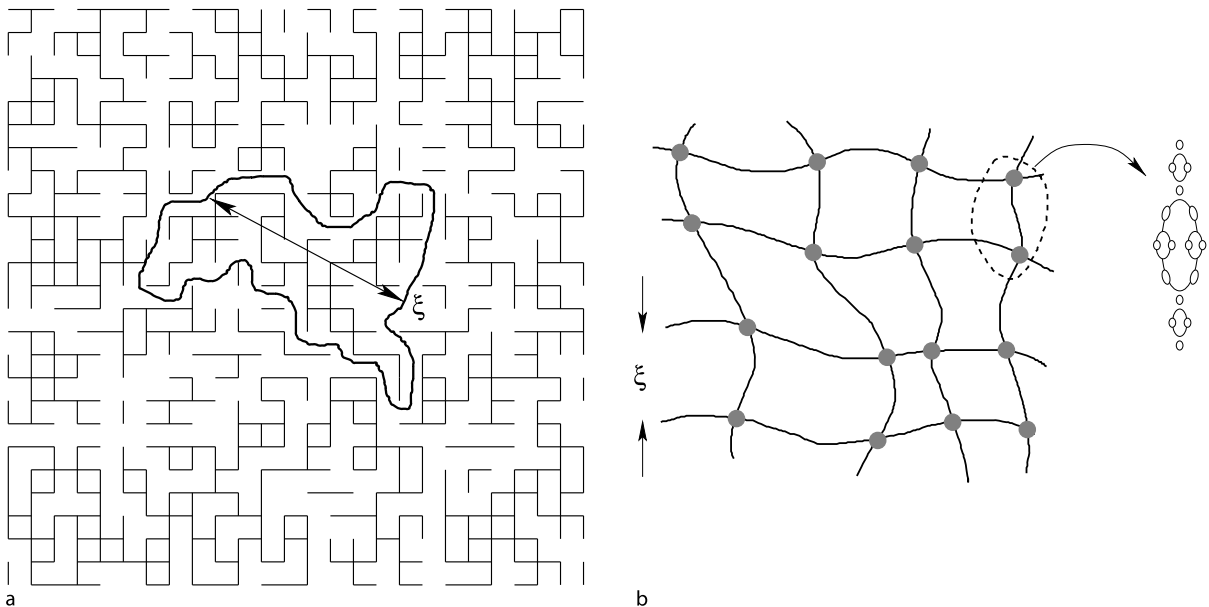
a   b

Fractal and Multifractal Scaling of Electrical Conduction in Random Resistor Networks, Figure 4
(*left*) Realization of bond percolation on a 25 × 25 square lattice at p=0.505. (*right*) Schematic picture of the nodes (shaded circles), links and blobs picture of percolation for $p \gtrsim p_c$

for $p \gtrsim p_c$, a large system has an irregular network-like topology that consists of quasi-linear chains that are separated by the correlation length ξ (Fig. 4). For a macroscopic sample of linear dimension L with a bus bar-geometry, the percolating cluster above p_c then consists of $(L/\xi)^{d-1}$ statistically identical chains in parallel, in which each chain consists of L/ξ macrolinks in series, and the macrolinks consists of nested blob-like structures.

The conductance of a macrolink is expected to vanish as $(p - p_c)^\zeta$, with ζ a new unknown exponent. Although a theory for the conductance of a single macrolink, and even a precise definition of a macrolink, is still lacking, the nodes, links, and blobs picture provides a starting point for understanding the dimension dependence of the conductance exponent. Using the rules for combining parallel and series conductances, the conductance of a large resistor network of linear dimension L is then

$$G(p, L) \sim \left(\frac{L}{\xi} \right)^{d-1} \frac{(p - p_c)^\zeta}{L/\xi}$$
$$\sim L^{d-2} (p - p_c)^{(d-2)\nu + \zeta} . \quad (18)$$

In the limit of large spatial dimension, we expect that a macrolink is merely a random walk between nodes. Since the spatial separation between nodes is ξ, the number of bonds in the macrolink, and hence its resistance, scales as ξ^2 [92]. Using the mean-field result $\xi \sim (p - p_c)^{-1/2}$, the resistance of the macrolink scales as $(p - p_c)^{-1}$ and thus the exponent $\zeta = 1$. Using the mean-field exponents $\nu = 1/2$ and $\zeta = 1$ at the upper critical dimension of $d_c = 6$, we then infer the mean-field value of the conductance exponent $t = 3$ [22,91,92].

Scaling also determines the conductance of a finite-size system of linear dimension L *exactly at* the percolation threshold. Although the correlation length formally diverges when $p - p_c = 0$, ξ is limited by L in a finite system of linear dimension L. Thus the only variable upon which the conductance can depend is L itself. Equivalently, deviations in $p - p_c$ that are smaller than $L^{-1/\nu}$ cannot influence critical behavior because ξ can never exceed L. Thus to determine the dependence of a singular observable for a finite-size system at p_c, we may replace $(p - p_c)$ by $L^{-1/\nu}$. By this prescription, the conductance at p_c of a large finite-size system of linear dimension L becomes

$$G(p_c, L) \sim L^{d-2} (L^{-1/\nu})^{(d-2)\nu + \zeta} \sim L^{-\zeta/\nu} . \quad (19)$$

In this *finite-size scaling* [85], we fix the occupation probability to be exactly at p_c and study the dependence of an observable on L to determine percolation exponents. This approach provides a convenient and more accurate method to determine the conductance exponent com-

pared to studying the dependence of the conductance of a large system as a function of $p - p_c$.

Conductance Exponent

In percolation and in the random resistor network, much effort has been devoted to computing the exponents that characterize basic physical observables – such as the correlation length ξ and the conductance G – to high precision. There are several reasons for this focus on exponents. First, because of the universality hypothesis, exponents are a meaningful quantifier of phase transitions. Second, various observables near a phase transition can sometimes be related by a scaling argument that leads to a corresponding exponent relation. Such relations may provide a decisive test of a theory that can be checked numerically. Finally, there is the intellectual challenge of developing accurate numerical methods to determine critical exponents. The best such methods have become quite sophisticated in their execution.

A seminal contribution was the "theorists' experiment" of Last and Thouless [58] in which they punched holes at random in a conducting sheet of paper and measured the conductance of the sheet as a function of the area fraction of conducting material. They found that the conductance vanished faster than linearly with $(p - p_c)$; here p corresponds to the area fraction of the conductor. Until this experiment, there was a sentiment that the conductance should be related to the fraction of material in the percolating cluster [30] – the percolation probability $P(p)$ – a quantity that vanished slower than linearly with $(p - p_c)$. The reason for this disparity is that in a resistor network, much of the percolating cluster consists of *dangling ends* – bonds that carry no current – and thus make no contribution to the conductance. A natural geometrical quantity that ought to be related to the conductance is the fraction of bonds $B(p)$ in the *conducting backbone* – the subset of the percolating cluster without dangling ends. However, a clear relation between the conductivity and a geometrical property of the backbone has not yet been established.

Analytically, there are primary two methods that have been developed to compute the conductance exponent: the renormalization group [44,86,87,89] and low-density series expansions [1,2,32]. In the real-space version of the renormalization group, the evolution of conductance distribution under length rescaling is determined, while the momentum-space version involves a diagrammatic implementation of this length rescaling in momentum space. The latter is a perturbative approach away from mean-field theory in the variable $6 - d$ that become exact as $d \to 6$.

Considerable effort has been devoted to determining the conductance exponent by numerical and algorithmic methods. Typically, the conductance is computed for networks of various linear dimensions L at $p = p_c$, and the conductance exponent is extracted from the L dependence of the conductance, which should vanish as $L^{-\zeta/\nu}$. An exact approach, but computationally impractical for large networks, is Gauss elimination to invert the conductance matrix [79]. A simple approximate method is Gauss relaxation [59,68,80,90,97] (and its more efficient variant of Gauss–Seidel relaxation [71]). This method uses Eq. (2) as the basis for an iteration scheme, in which the voltage V_i at node i at the nth update step is computed from (2) using the values of V_j at the $(n-1)^{st}$ update in the right-hand side of this equation. However, one can do much better by the conjugate gradient algorithm [27] and speeding up this method still further by Fourier acceleration methods [7].

Another computational approach is based on the node elimination method, in which the Δ-Y and Y-Δ transforms are used to successively eliminate bonds from the network and ultimately reduce a large network to a single bond [33,35,62]. In a different vein, the transfer matrix method has proved to be extremely accurate and efficient [24,25,70,100]. The method is based on building up the network one bond at a time and immediately calculating the conductance of the network after each bond addition. This method is most useful when applied to very long strips of transverse dimension L so that a single realization gives an accurate value for the conductance.

As a result of these investigations, as well as by series expansions for the conductance, the following exponents have been found. For $d = 2$, where most of the computational effort has been applied, the best estimate [70] for the exponent t (using $\zeta = t$ in $d = 2$ only) is $t = 1.299 \pm 0.002$. One reason for the focus on two dimensions is that early estimates for t were tantalizingly close to the correlation length exponent ν that is now known to exactly equal $4/3$ [23]. Another such connection was the Alexander–Orbach conjecture [5], which predicted $t = 91/72 = 1.2638\ldots$, but again is incompatible with the best numerical estimate for t. In $d = 3$, the best available numerical estimate for t appears to be $t = 2.003 \pm 0.047$ [12,36], while the low concentration series method gives an equally precise result of $t = 2.02 \pm 0.05$ [1,2]. These estimates are just compatible with the rigorous bound that $t \leq 2$ in $d = 3$ [37,38]. In greater than three dimensions, these series expansions give $t = 2.40 \pm 0.03$ for $d = 4$ and $t = 2.74 \pm 0.03$ for $d = 5$, and the dimension dependence is consistent with $t = 3$ when d reaches 6.

Voltage Distribution in Random Networks

Multifractal Scaling

While much research has been devoted to understanding the critical behavior of the conductance, it was realized that the *distribution* of voltages across each resistor of the network was quite rich and exhibited *multifractal* scaling [17,19,72,73]. Multifractality is a generalization of fractal scaling in which the distribution of an observable is sufficiently broad that different moments of the distribution scale independently. Such multifractal scaling arises in phenomena as diverse as turbulence [45,66], localization [13], and diffusion-limited aggregation [41,42]. All these diverse examples showed scaling properties that were much richer than first anticipated.

To make the discussion of multifractality concrete, consider the example of the Maxwell–Boltzmann velocity distribution of a one-dimensional ideal gas

$$P(v) = \sqrt{\frac{m}{2\pi k_B T}}\, e^{-mv^2/2k_B T} \equiv \frac{1}{\sqrt{2\pi v_{th}^2}}\, e^{-v^2/2v_{th}^2}\,,$$

where k_B is Boltzmann's constant, m is the particle mass, T is the temperature, and $v_{th} = \sqrt{k_B T/m}$ is the characteristic thermal velocity. The even integer moments of the velocity distribution are

$$\langle (v^2)^n \rangle \propto (v_{th}^2)^n \equiv (v_{th}^2)^{p(n)}\,.$$

Thus a single velocity scale, v_{th}, characterizes all positive moments of the velocity distribution. Alternatively, the exponent $p(n)$ is linear in n. This linear dependence of successive moment exponents characterizes single-parameter scaling. The new feature of multifractal scaling is that a wide range of scales characterizes the voltage distribution (Fig. 5). As a consequence, the moment exponent $p(n)$ is a non-linear function of n.

One motivation for studying the voltage distribution is its relation to basic aspects of electrical conduction. If a voltage $V = 1$ is applied across a resistor network, then the conductance G and the total current flow I are equal: $I = G$. Consider now the power dissipated through the network $P = IV = GV^2 \rightarrow G$. We may also compute the dissipated power by adding up these losses in each resistor to give

$$P = G = \sum_{ij} g_{ij} V_{ij}^2 \rightarrow \sum_{ij} V_{ij}^2 = \sum_V V^2 N(V)\,. \quad (20)$$

Here $g_{ij} = 1$ is the conductance of resistor ij, and V_{ij} is the corresponding voltage drop across this bond. In the last equality, $N(V)$ is the number of resistors with a voltage
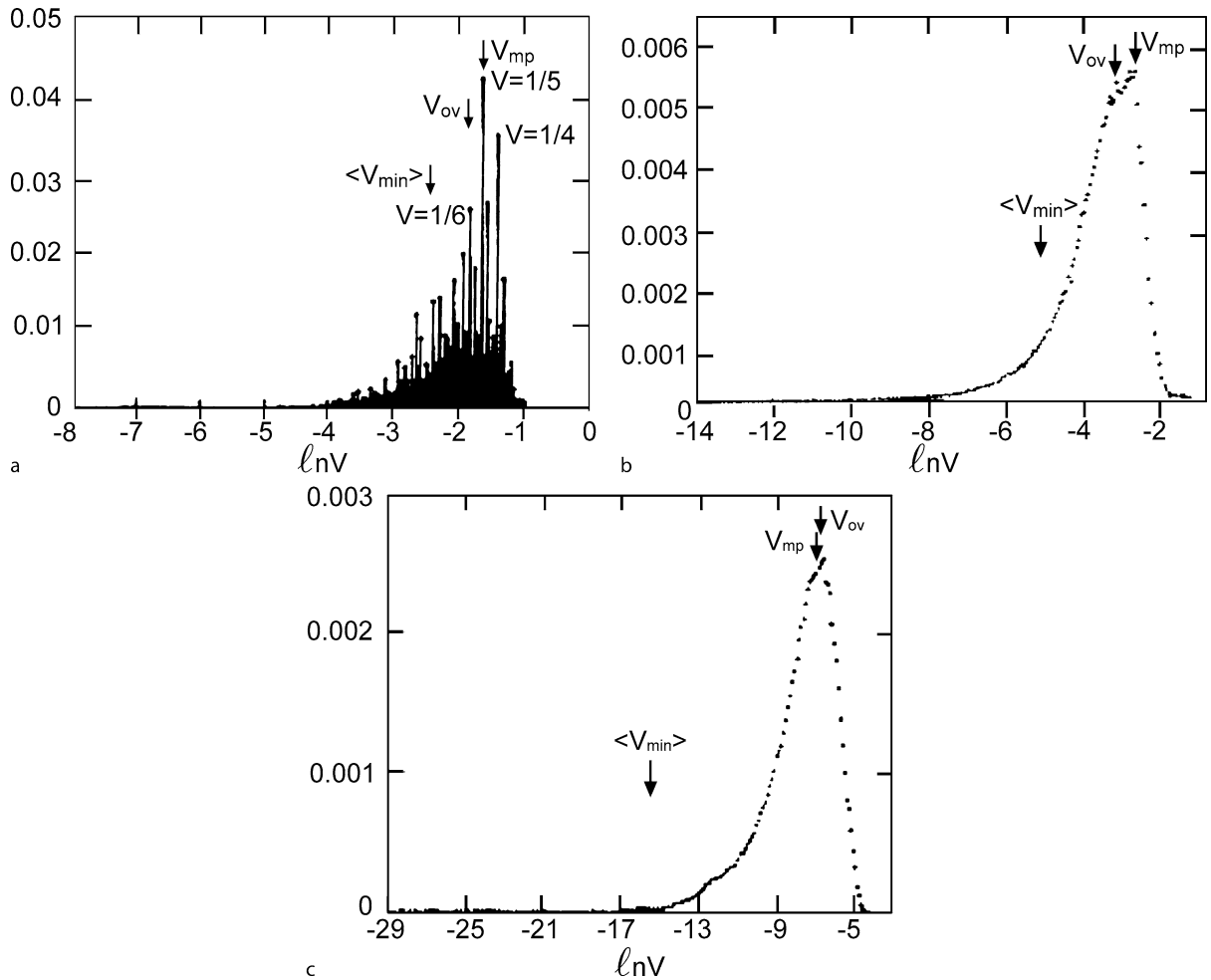
Fractal and Multifractal Scaling of Electrical Conduction in Random Resistor Networks, Figure 5
The voltage distribution on an $L \times L$ square lattice random resistor network at the percolation threshold for **a** $L = 4$ (exact), **b** $L = 10$, and **c** $L = 130$. The latter two plots are based on simulation data. For $L = 4$, a number of peaks, that correspond to simple rational fractions of the unit potential drop, are indicated. Also shown are the average voltage over all realizations, V_{av}, the most probable voltage, V_{mp}, and the average of the minimum voltage in each realization, $\langle V_{min} \rangle$. [Reprinted from Ref. [19]]

drop V. Thus the conductance is just the second moment of the distribution of voltage drops across each bond in the network.

From the statistical physics perspective it is natural to study other moments of the voltage distribution and the voltage distribution itself. Analogous to the velocity distribution, we define the family of exponents $p(k)$ for the scaling dependence of the voltage distribution at $p = p_c$ by

$$\mathcal{M}(k) \equiv \sum_V N(V) V^k \sim L^{-p(k)/\nu} . \qquad (21)$$

Since $\mathcal{M}(2)$ is just the network conductance, $p(2) = \zeta$. Other moments of the voltage distribution also have sim-

ple interpretations. For example, $\langle V^4 \rangle$ is related to the magnitude of the noise in the network [9,73], while $\langle V^k \rangle$ for $k \to \infty$ weights the bonds with the highest currents, or the "hottest" bonds of the network, most strongly, and they help understand the dynamics of fuse networks of failure [4,18]. On the other hand, negative moments weight low-current bonds more strongly and emphasize the low-voltage tail of the distribution. For example, $\mathcal{M}(-1)$ characterizes hydrodynamic dispersion [56], in which passive tracer particles disperse in a network due to a multiplicity of network paths. In hydrodynamics dispersion, the transit time across each bond is proportional to the inverse of the current in the bond, while the probability for tracer to enter a bond is proportional to the enter-
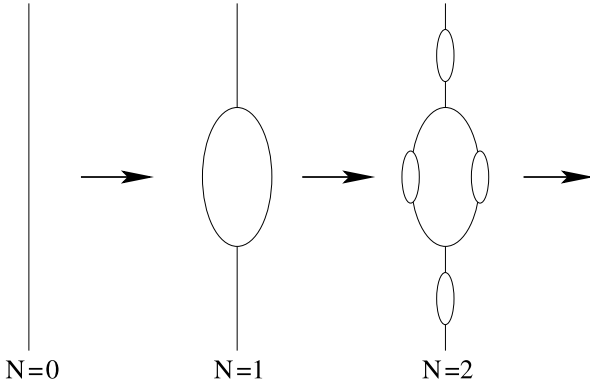
Fractal and Multifractal Scaling of Electrical Conduction in Random Resistor Networks, Figure 6
The first few iterations of a hierarchical model

ing current. As a result, the kth moment of the transit time distribution varies as $\mathcal{M}(-k + 1)$, so that the quantity that quantifies dispersion, $\langle t^2 \rangle - \langle t \rangle^2$, scales as $\mathcal{M}(-1)$.

A simple fractal model [3,11,67] of the conducting backbone (Fig. 6) illustrates the multifractal scaling of the voltage distribution near the percolation threshold [17]. To obtain the Nth-order structure, each bond in the $(N - 1)^{\text{st}}$ iteration is replaced by the first-order structure. The resulting fractal has a hierarchical embedding of links and blobs that captures the basic geometry of the percolating backbone. Between successive generations, the length scale changes by a factor of 3, while the number of bonds changes by a factor of 4. Defining the fractal dimension d_f as the scaling relation between mass $(M = 4^N)$ and the length scale $(\ell = 3^N)$ via $M \sim \ell^{d_f}$, gives a fractal dimension $d_f = \ln 4 / \ln 3$.

Now let's determine the distribution of voltage drops across the bonds. If a unit voltage is applied at the opposite ends of a first-order structure $(N = 1)$ and each bond is a 1 ohm resistor, then the two resistors in the central bubble each have a voltage drop of 1/5, while the two resistors at the ends have a voltage drop 2/5. In an Nth-order hierarchy, the voltage of any resistor is the product of these two factors, with number of times each factor occurs dependent on the level of embedding of a resistor within the blobs. It is a simple exercise to show that the voltage distribution is [17]

$$N(V(j)) = 2^N \binom{N}{j}, \qquad (22)$$

where the voltage $V(j)$ can take the values $2^j/5^N$ (with $j = 0, 1, \ldots, N$). Because j varies logarithmically in V, the voltage distribution is log binomial [75]. Using this distri-

bution in Eq. (21), the moments of the voltage distribution are

$$\mathcal{M}(k) = \left[\frac{2(1 + 2^k)}{5^k} \right]^N . \qquad (23)$$

In particular, the average voltage, $\mathcal{M}(1)/\mathcal{M}(0) \equiv V_{\text{av}}$ equals $((3/2)/5)^N$, which is very different from the most probable voltage, $V_{\text{mp}} = (\sqrt{2}/5)^N$ as $N \to \infty$. The underlying multiplicativity of the bond voltages is the ultimate source of the large disparity between the average and most probable values.

To calculate the moment exponent $p(k)$, we first need to relate the iteration index N to a physical length scale. For percolation, the appropriate relation is based on Coniglio's theorem [15], which is a simple but profound statement about the structure of the percolating cluster. This theorem states that the number of singly-connected bonds in a system of linear dimension L, $\mathcal{N}_s$, varies as $L^{1/\nu}$. Singly-connected bonds are those that would disconnect the network if they were cut. An equivalent form of the theorem is $\mathcal{N}_s = \partial p'/\partial p$, where p' is the probability that a spanning cluster exists in the system. This relation reflects the fact that when p is decreased slightly, p' changes only if a singly-connected bond happens to be deleted.

In the Nth-order hierarchy, the number of such singly-connected links is simply 2^N. Equating these two gives an effective linear dimension, $L = 2^{N\nu}$. Using this relation in (23), the moment exponent $p(k)$ is

$$p(k) = k - 1 + \left[k \ln (5/4) - \ln(1 + 2^{-k}) \right] / \ln 2 . \qquad (24)$$

Because each $p(k)$ is independent, the moments of the voltage distribution are characterized by an infinite set of exponents. Equation (24) is in excellent agreement with numerical data for the voltage distribution in two-dimensional random resistor networks at the percolation threshold [19]. A similar multifractal behavior was also found for the voltage distribution of the resistor network at the percolation threshold in three dimensions [8].

Maximum Voltage

An important aspect of the voltage distribution, both because of its peculiar scaling properties [27] and its application to breakdown problems [18,27], is the maximum voltage in a network. The salient features of this maximum voltage are: (i) logarithmic scaling as a function of system size [14,27,28,60,65], and (ii) non-monotonic dependence on the resistor concentration p [46]. The former property is a consequence of the expected size of the largest defect
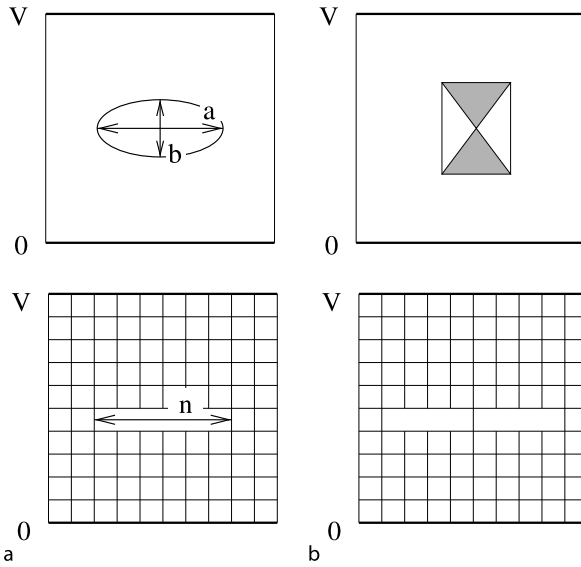
Fractal and Multifractal Scaling of Electrical Conduction in Random Resistor Networks, Figure 7
Defect configurations in two dimensions. a An ellipse and its square lattice counterpart, **b** a funnel, with the region of good conductor shown shaded, and a 2-slit configuration on the square lattice

in the network that gives maximal local currents. Here, we use the terms maximum local voltage and maximum local current interchangeably because they are equivalent.

To find the maximal current, we first need to identify the optimal defects that lead to large local currents. A natural candidate is an ellipse [27,28] with major and minor axes a and b (continuum), or its discrete analog of a linear crack (hyperplanar crack in greater than two dimensions) in which n resistors are missing (Fig. 7). Because current has to detour around the defect, the local current at the ends of the defect is magnified. For the continuum problem, the current at the tip of the ellipse is $I_{tip} = I_0(1 + a/b)$, where I_0 is the current in the unperturbed system [27]. For the maximum current in the lattice system, one must integrate the continuum current over a one lattice spacing and identify a/b with n [60]. This approach gives the maximal current at the tip of a crack $I_{max} \propto (1 + n^{1/2})$ in two dimensions and as $I_{max} \propto (1 + n^{1/2(d-1)})$ in d dimensions.

Next, we need to find the size of the largest defect, which is an extreme-value statistics exercise [39]. For a linear crack, each broken bond occurs with probability $1 - p$, so that the probability for a crack of length n is $(1 - p)^n \equiv e^{-an}$, with $a = -\ln(1 - p)$. In a network of volume L^d, we estimate the size of the largest defect by $L^d \int_{n_{max}}^{\infty} e^{-an} \, dx = 1$; that is, there exists of the order

of one defect of size n_{max} or larger in the network [39]. This estimate gives n_{max} varying as $\ln L$. Combining this result with the current at the tip of a crack of length n, the largest current in a system of linear dimension L scales as $(\ln L)^{1/2(d-1)}$.

A more thorough analysis shows, however, that a single crack is not quite optimal. For a continuum two-component network with conductors of resistance 1 with probability p and with resistance $r > 1$ with probability $1 - p$, the configuration that maximizes the local current is a funnel [14,65]. For a funnel of linear dimension ℓ, the maximum current at the apex of the funnel is proportional to $\ell^{1-\nu}$, where $\nu = (4/\pi) \tan^{-1}(r^{-1/2})$ [14,65]. The probability to find a funnel of linear dimension ℓ now scales as $e^{-b\ell^2}$ (exponentially in its area), with b a constant. By the same extreme statistics reasoning given above, the size of the largest funnel in a system of linear dimension L then scales as $(\ln L)^{1/2}$, and the largest expected current correspondingly scales as $(\ln L)^{(1-\nu)/2}$. In the limit $r \to \infty$, where one component is an insulator, the optimal discrete configuration in two dimensions becomes two parallel slits, each of length n, between which a single resistor remains [60]. For this two-slit configuration, the maximum current is proportional to n in two dimensions, rather than $n^{1/2}$ for the single crack. Thus the maximal current in a system of linear dimension L scales as $\ln L$ rather than as a fractional power of $\ln L$.

The p dependence of the maximum voltage is intriguing because it is non-monotonic. As p, the fraction of occupied bonds, decreases from 1, less total current flows (for a fixed overall voltage drop) because the conductance is decreasing, while local current in a funnel is enhanced because such defects grow larger. The competition between these two effects leads to V_{max} attaining its peak at p_{peak} above the percolation threshold that only slowly approaches p_c as $L \to \infty$. An experimental manifestation of this non-monotonicity in V_{max} occurred in a resistor-diode network [77], where the network reproducibly burned (solder connections melting and smoking) when $p \simeq 0.77$, compared to a percolation threshold of $p_c \simeq 0.58$. Although the directionality constraint imposed by diodes enhances funneling, similar behavior should occur in a random resistor network.

The non-monotonic p dependence of V_{max} can be understood within the quasi-one-dimensional "bubble" model [46] that captures the interplay between local funneling and overall current reduction as p decreases (Fig. 8). Although this system looks one-dimensional, it can be engineered to reproduce the percolation properties of a system in greater than one dimension by choosing the length L to scale exponentially with the width w. The prob-
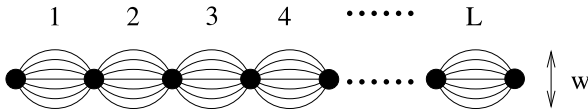
Fractal and Multifractal Scaling of Electrical Conduction in Random Resistor Networks, Figure 8
The bubble model: a chain of L bubbles in series, each consisting of w bonds in parallel. Each bond is independently present with probability p

ability for a spanning path in this structure is

$$
p' = \left[1 - (1-p)^w \right]^L
$$
$$
\rightarrow \exp[-L\,e^{-pw}] \qquad L, w \rightarrow \infty, \quad (25)
$$

which suddenly changes from 0 to 1 – indicative of percolation – at a threshold that lies strictly within (0,1) as $L \rightarrow \infty$ and $L \sim e^w$. In what follows, we take $L = 2^w$, which gives $p_c = 1/2$.

To determine the effect of bottlenecking, we appeal to the statement of Coniglio's theorem [15], $\partial p'/\partial p$ equals the average number of singly-connected bonds in the system. Evaluating $\partial p'/\partial p$ in Eq. (25) at the percolation threshold of $p_c = 1/2$ gives

$$
\frac{\partial p'}{\partial p} = w + \mathcal{O}(e^{-w}) \sim \ln L \,. \quad (26)
$$

Thus at p_c there are $w \sim \ln L$ bottlenecks. However, current focusing due to bottlenecks is substantially diluted because the conductance, and hence the total current through the network, is small at p_c. What is needed is a single bottleneck of width 1. One such bottleneck ensures the total current flow is still substantial, while the narrowing to width 1 endures that the focusing effect of the bottleneck is maximally effective.

Clearly, a single bottleneck of width 1 occurs above the percolation threshold. Thus let's determine when a such an isolated bottleneck of width 1 first appears as a function of p. The probability that a single non-empty bubble contains at least two bonds is $(1 - q^w - wpq^{w-1})/(1 - q^w)$, where $q = 1 - p$. Then the probability $P_1(p)$ that the width of the narrowest bottleneck has width 1 in a chain of L bubbles is

$$
P_1(p) = 1 - \left(1 - \frac{wpq^{w-1}}{1 - q^w} \right)^L
$$
$$
\sim 1 - \exp\left(-Lw \frac{p}{q(1-q^w)}\,e^{-pw} \right) \,. \quad (27)
$$

The subtracted term is the probability that L non-empty bubbles contain at least two bonds, and then $P_1(p)$ is the

complement of this quantity. As p decreases from 1, $P_1(p)$ sharply increases from 0 to 1 when the argument of the outer exponential becomes of the order of 1; this change occurs at $\hat{p} \sim p_c + \mathcal{O}(\ln(\ln L) / \ln L)$. At this point, a bottleneck of width 1 first appears and therefore $V_{\max}$ also occurs for this value of p.

Random Walks and Resistor Networks

The Basic Relation

We now discuss how the voltages at each node in a resistor network and the resistance of the network are directly related to *first-passage* properties of random walks [31,63,76,98]. To develop this connection, consider a random walk on a finite network that can hop between nearest-neighbor sites i to j with probability p_{ij} in a single step. We divide the boundary points of the network into two disjoint classes, $\mathcal{B}_+$ and $\mathcal{B}_-$, that we are free to choose; a typical situation is the geometry shown in Fig. 9. We now ask: starting at an arbitrary point i, what is the probability that the walk *eventually* reaches the boundary set $\mathcal{B}_+$ without first reaching any node in $\mathcal{B}_-$? This quantity is termed the exit probability $\mathcal{E}_+(i)$ (with an analogous definition for the exit probability $\mathcal{E}_-(i) = 1 - \mathcal{E}_+(i)$ to $\mathcal{B}_-$).

We obtain the exit probability $\mathcal{E}_+(i)$ by summing the probabilities for all walk trajectories that start at i and reach a site in $\mathcal{B}_+$ without touching any site in $\mathcal{B}_-$ (and similarly for $\mathcal{E}_-(i)$). Thus

$$
\mathcal{E}_\pm(i) = \sum_{P_\pm} \mathcal{P}_{P_\pm}(i) \,, \quad (28)
$$

where $\mathcal{P}_{P_\pm}(i)$ denotes the probability of a path from i to $\mathcal{B}_\pm$ that avoids $\mathcal{B}_\mp$. The sum over all these restricted paths
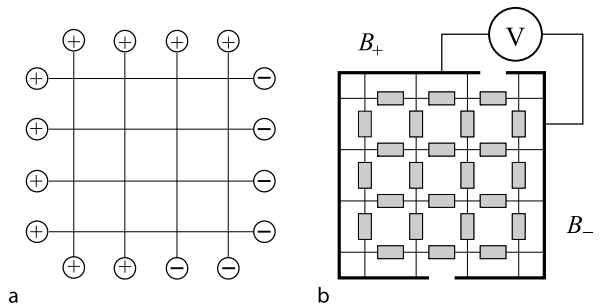
Fractal and Multifractal Scaling of Electrical Conduction in Random Resistor Networks, Figure 9
a Lattice network with boundary sites $\mathcal{B}_+$ or $\mathcal{B}_-$. **b** Corresponding resistor network in which each rectangle is a 1 ohm resistor. The sites in $\mathcal{B}_+$ are all fixed at potential $V = 1$, and sites in $\mathcal{B}_-$ are all grounded

can be decomposed into the outcome after one step, when the walk reaches some intermediate site j, and the sum over all path remainders from j to $\mathcal{B}_\pm$. This decomposition gives

$$\mathcal{E}_\pm(i) = \sum_j p_{ij}\, \mathcal{E}_\pm(j)\,. \tag{29}$$

Thus $\mathcal{E}_\pm(i)$ is a harmonic function because it equals a weighted average of $\mathcal{E}_\pm$ at neighboring points, with weighting function p_{ij}. This is exactly the same relation obeyed by the node voltages in Eq. (2) for the corresponding resistor network when we identify the single-step hopping probabilities p_{ij} with the conductances $g_{ij}/\sum_j g_{ij}$. We thus have the following equivalence:

- Let the boundary sets $\mathcal{B}_+$ and $\mathcal{B}_-$ in a resistor network be fixed at voltages 1 and 0 respectively, with g_{ij} the conductance of the bond between sites i and j. Then the voltage at any interior site i coincides with the probability for a random walk, which starts at i, to reach $\mathcal{B}_+$ before reaching $\mathcal{B}_-$, when the hopping probability from i to j is $p_{ij} = g_{ij}/\sum_j g_{ij}$.

If all the bond conductances are the same – corresponding to single-step hopping probabilities in the equivalent random walk being identical – then Eq. (29) is just the discrete Laplace equation. We can then exploit this correspondence between conductances and hopping probabilities to infer non-trivial results about random walks and about resistor networks from basic electrostatics. This correspondence can also be extended in a natural way to general random walks with a spatially-varying bias and diffusion coefficient, and to continuous media.

The consequences of this equivalence between random walks and resistor networks is profound. As an example [76], consider a diffusing particle that is initially at distance r_0 from the center of a sphere of radius $a < r_0$ in otherwise empty d-dimensional space. By the correspondence with electrostatics, the probability that this particle eventually hits the sphere is simply the electrostatic potential at r_0, $\mathcal{E}_-(r_0) = (a/r_0)^{d-2}$!

Network Resistance and Pólya's Theorem

An important extension of the relation between exit probability and node voltages is to infinite resistor networks. This extension provides a simple connection between the classic recurrence/transience transition of random walks on a given network [31,63,76,98] and the electrical resistance of this same network [26]. Consider a symmetric random walk on a regular lattice in d spatial dimensions.

Suppose that the walk starts at the origin at $t = 0$. What is the probability that the walk *eventually* returns to its starting point? The answer is strikingly simple:

- For $d \leq 2$, a random walk is *certain* to eventually return to the origin. This property is known as *recurrence*.
- For $d > 2$, there is a non-zero probability that the random walk will *never* return to the origin. This property is known as *transience*.

Let's now derive the transience and recurrence properties of random walks in terms of the equivalent resistor network problem. Suppose that the voltage V at the boundary sites $\mathcal{B}_+$ is set to one. Then by Kirchhoff's law, the total current entering the network is

$$I = \sum_j (1 - V_j)g_{+j} = \sum_j (1 - V_j)p_{+j}\sum_k g_{+k}\,. \tag{30}$$

Here g_{+j} is the conductance of the resistor between $\mathcal{B}_+$ and a neighboring site j, and $p_{+j} = g_{+j}/\sum_j g_{+j}$. Because the voltage V_j also equals the probability for the corresponding random walk to reach $\mathcal{B}_+$ without reaching $\mathcal{B}_-$, the term $V_j\, p_{+j}$ is just the probability that a random walk starts at $\mathcal{B}_+$, makes a single step to one of the sites j adjacent to $\mathcal{B}_+$ (with hopping probability p_{ij}), and then returns to $\mathcal{B}_+$ without reaching $\mathcal{B}_-$. We therefore deduce that

$$
\begin{aligned}
I &= \sum_j (1 - V_j)g_{+j} \\
&= \sum_k g_{+k} \sum_j (1 - V_j)p_{+j} \\
&= \sum_k g_{+k} \times (1 - \text{return probability}) \\
&= \sum_k g_{+k} \times \text{escape probability}\,. \tag{31}
\end{aligned}
$$

Here "escape" means that the random walk reaches the set $\mathcal{B}_-$ without returning to a node in $\mathcal{B}_+$.

On the other hand, the current and the voltage drop across the network are related to the conductance G between the two boundary sets by $I = GV = G$. From this fact, Eq. (31) gives the fundamental result

$$\text{escape probability} \equiv P_{\text{escape}} = \frac{G}{\sum_k g_{+k}}\,. \tag{32}$$

Suppose now that a current I is injected at a single point of an infinite network, with outflow at infinity (Fig. 10). Thus the probability for a random walk to never return to its
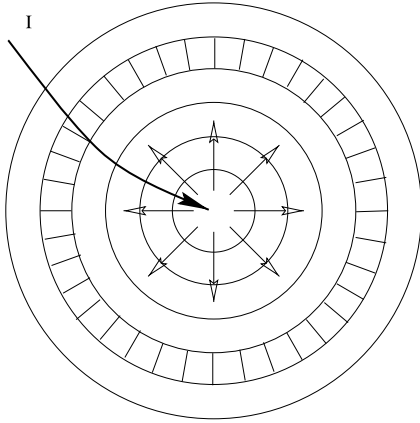
Fractal and Multifractal Scaling of Electrical Conduction in Random Resistor Networks, Figure 10
Decomposition of a conducting medium into concentric shells, each of which consists of fixed-conductance blocks. A current I is injected at the origin and flows radially outward through the medium

starting point, is simply proportional to the conductance G from this starting point to infinity of the same network. Thus a subtle feature of random walks, namely, the escape probability, is directly related to currents and voltages in an equivalent resistor network.

Part of the reason why this connection is so useful is that the conductance of the infinite network for various spatial dimensions can be easily determined, while a direct calculation of the return probability for a random walk is more difficult. In one dimension, the conductance of an infinitely long chain of identical resistors is clearly zero. Thus $P_{\text{escape}} = 0$ or, equivalently, $P_{\text{return}} = 1$. Thus a random walk in one dimension is recurrent. As alluded to at the outset of Sect. "Introduction to Current Flows", the conductance between one point and infinity in an infinite resistor lattice in general spatial dimension is somewhat challenging. However, to merely determine the recurrence or transience of a random walk, we only need to know if the return probability is zero or greater than zero. Such a simple question can be answered by a crude physical estimate of the network conductance.

To estimate the conductance from one point to infinity, we replace the discrete lattice by a continuum medium of constant conductance. We then estimate the conductance of the infinite medium by decomposing it into a series of concentric shells of fixed thickness dr. A shell at radius r can be regarded as a parallel array of r^{d-1} volume elements, each of which has a fixed conductance. The conductance of one such shell is proportional to its surface area, and the overall resistance is the sum of these shell re-

sistances. This reasoning gives

$$
\begin{aligned}
R &\sim \int^{\infty} R_{\text{shell}}(r) dr \\
&\sim \int^{\infty} \frac{dr}{r^{d-1}} = \begin{cases} \infty & \text{for } d \leq 2 \\ (P_{\text{escape}} \sum_j g_{+j})^{-1} & \text{for } d > 2 \, . \end{cases}
\end{aligned} \tag{33}
$$

The above estimate gives an easy solution to the recurrence/transience transition of random walks. For $d \leq 2$, the conductance to infinity is zero because there are an insufficient number of independent paths from the origin to infinity. Correspondingly, the escape probability is zero and the random walk is recurrent. The case $d = 2$ is more delicate because the integral in Eq. (33) diverges only logarithmically at the upper limit. Nevertheless, the conductance to infinity is still zero and the corresponding random walk is recurrent (but just barely). For $d > 2$, the conductance between a single point and infinity in an infinite homogeneous resistor network is non zero and therefore the escape probability of the corresponding random walk is also non zero – the walk is now transient.

There are many amusing ramifications of the recurrence of random walks and we mention two such properties. First, for $d \leq 2$, even though a random walk eventually returns to its starting point, the mean time for this event is infinite! This divergence stems from a power-law tail in the time dependence of the first-passage probability [31,76], namely, the probability that a random walk returns to the origin for the first time. Another striking aspect of recurrence is that because a random walk returns to its starting point with certainty, it necessarily returns an infinite number of times.

Future Directions

There is a good general understanding of the conductance of resistor networks, both far from the percolation threshold, where effective medium theory applies, and close to percolation, where the conductance G vanishes as $(p - p_c)^t$. Many advancements in numerical techniques have been developed to determine the conductance accurately and thereby obtain precise values for the conductance exponent, especially in two dimensions. In spite of this progress, we still do not yet have the right way, if it exists at all, to link the geometry of the percolation cluster or the conducting backbone to the conductivity itself. Furthermore, many exponents of two-dimensional percolation are known exactly. Is it possible that the exact approaches developed to determine percolation exponents can be extended to give the exact conductance exponent?

Finally, there are aspects about conduction in random networks that are worth highlighting. The first falls under the rubric of *directed percolation* [51]. Here each link in a network has an intrinsic directionality that allows current to flow in one direction only – a resistor and diode in series. Links are also globally oriented; on the square lattice for example, current can flow rightward and upward. A qualitative understanding of directed percolation and directed conduction has been achieved that parallels that of isotropic percolation. However, there is one facet of directed conduction that is barely explored. Namely, the state of the network (the bonds that are forward biased) must be determined self consistently from the current flows. This type of non-linearity is much more serious when the circuit elements are randomly oriented. These questions about the coupling between the state of the network and its conductance are central when the circuit elements are intrinsically non-linear [49,78]. This is a topic that seems ripe for new developments.

Bibliography

1. Adler J (1985) Conductance Exponents From the Analysis of Series Expansions for Random Resistor Networks. J Phys A Math Gen 18:307–314
2. Adler J, Meir Y, Aharony A, Harris AB, Klein L (1990) Low-Concentration Series in General Dimension. J Stat Phys 58:511–538
3. Aharony A, Feder J (eds) (1989) Fractals in Physics. Phys D 38:1–398
4. Alava MJ, Nukala PKVV, Zapperi S (2006) Statistical Models for Fracture. Adv Phys 55:349–476
5. Alexander S, Orbach R (1982) Density of States of Fractals: Fractons. J Phys Lett 43:L625–L631
6. Atkinson D, van Steenwijk FJ (1999) Infinite Resistive Lattice. Am J Phys 67:486–492
7. Batrouni GG, Hansen A, Nelkin M (1986) Fourier Acceleration of Relaxation Processes in Disordered Systems. Phys Rev Lett 57:1336–1339
8. Batrouni GG, Hansen A, Larson B (1996) Current Distribution in the Three-Dimensional Random Resistor Network at the Percolation Threshold. Phys Rev E 53:2292–2297
9. Blumenfeld R, Meir Y, Aharony A, Harris AB (1987) Resistance Fluctuations in Randomly Diluted Networks. Phys Rev B 35:3524–3535
10. Bruggeman DAG (1935) Berechnung verschiedener physikalischer Konstanten von heterogenen Substanzen. I. Dielektrizitätskonstanten und Leitfähigkeiten der Mischkörper aus isotropen Substanzen. Ann Phys (Leipzig) 24:636–679. [Engl Trans: Computation of Different Physical Constants of Heterogeneous Substances. I. Dielectric Constants and Conductivenesses of the Mixing Bodies from Isotropic Substances.]
11. Bunde A, Havlin S (eds) (1991) Fractals and Disordered Systems. Springer, Berlin
12. Byshkin MS, Turkin AA (2005) A new method for the calculation of the conductance of inhomogeneous systems. J Phys A Math Gen 38:5057–5067
13. Castellani C, Peliti L (1986) Multifractal Wavefunction at the Localisation Threshold. J Phys A Math Gen 19:L429–L432
14. Chan SK, Machta J, Guyer RA (1989) Large Currents in Random Resistor Networks. Phys Rev B 39:9236–9239
15. Coniglio A (1981) Thermal Phase Transition of the Dilute s-State Potts and n-Vector Models at the Percolation Threshold. Phys Rev Lett 46:250–253
16. Cserti J (2000) Application of the lattice Green's function of calculating the resistance of an infinite network of resistors. Am J Phys 68:896–906
17. de Arcangelis L, Redner S, Coniglio A (1985) Anomalous Voltage Distribution of Random Resistor Networks and a New Model for the Backbone at the Percolation Threshold. Phys Rev B 3:4725–4727
18. de Arcangelis L, Redner S, Herrmann HJ (1985) A Random Fuse Model for Breaking Processes. J Phys 46:L585–L590
19. de Arcangelis L, Redner S, Coniglio A (1986) Multiscaling Approach in Random Resistor and Random Superconducting Networks. Phys Rev B 34:4656–4673
20. de Gennes PG (1972) Exponents for the Excluded vol Problem as Derived by the Wilson Method. Phys Lett A 38:339–340
21. de Gennes PG (1976) La Notion de Percolation: Un Concept Unificateur. La Recherche 7:919–927
22. de Gennes PG (1976) On a Relation Between Percolation Theory and the Elasticity of Gels. J Phys Lett 37:L1–L3
23. den Nijs M (1979) A Relation Between the Temperature Exponents of the Eight-Vertex and q-state Potts Model. J Phys A Math Gen 12:1857–1868
24. Derrida B, Vannimenus J (1982) Transfer–Matrix Approach to Random Resistor Networks. J Phys A: Math Gen 15:L557–L564
25. Derrida B, Zabolitzky JG, Vannimenus J, Stauffer D (1984) A Transfer Matrix Program to Calculate the Conductance of Random Resistor Networks. J Stat Phys 36:31–42
26. Doyle PG, Snell JL (1984) Random Walks and Electric Networks. The Carus Mathematical Monograph, Series 22. The Mathematical Association of America, USA
27. Duxbury PM, Beale PD, Leath PL (1986) Size Effects of Electrical Breakdown in Quenched Random Media. Phys Rev Lett 57:1052–1055
28. Duxbury PM, Leath PL, Beale PD (1987) Breakdown Properties of Quenched Random Systems: The Random-Fuse Network. Phys Rev B 36:367–380
29. Dykhne AM (1970) Conductivity of a Two-Dimensional Two-Phase System. Zh Eksp Teor Fiz 59:110–115 [Engl Transl: (1971) Sov Phys-JETP 32:63–65]
30. Eggarter TP, Cohen MH (1970) Simple Model for Density of States and Mobility of an Electron in a Gas of Hard-Core Scatterers. Phys Rev Lett 25:807–810
31. Feller W (1968) An Introduction to Probability Theory and Its Applications, vol 1. Wiley, New York
32. Fisch R, Harris AB (1978) Critical Behavior of Random Resistor Networks Near the Percolation Threshold. Phys Rev B 18:416–420
33. Fogelholm R (1980) The Conductance of Large Percolation Network Samples. J Phys C 13:L571–L574
34. Fortuin CM, Kasteleyn PW (1972) On the Random Cluster Model. I. Introduction and Relation to Other Models. Phys 57:536–564
35. Frank DJ, Lobb CJ (1988) Highly Efficient Algorithm for Percolative Transport Studies in Two Dimensions. Phys Rev B 37:302–307

36. Gingold DB, Lobb CJ (1990) Percolative Conduction in Three Dimensions. Phys Rev B 42:8220–8224

37. Golden K (1989) Convexity in Random Resistor Networks. In: Kohn RV, Milton GW (eds) Random Media and Composites. SIAM, Philadelphia, pp 149–170

38. Golden K (1990) Convexity and Exponent Inequalities for Conduction Near Percolation. Phys Rev Lett 65:2923–2926

39. Gumbel EJ (1958) Statistics of Extremes. Columbia University Press, New York

40. Halperin BI, Feng S, Sen PN (1985) Differences Between Lattice and Continuum Percolation Transport Exponents. Phys Rev Lett 54:2391–2394

41. Halsey TC, Jensen MH, Kadanoff LP, Procaccia I, Shraiman BI (1986) Fractal Measures and Their Singularities: The Characterization of Strange Sets. Phys Rev A 33:1141–1151

42. Halsey TC, Meakin P, Procaccia I (1986) Scaling Structure of the Surface Layer of Diffusion-Limited Aggregates. Phys Rev Lett 56:854–857

43. Harary F (1969) Graph Theory. Addison Wesley, Reading, MA

44. Harris AB, Kim S, Lubensky TC (1984) ε Expansion for the Conductance of a Random Resistor Network. Phys Rev Lett 53:743–746

45. Hentschel HGE, Procaccia I (1983) The Infinite Number of Generalized Dimensions of Fractals and Strange Attractors. Phys D 8:435–444

46. Kahng B, Batrouni GG, Redner S (1987) Logarithmic Voltage Anomalies in Random Resistor Networks. J Phys A: Math Gen 20:L827–834

47. Kasteleyn PW, Fortuin CM (1969) Phase Transitions in Lattice Systems with Random Local Properties. J Phys Soc Japan (Suppl) 26:11–14

48. Keller JB (1964) A Theorem on the Conductance of a Composite Medium. J Math Phys 5:548–549

49. Kenkel SW, Straley JP (1982) Percolation Theory of Nonlinear Circuit Elements. Phys Rev Lett 49:767–770

50. Kennelly AE (1899) The Equivalence of Triangles and Three-Pointed Stars in Conducting Networks. Electr World Eng 34:413–414

51. Kinzel W (1983) Directed Percolation in Percolation Structures and Processes. In: Deutscher G, Zallen R, Adler J, Hilger A, Bristol UK, Redner S (eds) Annals of the Israel Physical Society, vol 5, Percolation and Conduction in Random Resistor-Diode Networks, ibid, pp 447–475

52. Kirchhoff G (1847) Über die Auflösung der Gleichungen, auf welche man bei der Untersuchung der linearen Verteilung galvanischer Ströme geführt wird. Ann Phys Chem 72:497–508. [English translation by O'Toole JB, Kirchhoff G (1958) On the solution of the equations obtained from the investigation of the linear distribution of galvanic currents. IRE Trans Circuit Theory CT5:4–8.]

53. Kirkpatrick S (1971) Classical Transport in Disordered Media: Scaling and Effective-Medium Theories. Phys Rev Lett 27:1722–1725

54. Kirkpatrick S (1973) Percolation and Conduction. Rev Mod Phys 45:574–588

55. Koplik J (1981) On the Effective Medium Theory of Random Linear Networks. J Phys C 14:4821–4837

56. Koplik J, Redner S, Wilkinson D (1988) Transport and Dispersion in Random Networks with Percolation Disorder. Phys Rev A 37:2619–2636

57. Landauer R (1952) The Electrical Resistance of Binary Metallic Mixtures. J Appl Phys 23:779–784

58. Last BL, Thouless DJ (1971) Percolation Theory and Electrical Conductance. Phys Rev Lett 27:1719–1721

59. Li PS, Strieder W (1982) Critical Exponents for Conduction in a Honeycomb Random Site Lattice. J Phys C 15:L1235–L1238; Also in: Li PS, Strieder W (1982) Monte Carlo Simulation of the Conductance of the Two-Dimensional Triangular Site Network. J Phys C 15:6591–6595

60. Li YS, Duxbury PM (1987) Size and Location of the Largest Current in a Random Resistor Network. Phys Rev B 36:5411–5419

61. Lobb CJ, Frank DJ (1979) A Large-Cell Renormalisation Group Calculation of the Percolation Conduction Critical Exponent. J Phys C 12:L827–L830

62. Lobb CJ, Frank DJ (1982) Percolative Conduction and the Alexander–Orbach Conjecture in Two Dimensions. Phys Rev B 30:4090–4092

63. Lovasz L (1993) Random Walks on Graphs: A Survey. In: Miklós D, Sós VT, Szönyi T (eds) Combinatorics, Paul Erdös is Eighty, vol 2. János Bolyai Mathematical Society, Budapest 2, pp 1–46

64. Ma SK (1976) Modern Theory of Critical Phenomena. WA Benjamin, Reading, MA

65. Machta J, Guyer RA (1987) Largest Current in a Random Resistor Network. Phys Rev B 36:2142–2146

66. Mandelbrot BB (1974) Intermittent Turbulence in Self-Similar Cascades: Divergence of High Moments and Dimension of the Carrier. J Fluid Mech 62:331–358

67. Mandelbrot BB (1982) The Fractal Geometry of Nature. WH Freeman, San Francisco

68. Mitescu CD, Allain A, Guyon E, Clerc J (1982) Electrical Conductance of Finite-Size Percolation Networks. J Phys A: Math Gen 15:2523–2532

69. Nevard J, Keller JB (1985) Reciprocal Relations for Effective Conductivities of Anisotropic Media. J Math Phys 26:2761–2765

70. Normand JM, Herrmann HJ, Hajjar M (1988) Precise Calculation of the Dynamical Exponent of Two-Dimensional Percolation. J Stat Phys 52:441–446

71. Press W, Teukolsky S, Vetterling W, Flannery B (1992) Numerical Recipes in Fortran 90. The Art of Parallel Scientific Computing. Cambridge University Press, New York

72. Rammal R, Tannous C, Breton P, Tremblay A-MS (1985) Flicker (1/f) Noise in Percolation Networks: A New Hierarchy of Exponents Phys Rev Lett 54:1718–1721

73. Rammal R, Tannous C, Tremblay A-MS (1985) 1/f Noise in Random Resistor Networks: Fractals and Percolating Systems. Phys Rev A 31:2662–2671

74. Rayleigh JW (1892) On the Influence of Obstacles Arranged in Rectangular Order upon the Properties of a Medium. Philos Mag 34:481–502

75. Redner S (1990) Random Multiplicative Processes: An Elementary Tutorial. Am J Phys 58:267–272

76. Redner S (2001) A Guide to First-Passage Processes. Cambridge University Press, New York

77. Redner S, Brooks JS (1982) Analog Experiments and Computer Simulations for Directed Conductance. J Phys A Math Gen 15:L605–L610

78. Roux S, Herrmann HJ (1987) Disorder-Induced Nonlinear Conductivity. Europhys Lett 4:1227–1231

79. Sahimi M, Hughes BD, Scriven LE, Davis HT (1983) Critical

Exponent of Percolation Conductance by Finite-Size Scaling. J Phys C 16:L521–L527

80. Sarychev AK, Vinogradoff AP (1981) Drop Model of Infinite Cluster for 2d Percolation. J Phys C 14:L487–L490
81. Senturia SB, Wedlock BD (1975) Electronic Circuits and Applications. Wiley, New York, pp 75
82. Skal AS, Shklovskii BI (1975) Topology of the Infinite Cluster of The Percolation Theory and its Relationship to the Theory of Hopping Conduction. Fiz Tekh Poluprov 8:1586–1589 [Engl. transl.: Sov Phys-Semicond 8:1029–1032]
83. Stanley HE (1971) Introduction to Phase Transition and Critical Phenomena. Oxford University Press, Oxford, UK
84. Stanley HE (1977) Cluster Shapes at the Percolation Threshold: An Effective Cluster Dimensionality and its Connection with Critical-Point Exponents. J Phys A Math Gen 10: L211–L220
85. Stauffer D, Aharony A (1994) Introduction to Percolation Theory, 2nd edn. Taylor & Francis, London, Bristol, PA
86. Stenull O, Janssen HK, Oerding K (1999) Critical Exponents for Diluted Resistor Networks. Phys Rev E 59:4919–4930
87. Stephen M (1978) Mean-Field Theory and Critical Exponents for a Random Resistor Network. Phys Rev B 17:4444–4453
88. Stephen MJ (1976) Percolation problems and the Potts model. Phys Lett A 56:149–150
89. Stinchcombe RB, Watson BP (1976) Renormalization Group Approach for Percolation Conductance. J Phys C 9:3221–3247
90. Straley JP (1977) Critical Exponents for the Conductance of Random Resistor Lattices. Phys Rev B 15:5733–5737
91. Straley JP (1982) Random Resistor Tree in an Applied Field. J Phys C 10:3009–3014
92. Straley JP (1982) Threshold Behaviour of Random Resistor Networks: A Synthesis of Theoretical Approaches. J Phys C 10:2333–2341
93. Trugman SA, Weinrib A (1985) Percolation with a Threshold at Zero: A New Universality Class. Phys Rev B 31:2974–2980
94. van der Pol B, Bremmer H (1955) Operational Calculus Based on the Two-Sided Laplace Integral. Cambridge University Press, Cambridge, UK
95. Venezian G (1994) On the resistance between two points on a grid. Am J Phys 62:1000–1004
96. Watson GN (1939) Three Triple Integrals. Oxford Ser 2. Quart J Math 10:266–276
97. Webman I, Jortner J, Cohen MH (1975) Numerical Simulation of Electrical Conductance in Microscopically Inhomogeneous Materials. Phys Rev B 11:2885–2892
98. Weiss GH (1994) Aspects and Applications of the Random Walk. Elsevier Science Publishing Co, New York
99. Wu FY (1982) The Potts Model. Rev Mod Phys 54:235–268
100. Zabolitsky JG (1982) Monte Carlo Evidence Against the Alexander–Orbach Conjecture for Percolation Conductance. Phys Rev B 30:4077–4079

Fractal and Multifractal Time Series

Jan W. Kantelhardt
Institute of Physics,
Martin-Luther-University Halle-Wittenberg,
Halle, Germany

Article Outline

Glossary

Time series One dimensional array of numbers (x_i), $i = 1, \ldots, N$, representing values of an observable x usually measured equidistant (or nearly equidistant) in time.

Complex system A system consisting of many non-linearly interacting components. It cannot be split into simpler sub-systems without tampering with the dynamical properties.

Scaling law A power law with a scaling exponent (e. g. α) describing the behavior of a quantity F (e. g., fluctuation, spectral power) as function of a scale parameter s (e. g., time scale, frequency) at least asymptotically: $F(s) \sim s^{\alpha}$. The power law should be valid for a large range of s values, e. g., at least for one order of magnitude.

Fractal system A system characterized by a scaling law with a fractal, i. e., non-integer exponent. Fractal systems are self-similar, i. e., a magnification of a small part is statistically equivalent to the whole.

Self-affine system Generalization of a fractal system, where different magnifications s and $s' = s^H$ have to be used for different directions in order to obtain a statistically equivalent magnification. The exponent H is called Hurst exponent. Self-affine time series and time series becoming self-affine upon integration are commonly denoted as fractal using a less strict terminology.

Multifractal system A system characterized by scaling laws with an infinite number of different fractal exponents. The scaling laws must be valid for the same range of the scale parameter.

Crossover Change point in a scaling law, where one scaling exponent applies for small scale parameters and another scaling exponent applies for large scale pa-

rameters. The center of the crossover is denoted by its characteristic scale parameter $s_\times$ in this article.

Persistence In a persistent time series, a large value is usually (i. e., with high statistical preference) followed by a large value and a small value is followed by a small value. A fractal scaling law holds at least for a limited range of scales.

Short-term correlations Correlations that decay sufficiently fast that they can be described by a characteristic correlation time scale; e. g., exponentially decaying correlations. A crossover to uncorrelated behavior is observed on larger scales.

Long-term correlations Correlations that decay sufficiently slow that a characteristic correlation time scale cannot be defined; e. g., power-law correlations with an exponent between 0 and 1. Power-law scaling is observed on large time scales and asymptotically. The term long-range correlations should be used if the data is not a time series.

Non-stationarities If the mean or the standard deviation of the data values change with time, the weak definition of stationarity is violated. The strong definition of stationarity requires that all moments remain constant, i. e., the distribution density of the values does not change with time. Non-stationarities like monotonous, periodic, or step-like trends are often caused by external effects. In a more general sense, changes in the dynamics of the system also represent non-stationarities.

Definition of the Subject

Data series generated by complex systems exhibit fluctuations on a wide range of time scales and/or broad distributions of the values. In both equilibrium and non-equilibrium situations, the natural fluctuations are often found to follow a scaling relation over several orders of magnitude. Such scaling laws allow for a characterization of the data and the generating complex system by fractal (or multifractal) scaling exponents, which can serve as characteristic fingerprints of the systems in comparisons with other systems and with models. Fractal scaling behavior has been observed, e. g., in many data series from experimental physics, geophysics, medicine, physiology, and even social sciences. Although the underlying causes of the observed fractal scaling are often not known in detail, the fractal or multifractal characterization can be used for generating surrogate (test) data, modeling the time series, and deriving predictions regarding extreme events or future behavior. The main application, however, is still the characterization of different states or phases of the complex

system based on the observed scaling behavior. For example, the health status and different physiological states of the human cardiovascular system are represented by the fractal scaling behavior of the time series of intervals between successive heartbeats, and the coarsening dynamics in metal alloys are represented by the fractal scaling of the time-dependent speckle intensities observed in coherent X-ray spectroscopy.

In order to observe fractal and multifractal scaling behavior in time series, several tools have been developed. Besides older techniques assuming stationary data, there are more recently established methods differentiating truly fractal dynamics from fake scaling behavior caused by non-stationarities in the data. In addition, short-term and long-term correlations have to be clearly distinguished to show fractal scaling behavior unambiguously. This article describes several methods originating from statistical physics and applied mathematics, which have been used for fractal and multifractal time series analysis in stationary and non-stationary data.

Introduction

The characterization and understanding of complex systems is a difficult task, since they cannot be split into simpler subsystems without tampering with the dynamical properties. One approach in studying such systems is the recording of long *time series* of several selected variables (observables), which reflect the state of the system in a dimensionally reduced representation. Some systems are characterized by periodic or nearly periodic behavior, which might be caused by oscillatory components or closed-loop regulation chains. However, in truly complex systems such periodic components are usually not limited to one or two characteristic frequencies or frequency bands. They rather extend over a wide spectrum, and fluctuations on many time scales as well as broad distributions of the values are found. Often no specific lower frequency limit – or, equivalently, upper characteristic time scale – can be observed. In these cases, the dynamics can be characterized by *scaling laws* which are valid over a wide (possibly even unlimited) range of time scales or frequencies; at least over orders of magnitude. Such dynamics are usually denoted as *fractal* or *multifractal,* depending on the question if they are characterized by one scaling exponent or by a multitude of scaling exponents.

The first scientist who applied fractal analysis to natural time series was Benoit B. Mandelbrot [1,2,3], who included early approaches by H.E. Hurst regarding hydrological systems [4,5]. For extensive introductions describing fractal scaling in complex systems, we refer to [6,7,8,

9,10,11,12,13]. In the last decade, fractal and multifractal scaling behavior has been reported in many natural time series generated by complex systems, including

- Geophysics time series (recordings of temperature, precipitation, water runoff, ozone levels, wind speed, seismic events, vegetational patterns, and climate dynamics),
- Medical and physiological time series (recordings of heartbeat, respiration, blood pressure, blood flow, nerve spike intervals, human gait, glucose levels, and gene expression data),
- DNA sequences (they are not actually *time* series),
- Astrophysical time series (X-ray light sources and sunspot numbers),
- Technical time series (internet traffic, highway traffic, and neutronic power from a reactor),
- Social time series (finance and economy, language characteristics, fatalities in conflicts), as well as
- Physics data (also going beyond *time* series), e. g., surface roughness, chaotic spectra of atoms, and photon correlation spectroscopy recordings.

If one finds that a complex system is characterized by fractal (or multifractal) dynamics with particular scaling exponents, this finding will help in obtaining predictions on the future behavior of the system and on its reaction to external perturbations or changes in the boundary conditions. Phase transitions in the regulation behavior of a complex system are often associated with changes in their fractal dynamics, allowing for a detection of such transitions (or the corresponding states) by fractal analysis. One example for a successful application of this approach is the human cardiovascular system, where the fractality of heartbeat interval time series was shown to reflect certain cardiac impairments as well as sleep stages [14,15]. In addition, one can test and iteratively improve models of the system until they reproduce the observed scaling behavior. One example for such an approach is climate modeling, where the models were shown to need input from volcanos and solar radiation in order to reproduce the long-term correlated (fractal) scaling behavior [16] previously found in observational temperature data [17].

Fractal (or multifractal) scaling behavior certainly cannot be assumed a priori, but has to be established. Hence, there is a need for refined analysis techniques, which help to differentiate truly fractal dynamics from fake scaling behavior caused, e. g., by non-stationarities in the data. If conventional statistical methods are applied for the analysis of time series representing the dynamics of a complex system [18,19], there are two major problems. (i) The number of data series and their durations (lengths) are usually very limited, making it difficult to extract significant information on the dynamics of the system in a reliable way. (ii) If the length of the data is extended using computer-based recording techniques or historical (proxy) data, non-stationarities in the signals tend to be superimposed upon the intrinsic fluctuation properties and measurement noise. Non-stationarities are caused by external or internal effects that lead to either continuous or sudden changes in the average values, standard deviations or regulation mechanism. They are a major problem for the characterization of the dynamics, in particular for finding the scaling properties of given data.

Fractal and Multifractal Time Series

Fractality, Self-Affinity, and Scaling

The topic of this article is the fractality (and/or multifractality) of time series. Since fractals and multifractals in general are discussed in many other articles of the encyclopedia, the concept is not thoroughly explained here. In particular, we refer to the articles ▶ Fractal Geometry, A Brief Introduction to and ▶ Fractals and Multifractals, Introduction to for the formalism describing fractal and multifractal structures, respectively.

In a strict sense, most time series are one dimensional, since the values of the considered observable are measured in homogeneous time intervals. Hence, unless there are missing values, the fractal dimension of the support is $D(0) = 1$. However, there are rare cases where most of the values of a time series are very small or even zero, causing a dimension $D(0) < 1$ of the support. In these cases, one has to be very careful in selecting appropriate analysis techniques, since many of the methods presented in this article are not accurate for such data; the wavelet transform modulus maxima technique (see Subsect. "Wavelet Transform Modulus Maxima (WTMM) Method") is the most advanced applicable method.

Even if the fractal dimension of support is one, the information dimension $D(1)$ and the correlation dimension $D(2)$ can be studied. As we will see in Subsect. "The Structure Function Approach and Singularity Spectra", $D(2)$ is in fact explicitly related to all exponents studied in monofractal time series analysis. However, usually a slightly different approach is employed based on the notion of self-affinity instead of (multi-) fractality. Here, one takes into account that the time axis and the axis of the measured values $x(t)$ are not equivalent. Hence, a rescaling of time t by a factor a may require rescaling of the series values $x(t)$ by a different factor a^H in order to obtain a statistically similar (i. e., self-similar) picture. In this case the

scaling relation

$$x(t) \rightarrow a^H x(at) \qquad (1)$$

holds for an arbitrary factor a, describing the data as self-affine (see, e. g., [6]). The *Hurst* exponent H (after the water engineer H.E. Hurst [4]) characterizes the type of self affinity. Figure 1a shows several examples of self-affine time series with different H. The trace of a random walk (Brownian motion, third line in Fig. 1a), for example, is characterized by $H = 0.5$, implying that the position axis must be rescaled by a factor of two if the time axis is rescaled by a factor of four. Note that self-affine series are often denoted as fractal even though they are not fractal in the strict sense. In this article the term "fractal" will be used in the more general sense including all data, where a Hurst exponent H can be reasonably defined.

The scaling behavior of self-affine data can also be characterized by looking at their mean-square displacement. Since the mean-square displacement of a random walker is known to increase linearly in time, $\langle x^2(t) \rangle \sim t$, deviations from this law will indicate the presence of self-affine scaling. As we will see in Subsect. "Fluctuation Analysis (FA)", one can thus retrieve the Hurst (or self-affinity) exponent H by studying the scaling behavior of the mean-square displacement, or the mean-square fluctuations $\langle x^2(t) \rangle \sim t^{2H}$.

Persistence, Long- and Short-Term Correlations

Self-affine data are persistent in the sense that a large value is usually (i. e., with high statistical preference) followed by a large value and a small value is followed by a small value. For the trace of a random walk, persistence on all time scales is trivial, since a later position is just a former one plus some random increment(s). The persistence holds for all time scales, where the self-affinity relation (1) holds. However, the degree of persistence can also vary on different time scales. Weather is a typical example: while the weather tomorrow or in one week is probably similar to the weather today (due to a stable general weather condition), persistence is much harder to be seen on longer time scales.

Considering the increments $\Delta x_i = x_i - x_{i-1}$ of a self-affine series, (x_i), $i = 1, \ldots, N$ with N values measured equidistant in time, one finds that the Δx_i can be either persistent, independent, or anti-persistent. Examples for all cases are shown in Fig. 1b. In our example of the random walk with $H = 0.5$ (third line in the figure), the increments (steps) are fully independent of each other. Persistent and anti-persistent increments, where a positive increment is likely to be followed by another positive or negative increment, respectively, are also leading to persistent integrated series $x_i = \sum_{j=1}^{i} \Delta x_j$.

For stationary data with constant mean and standard deviation the auto-covariance function of the increments,

$$C(s) = \langle \Delta x_i \Delta x_{i+s} \rangle = \frac{1}{N-s} \sum_{i=1}^{N-s} \Delta x_i \Delta x_{i+s} \qquad (2)$$

can be studied to determine the degree of persistence. If $C(s)$ is divided by the variance $\langle (\Delta x_i)^2 \rangle$, it becomes the auto-correlation function; both are identical if the data are
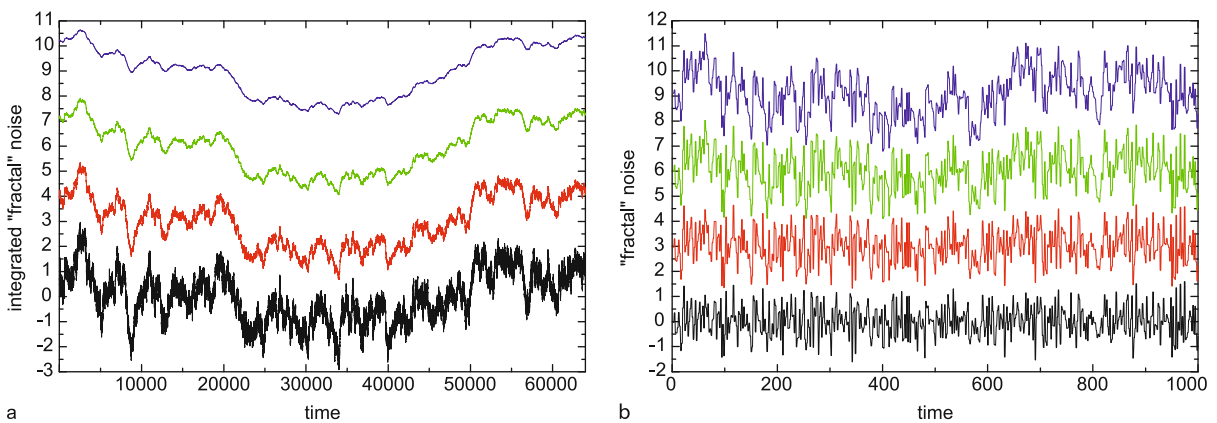
Fractal and Multifractal Time Series, Figure 1
a Examples of self-affine series x_i characterized by different Hurst exponents H = 0.9, 0.7, 0.5, 0.3 (from *top* to *bottom*). The data has been generated by Fourier filtering using the same seed for the random number generator. **b** Differentiated series Δx_i of the data from **a**; the Δx_i are characterized by positive long-term correlations (persistence) with γ = 0.2 and 0.6 (*first* and *second line*), uncorrelated behavior (*third line*), and anti-correlations (*bottom line*), respectively
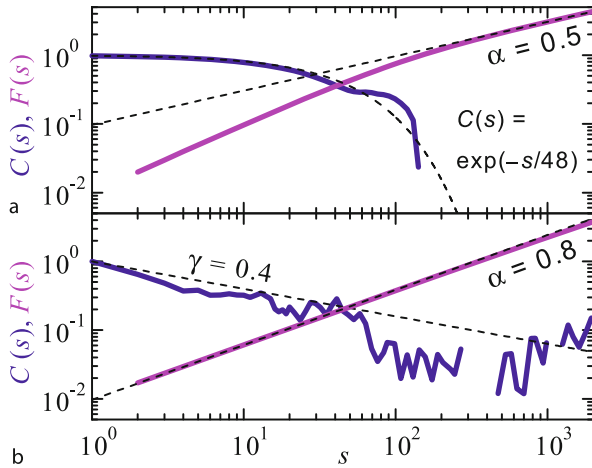
Fractal and Multifractal Time Series, Figure 2
Comparison of the autocorrelation functions $C(s)$ (decreasing functions) and fluctuation functions $F_2(s)$ (increasing functions) for short-term correlated data (*top panel*) and long-term correlated data ($\gamma = 0.4$, *bottom panel*). The asymptotic slope $H \approx \alpha = 0.5$ of $F_2(s)$ clearly indicates missing long-term correlations, while $H \approx \alpha = 1 - \gamma/2 > 0.5$ indicates long-term correlations. The difference is much harder to observe in $C(s)$, where statistical fluctuations and negative values start occurring above $s \approx 100$. The data have been generated by an AR process Eq. (4) and Fourier filtering, respectively. The *dashed lines* indicate the theoretical curves

normalized with unit variance. If the Δx_i are uncorrelated (as for the random walk), $C(s)$ is zero for $s > 0$. Short-range correlations of the increments Δx_i are usually described by $C(s)$ declining exponentially,

$$C(s) \sim \exp(-s/t_\times) \tag{3}$$

with a characteristic decay time $t_\times$. Such behavior is typical for increments generated by an auto-regressive (AR) process

$$\Delta x_i = c\Delta x_{i-1} + \varepsilon_i \tag{4}$$

with random uncorrelated offsets ε_i and $c = \exp(-1/t_\times)$. Figure 2a shows the auto-correlation function for one configuration of an AR process with $t_\times = 48$.

For so-called *long-range correlations* $\int_0^\infty C(s)\mathrm{d}s$ diverges in the limit of infinitely long series ($N \to \infty$). In practice, this means that $t_\times$ cannot be defined because it increases with increasing N. For example, $C(s)$ declines as a power-law

$$C(s) \propto s^{-\gamma} \tag{5}$$

with an exponent $0 < \gamma < 1$. Figure 2b shows $C(s)$ for one configuration with $\gamma = 0.4$. This type of behavior can

be modeled by the Fourier filtering technique (see Subsect. "Fourier Filtering"). Long-term correlated, i. e. persistent, behavior of the Δx_i leads to self-affine scaling behavior of the x_i, characterized by $H = 1 - \gamma/2$, as will be shown below.

Crossovers and Non-Stationarities in Time Series

Short-term correlated increments Δx_i characterized by a finite characteristic correlation decay time $t_\times$ lead to a crossover in the scaling behavior of the integrated series $x_i = \sum_{j=1}^i \Delta x_j$, see Fig. 2a for an example. Since the position of the crossover might be numerically different from $t_\times$, we denote it by $s_\times$ here. Time series with a crossover are not self-affine and there is no unique Hurst exponent H characterizing them. While $H > 0.5$ is observed on small time scales (indicating correlations in the increments), the asymptotic behavior (for large time scales $s \gg t_\times$ and $\gg s_\times$) is always characterized by $H = 0.5$, since all correlations have decayed. Many natural recordings are characterized by pronounced short-term correlations in addition to scaling long-term correlations. For example, there are short-term correlations due to particular general weather situations in temperature data and due to respirational effects in heartbeat data. Crossovers in the scaling behavior of complex time series can also be caused by different regulation mechanisms on fast and slow time scales. Fluctuations of river runoff, for example, show different scaling behavior on time scales below and above approximately one year.

Non-stationarities can also cause crossovers in the scaling behavior of data if they are not properly taken into account. In the most strict sense, non-stationarities are variations in the mean or the standard deviation of the data (violating weak stationarity) or the distribution of the data values (violating strong stationarity). Non-stationarities like monotonous, periodic or step-like trends are often caused by external effects, e.g., by the greenhouse warming and seasonal variations for temperature records, different levels of activity in long-term physiological data, or unstable light sources in photon correlation spectroscopy. Another example for non-stationary data is a record consisting of segments with strong fluctuations alternating with segments with weak fluctuations. Such behavior will cause a crossover in scaling at the time scale corresponding to the typical duration of the homogeneous segments. Different mechanisms of regulation during different time segments – like, e. g., different heartbeat regulation during different sleep stages at night – can also cause crossovers; they are regarded as non-stationarities here, too. Hence, if crossovers in the scaling behavior of

data are observed, more detailed studies are needed to find out the cause of the crossovers. One can try to obtain homogenous data by splitting the original series and employing methods that are at least insensitive to monotonous (polynomially shaped) trends.

To characterize a complex system based on time series, trends and fluctuations are usually studied separately (see, e. g., [20] for a discussion). Strong trends in data can lead to a false detection of long-range statistical persistence if only one (non-detrending) method is used or if the results are not carefully interpreted. Using several advanced techniques of scaling time series analysis (as described in Sect. "Methods for Non-stationary Fractal Time Series Analysis") crossovers due to trends can be distinguished from crossovers due to different regulation mechanisms on fast and slow time scales. The techniques can thus assist in gaining insight into the scaling behavior of the natural variability as well as into the kind of trends of the considered time series.

It has to be stressed that crossovers in scaling behavior must not be confused with multifractality. Even though several scaling exponents are needed, they are not applicable for the same regime (i. e., the same range of time scales). Real multifractality, on the other hand, is characterized by different scaling behavior of different moments over the full range of time scales (see next section).

Multifractal Time Series

Many records do not exhibit a simple monofractal scaling behavior, which can be accounted for by a single scaling exponent. As discussed in the previous section, there might exist crossover (time-) scales $s_\times$ separating regimes with different scaling exponents. In other cases, the scaling behavior is more complicated, and different scaling exponents are required for different parts of the series. In even more complicated cases, such different scaling behavior can be observed for many interwoven fractal subsets of the time series. In this case a multitude of scaling exponents is required for a full description of the scaling behavior in the same range of time scales, and a multifractal analysis must be applied.

Two general types of multifractality in time series can be distinguished: (i) Multifractality due to a broad probability distribution (density function) for the values of the time series, e. g. a Levy distribution. In this case the multifractality cannot be removed by shuffling the series. (ii) Multifractality due to different long-term correlations of the small and large fluctuations. In this case the probability density function of the values can be a regular distribution with finite moments, e. g., a Gaussian distribu-

tion. The corresponding shuffled series will exhibit non-multifractal scaling, since all long-range correlations are destroyed by the shuffling procedure. Randomly shuffling the order of the values in the time series is the easiest way of generating surrogate data; however, there are more advanced alternatives (see Sect. "Simple Models for Fractal and Multifractal Time Series"). If both kinds of multifractality are present, the shuffled series will show weaker multifractality than the original series.

A multifractal analysis of time series will also reveal higher order correlations. Multifractal scaling can be observed if, e. g., three or four-point correlations scale differently from the standard two-point correlations studied by classical autocorrelation analysis (Eq. (2)). In addition, multifractal scaling is observed if the scaling behavior of small and large fluctuations is different. For example, extreme events might be more or less correlated than typical events.

Methods for Stationary Fractal Time Series Analysis

In this section we describe four traditional approaches for the fractal analysis of stationary time series, see [21,22,23] for comparative studies. The main focus is on the determination of the scaling exponents H or γ, defined in Eqs. (1) and (5), respectively, and linked by $H = 1 - \gamma/2$ in long-term persistent data. Methods taking non-stationarities into account will be discussed in the next chapter.

Autocorrelation Function Analysis

We consider a record (x_i) of $i = 1, \ldots, N$ equidistant measurements. In most applications, the index i will correspond to the time of the measurements. We are interested in the correlation of the values x_i and x_{i+s} for different time lags, i. e. correlations over different time scales s. In order to remove a constant offset in the data, the mean $\langle x \rangle = \frac{1}{N} \sum_{i=1}^{N} x_i$ is usually subtracted, $\tilde{x}_i \equiv x_i - \langle x \rangle$. Alternatively, the correlation properties of increments $\tilde{x}_i = \Delta x_i = x_i - x_{i-1}$ of the original series can be studied (see also Subsect. "Persistence, Long- and Short-Term Correlations"). Quantitatively, correlations between $\tilde{x}$-values separated by s steps are defined by the (auto-) covariance function $C(s) = \langle \tilde{x}_i \tilde{x}_{i+s} \rangle$ or the (auto-) correlation function $C(s)/\langle \tilde{x}_i^2 \rangle$, see also Eq. (2).

As already mentioned in Subsect. "Persistence, Long- and Short-Term Correlations", the $\tilde{x}_i$ are short-term correlated if $C(s)$ declines exponentially, $C(s) \sim \exp(-s/t_\times)$, and long-term correlated if $C(s)$ declines as a power-law $C(s) \propto s^{-\gamma}$ with a correlation exponent $0 < \gamma < 1$ (see Eqs. (3) and (5), respectively). As illustrated by the two examples shown in Fig. 2, a direct calculation of $C(s)$ is

usually not appropriate due to noise superimposed on the data $\tilde{x}_i$ and due to underlying non-stationarities of unknown origin. Non-stationarities make the definition of $C(s)$ problematic, because the average $\langle x \rangle$ is not well-defined. Furthermore, $C(s)$ strongly fluctuates around zero on large scales s (see Fig. 2b), making it impossible to find the correct correlation exponent γ. Thus, one has to determine the value of γ indirectly.

Spectral Analysis

If the time series is stationary, we can apply standard spectral analysis techniques (Fourier transform) and calculate the power spectrum $S(f)$ of the time series $(\tilde{x}_i)$ as a function of the frequency f to determine self-affine scaling behavior [24]. For long-term correlated data characterized by the correlation exponent γ, we have

$$S(f) \sim f^{-\beta} \quad \text{with } \beta = 1 - \gamma . \tag{6}$$

The spectral exponent β and the correlation exponent γ can thus be obtained by fitting a power-law to a double logarithmic plot of the power spectrum $S(f)$. An example is shown in Fig. 3. The relation (6) can be derived from the Wiener–Khinchin theorem (see, e. g., [25]). If, instead of $\tilde{x}_i = \Delta x_i$ the integrated runoff time series is Fourier transformed, i. e., $\tilde{x}_i = x_i$ $X = \sum_{j=1}^{i} \Delta x_j$, the resulting power spectrum scales as $S(f) \sim f^{-2-\beta}$.

Spectral analysis, however, does not yield more reliable results than auto-correlation analysis unless a logarithmic binning procedure is applied to the double loga-
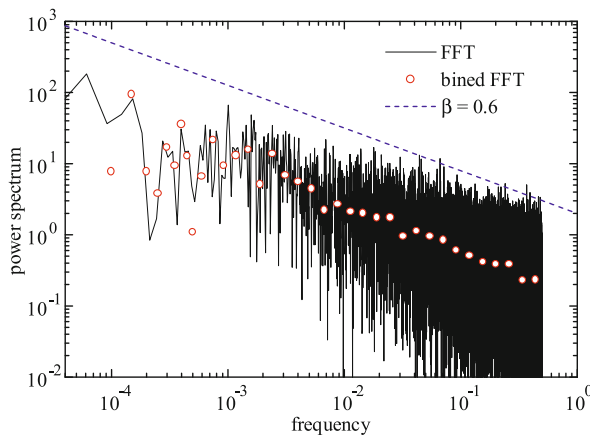
Fractal and Multifractal Time Series, Figure 3
Spectral analysis of a fractal time series characterized by long-term correlations with $\gamma = 0.4$ ($\beta = 0.6$). The expected scaling behavior (*dashed line* indicating the slope $-\beta$) is observed only after binning of the spectrum (*circles*). The data has been generated by Fourier filtering

rithmic plot of $S(f)$ [21], see also Fig. 3. I. e., the average of $\log S(f)$ is calculated in successive, logarithmically wide bands from $a^n f_0$ to $a^{n+1} f_0$, where f_0 is the minimum frequency, $a > 1$ is a factor (e. g., $a = 1.1$), and the index n is counting the bins. Spectral analysis also requires stationarity of the data.

Hurst's Rescaled-Range Analysis

The first method for the analysis of long-term persistence in time series based on random walk theory has been proposed by the water construction engineer Harold Edwin Hurst (1880–1978), who developed it while working in Egypt. His so-called rescaled range analysis (R/S analysis) [1,2,4,5,6] begins with splitting of the time series $(\tilde{x}_i)$ into non-overlapping segments ν of size (time scale) s (first step), yielding $N_s = \text{int}(N/s)$ segments altogether. In the second step, the *profile* (integrated data) is calculated in each segment $\nu = 0, \ldots, N_s - 1$,

$$Y_\nu(j) = \sum_{i=1}^{j} (\tilde{x}_{\nu s + i} - \langle \tilde{x}_{\nu s + i} \rangle_s)$$

$$= \sum_{i=1}^{j} \tilde{x}_{\nu s + i} - \frac{j}{s} \sum_{i=1}^{s} \tilde{x}_{\nu s + i} . \tag{7}$$

By the subtraction of the local averages, piecewise constant trends in the data are eliminated. In the third step, the differences between minimum and maximum value (*ranges*) $R_\nu(s)$ and the standard deviations $S_\nu(s)$ in each segment are calculated,

$$R_\nu(s) = \max_{j=1}^{s} Y_\nu(j) - \min_{j=1}^{s} Y_\nu(j),$$

$$S_\nu(s) = \sqrt{\frac{1}{s} \sum_{j=1}^{s} Y_\nu^2(j)} . \tag{8}$$

Finally, the rescaled range is averaged over all segments to obtain the fluctuation function $F(s)$,

$$F_{RS}(s) = \frac{1}{N_s} \sum_{\nu=0}^{N_s-1} \frac{R_\nu(s)}{S_\nu(s)} \sim s^H \quad \text{for } s \gg 1 , \tag{9}$$

where H is the Hurst exponent already introduced in Eq. (1). One can show [1,24] that H is related to β and γ by $2H \approx 1 + \beta = 2 - \gamma$ (see also Eqs. (6) and (14)). Note that $0 < \gamma < 1$, so that the right part of the equation does not hold unless $0.5 < H < 1$. The relationship does *not* hold in general for multifractal data. Note also that H actually characterizes the self-affinity of the profile function (7), while β and γ refer to the original data.

The values of H, that can be obtained by Hurst's rescaled range analysis, are limited to $0 < H < 2$, and significant inaccuracies are to be expected close to the bounds. Since H can be increased or decreased by one if the data is integrated ($\tilde{x}_j \to \sum_{i=1}^{j} \tilde{x}_i$) or differentiated ($\tilde{x}_i \to \tilde{x}_i - \tilde{x}_{i-1}$), respectively, one can always find a way to calculate H by rescaled range analysis provided the data is stationary. While values $H < 1/2$ indicate long-term anti-correlated behavior of the data $\tilde{x}_i$, $H > 1/2$ indicates long-term positively correlated behavior. For power-law correlations decaying faster than $1/s$, we have $H = 1/2$ for large s values, like for uncorrelated data.

Compared with spectral analysis, Hurst's rescaled range analysis yields smoother curves with less effort (no binning procedure is necessary) and works also for data with piecewise constant trends.

Fluctuation Analysis (FA)

The standard fluctuation analysis (FA) [8,26] is also based on random walk theory. For a time series $(\tilde{x}_i)$, $i = 1, \ldots, N$, with zero mean, we consider the global profile, i. e., the cumulative sum (cf. Eq. (7))

$$Y(j) = \sum_{i=1}^{j} \tilde{x}_i, \quad j = 0, 1, 2, \ldots, N, \tag{10}$$

and study how the fluctuations of the profile, in a given time window of size s, increase with s. The procedure is illustrated in Fig. 4 for two values of s. We can consider the profile $Y(j)$ as the position of a random walker on a linear chain after j steps. The random walker starts at the origin and performs, in the ith step, a jump of length $\tilde{x}_i$ to the bottom, if $\tilde{x}_i$ is positive, and to the top, if $\tilde{x}_i$ is negative.

To find how the square-fluctuations of the profile scale with s, we first divide the record of N elements into $N_s = \text{int}(N/s)$ non-overlapping segments of size s starting from the beginning (see Fig. 4) and another N_s non-overlapping segments of size s starting from the end of the considered series. This way neither data at the end nor at the beginning of the record is neglected. Then we determine the fluctuations in each segment v.

In the standard FA, we obtain the fluctuations just from the values of the profile at both endpoints of each segment $v = 1, \ldots, N_s$,

$$F_{FA}^2(v, s) = [Y(vs) - Y((v1)s)]^2, \tag{11}$$

(see Fig. 4) and analogous for $v = N_{s+1}, \ldots, 2N_s$,

$$F_{FA}^2(v, s) = [Y(N-(v-N_s)s)-Y(N-(v-1-N_s)s)]^2. \tag{12}$$
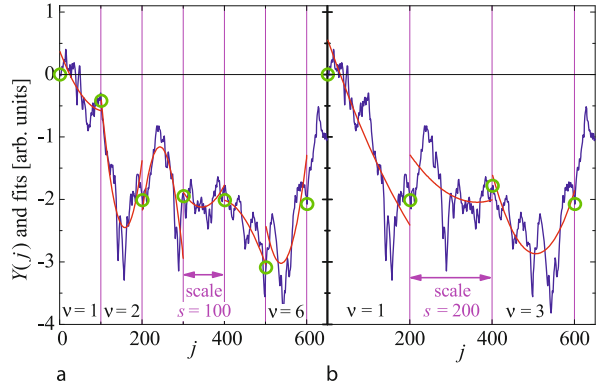
Fractal and Multifractal Time Series, Figure 4
Illustration of the fluctuation analysis (FA) and the detrended fluctuation analysis (DFA). For two segment durations (time scales) $s = 100$ (a) and 200 (b), the profiles $Y(j)$ (*blue lines*; defined in Eq. (11)), the values used for fluctuation analysis in Eq. (12) (*green circles*), and least-square quadratic fits to the profiles (*red lines*) are shown

Then we average $F_{FA}^2(v, s)$ over all subsequences to obtain the mean fluctuation $F_2(s)$,

$$F_2(s) = \left[\frac{1}{2N_s} \sum_{v=1}^{2N_s} F_{FA}^2(v, s) \right]^{1/2} \sim s^\alpha. \tag{13}$$

By definition, $F_2(s)$ can be viewed as the root-mean-square displacement of the random walker on the chain, after s steps (the reason for the index 2 will become clear later). For uncorrelated x_i values, we obtain Fick's diffusion law $F_2(s) \sim s^{1/2}$. For the relevant case of long-term correlations, in which $C(s)$ follows the power-law behavior of Eq. (5), $F_2(s)$ increases by a power law,

$$F_2(s) \sim s^\alpha \quad \text{with } \alpha \approx H, \tag{14}$$

where the fluctuation exponent α is identical with the Hurst exponent H for mono-fractal data and related to γ and β by

$$2\alpha = 1 + \beta = 2 - \gamma. \tag{15}$$

The typical behavior of $F_2(s)$ for short-term correlated and long-term correlated data is illustrated in Fig. 2. The relation (15) can be derived straightforwardly by inserting Eqs. (10), (2), and (5) into Eq. (11) and separating sums over products $\tilde{x}_i \tilde{x}_j$ with identical and different i and j, respectively.

The range of the α values that can be studied by standard FA is limited to $0 < \alpha < 1$, again with significant inaccuracies close to the bounds. Regarding integration or

differentiation of the data, the same rules apply as listed for H in the previous subsection. The results of FA become statistically unreliable for scales s larger than one tenth of the length of the data, i. e. the analysis should be limited by $s < N/10$.

Methods for Non-stationary Fractal Time Series Analysis

Wavelet Analysis

The origins of wavelet analysis come from signal theory, where frequency decompositions of time series were studied [27,28]. Like the Fourier transform, the wavelet transform of a signal $x(t)$ is a convolution integral to be replaced by a summation in case of a discrete time series $(\tilde{x}_i)$, $i = 1, \ldots, N$,

$$L_\psi(\tau, s) = \frac{1}{s} \int_{-\infty}^{\infty} x(t)\psi[(t - \tau)/s] \, dt$$

$$= \frac{1}{s} \sum_{i=1}^{N} \tilde{x}_i \, \psi[(i - \tau)/s] . \qquad (16)$$

Here, $\psi(t)$ is a so-called mother wavelet, from which all daughter wavelets $\psi_{\tau,s}(t) = \psi((t - \tau)/s)$ evolve by shifting and stretching of the time axis. The wavelet coefficients $L_\psi(\tau, s)$ thus depend on both time position τ and scale s. Hence, the local frequency decomposition of the signal is described with a time resolution appropriate for the considered frequency $f = 1/s$ (i. e., inverse time scale).

All wavelets $\psi(t)$ must have zero mean. They are often chosen to be orthogonal to polynomial trends, so that the analysis method becomes insensitive to possible trends in the data. Simple examples are derivatives of a Gaussian, $\psi_{\mathrm{Gauss}}^{(n)}(t) = \frac{d^n}{dt^n} \exp(-x^2/2)$, like the Mexican hat wavelet $-\psi_{\mathrm{Gauss}}^{(2)}$ and the Haar wavelet, $\psi_{\mathrm{Haar}}^{(1)}(t) = +1$ if $0 \le t < 1$, -1 if $1 \le t < 2$, and 0 otherwise. It is straightforward to construct higher order Haar wavelets that are orthogonal to linear, quadratic and cubic trends, e. g., $\psi_{\mathrm{Haar}}^{(2)}(t) = 1$ for $t \in [0, 1) \cup [2, 3)$, -2 for $t \in [1, 2)$, and 0 otherwise, or $\psi_{\mathrm{Haar}}^{(3)}(t) = 1$ for $t \in [0, 1)$, -3 for $t \in [1, 2)$, $+3$ for $t \in [2, 3)$, -1 for $t \in [3, 4)$, and 0 otherwise.

Discrete Wavelet Transform (WT) Approach

A detrending fractal analysis of time series can be easily implemented by considering Haar wavelet coefficients of the profile $Y(j)$, Eq. (10) [17,30]. In this case the convolution (16) corresponds to the addition and subtraction of mean values of $Y(j)$ within segments of size s. Hence, defining $\bar{Y}_\nu(s) = \frac{1}{s} \sum_{j=1}^{s} Y(\nu s + j)$, the coefficients can

be written as

$$F_{\mathrm{WT1}}(\nu, s) \equiv L_{\psi_{\mathrm{Haar}}^{(0)}}(\nu s, s) = \bar{Y}_\nu(s) - \bar{Y}_{\nu+1}(s) , \qquad (17)$$

$$F_{\mathrm{WT2}}(\nu, s) \equiv L_{\psi_{\mathrm{Haar}}^{(1)}}(\nu s, s)$$
$$= \bar{Y}_\nu(s) - 2\bar{Y}_{\nu+1}(s) + \bar{Y}_{\nu+2}(s) , \qquad (18)$$

and

$$F_{\mathrm{WT3}}(\nu, s) \equiv L_{\psi_{\mathrm{Haar}}^{(2)}}(\nu s, s)$$
$$= \bar{Y}_\nu(s) - 3\bar{Y}_{\nu+1}(s) + 3\bar{Y}_{\nu+2}(s) - \bar{Y}_{\nu+3}(s) \qquad (19)$$

for constant, linear and quadratic detrending, respectively. The generalization for higher orders of detrending is obvious. The resulting mean-square fluctuations $F_{\mathrm{WT}n}^2(\nu, s)$ are averaged over all ν to obtain the mean fluctuation $F_2(s)$, see Eq. (13). Figure 5 shows typical results for WT analysis of long-term correlated, short-term correlated and uncorrelated data.

Regarding trend-elimination, wavelet transform WT0 corresponds to standard FA (see Subsect. "Fluctuation Analysis (FA)"), and only constant trends in the profile are eliminated. WT1 is similar to Hurst's rescaled range analysis (see Subsect. "Hurst's Rescaled-Range Analysis"): linear trends in the profile and constant trends in the data are eliminated, and the range of the fluctuation exponent
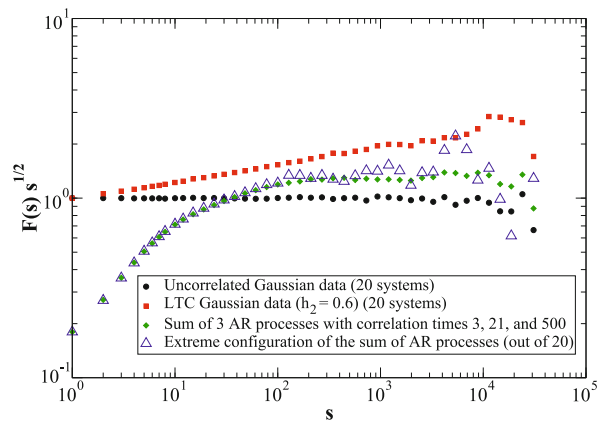
Fractal and Multifractal Time Series, Figure 5
Application of discrete wavelet transform (WT) analysis on uncorrelated data (*black circles*), long-term correlated data ($\gamma = 0.8$, $\alpha = 0.6$, *red squares*), and short-term correlated data (summation of three AR processes, *green diamonds*). Averages of $F_2(s)$ averaged over 20 series with $N = 2^{16}$ points and divided by $s^{1/2}$ are shown, so that a horizontal line corresponds to uncorrelated behavior. The *blue open triangles* show the result for one selected extreme configuration, where it is hard to decide about the existence of long-term correlations (figure after [29])

$\alpha \approx H$ is up to 2. In general, WTn determines the fluctuations from the nth derivative, this way eliminating trends described by $(n-1)$st-order polynomials in the data. The results become statistically unreliable for scales s larger than one tenth of the length of the data, just as for FA.

Detrended Fluctuation Analysis (DFA)

In the last 14 years *Detrended Fluctuation Analysis* (DFA), originally introduced by Peng et al. [31], has been established as an important method to reliably detect long-range (auto-) correlations in non-stationary time series. The method is based on random walk theory and basically represents a linear detrending version of FA (see Subsect. "Fluctuation Analysis (FA)"). DFA was later generalized for higher order detrending [15], separate analysis of sign and magnitude series [32] (see Subsect. "Sign and Magnitude (Volatility) DFA"), multifractal analysis [33] (see Subsect. "Multifractal Detrended Fluctuation Analysis (MFDFA)"), and data with more than one dimension [34]. Its features have been studied in many articles [35,36,37,38,39,40]. In addition, several comparisons of DFA with other methods for stationary and non-stationary time-series analysis have been published, see, e. g., [21,23,41,42] and in particular [22], where DFA is compared with many other established methods for short data sets, and [43], where it is compared with recently suggested improved methods. Altogether, there are about 600 papers applying DFA (till September 2008). In most cases positive auto-correlations were reported leaving only a few exceptions with anti-correlations, see, e. g., [44,45,46].

Like in the FA method, one first calculates the global profile according to Eq. (10) and divides the profile into $N_s = \mathrm{int}(N/s)$ non-overlapping segments of size s starting from the beginning and another N_s segments starting from the end of the considered series. DFA explicitly deals with monotonous trends in a detrending procedure. This is done by estimating a polynomial trend $y^m_{\nu,s}(j)$ within each segment ν by least-square fitting and subtracting this trend from the original profile ('detrending'),

$$\tilde{Y}_s(j) = Y(j) - y^m_{\nu,s}(j) . \tag{20}$$

The degree of the polynomial can be varied in order to eliminate constant ($m = 0$), linear ($m = 1$), quadratic ($m = 2$) or higher order trends of the profile function [15]. Conventionally the DFA is named after the order of the fitting polynomial (DFA0, DFA1, DFA2, …). In DFAm, trends of order m in the profile $Y(j)$ and of order $m-1$ in the original record $\tilde{x}_i$ are eliminated. The variance of the detrended profile $\tilde{Y}_s(j)$ in each segment ν yields the mean-square fluctuations,
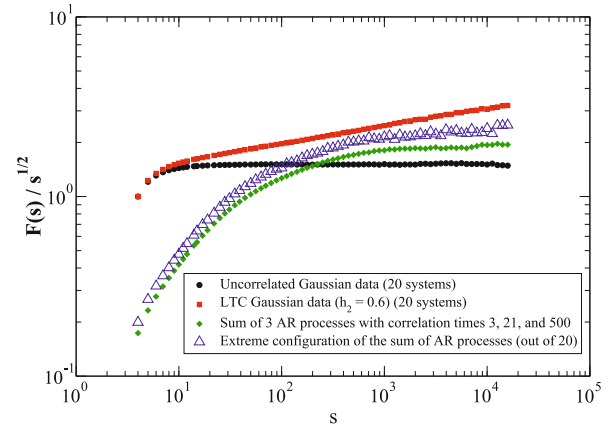
Fractal and Multifractal Time Series, Figure 6
Application of Detrended Fluctuation Analysis (DFA) on the data already studied in Fig. 5 (figure after [29])

$$F^2_{\mathrm{DFA}m}(\nu, s) = \frac{1}{s} \sum_{j=1}^{s} \tilde{Y}^2_s(j) . \tag{21}$$

As for FA and discrete wavelet analysis, the $F^2_{\mathrm{DFA}m}(\nu, s)$ are averaged over all segments ν to obtain the mean fluctuations $F_2(s)$, see Eq. (13). Calculating $F_2(s)$ for many s, the fluctuation scaling exponent α can be determined just as with FA, see Eq. (14). Figure 6 shows typical results for DFA of the same long-term correlated, short-term correlated and uncorrelated data studied already in Fig. 5.

We note that in studies that include averaging over many records (or one record cut into many separate pieces by the elimination of some unreliable intermediate data points) the averaging procedure (13) must be performed for all data. Taking the square root is always the final step after all averaging is finished. It is not appropriate to calculate $F_2(s)$ for parts of the data and then average the $F_2(s)$ values, since such a procedure will bias the results towards smaller scaling exponents on large time scales.

If $F_2(s)$ increases for increasing s by $F_2(s) \sim s^\alpha$ with $0.5 < \alpha < 1$, one finds that the scaling exponent $\alpha \approx H$ is related to the correlation exponent γ by $\alpha = 1 - \gamma/2$ (see Eq. (15)). A value of $\alpha = 0.5$ thus indicates that there are no (or only short-range) correlations. If $\alpha > 0.5$ for all scales s, the data are long-term correlated. The higher α, the stronger the correlations in the signal are. $\alpha > 1$ indicates a non-stationary local average of the data; in this case, FA fails and yields only $\alpha = 1$. The case $\alpha < 0.5$ corresponds to long-term anti-correlations, meaning that large values are most likely to be followed by small values and vice versa. α values below 0 are not possible. Since the

maximum value for α in DFAm is $m + 1$, higher detrending orders should be used for very non-stationary data with large α. Like in FA and Hurst's analysis, α will decrease or increase by one upon additional differentiation or integration of the data, respectively.

Small deviations from the scaling law (14), i.e. deviations from a straight line in a double logarithmic plot, occur for small scales s, in particular for DFAm with large detrending order m. These deviations are intrinsic to the usual DFA method, since the scaling behavior is only approached asymptotically. The deviations limit the capability of DFA to determine the correct correlation behavior in very short records and in the regime of small s. DFA6, e.g., is only defined for $s \geq 8$, and significant deviations from the scaling law $F_2(s) \sim s^\alpha$ occur even up to $s \approx 30$. They will lead to an over-estimation of the fluctuation exponent α, if the regime of small s is used in a fitting procedure. An approach for correction of this systematic artefact in DFA is described in [35].

The number of independent segments of length s is larger in DFA than in WT, and the fluctuations in FA are larger than in DFA. Hence, the analysis has to be based on s values lower than $s_{\max} = N/4$ for DFA compared with $s_{\max} = N/10$ for FA and WT. The accuracy of scaling exponents α determined by DFA was recently studied as a function of the length N of the data [43] (fitting range $s \in [10, N/2]$ was used). The results show that statistical standard errors of α (one standard deviation) are approximately 0.1 for $N = 500$, 0.05 for $N = 3000$, and reach 0.03 for $N = 10,000$. Findings of long-term correlations with $\alpha = 0.6$ in data with only 500 points are thus not significant.

A generalization of DFA for two-dimensional data (or even higher dimensions d) was recently suggested [34]. The generalization works well when tested with synthetic surfaces including fractional Brownian surfaces and multifractal surfaces. In the 2D procedure, a double cumulative sum (profile) is calculated by summing over both directional indices analogous with Eq. (10), $Y(k, l) = \sum_{i=1}^{k} \sum_{j=1}^{l} \tilde{x}_{i,j}$. This surface is partitioned into squares of size $s \times s$ with indices ν and μ, in which polynomials like $y^2_{\nu,\mu,s}(i, j) = ai^2 + bj^2 + cij + di + ej + f$ are fitted. The fluctuation function $F_2(s)$ is again obtained by calculating the variance of the profile from the fits.

Detection of Trends and Crossovers with DFA

Frequently, the correlations of recorded data do not follow the same scaling law for all time scales s, but one or sometimes even more crossovers between different scaling regimes are observed (see Subsect. "Crossovers and Non-stationarities in Time Series"). Time series with a well-defined crossover at $s_\times$ and vanishing correlations above $s_\times$ are most easily generated by Fourier filtering (see Subsect. "Fourier Filtering"). The power spectrum $S(f)$ of an uncorrelated random series is multiplied by $(f/f_\times)^{-\beta}$ with $\beta = 2\alpha - 1$ for frequencies $f > f_\times = 1/s_\times$ only. The series obtained by inverse Fourier transform of this modified power spectrum exhibits power-law correlations on time scales $s < s_\times$ only, while the behavior becomes uncorrelated on larger time scales $s > s_\times$.

The crossover from $F_2(s) \sim s^\alpha$ to $F_2(s) \sim s^{1/2}$ is clearly visible in double logarithmic plots of the DFA fluctuation function for such short-term correlated data. However, it occurs at times $s_\times^{(m)}$ that are different from the original $s_\times$ used for the generation of the data and that depend on the detrending order m. This systematic deviation is most significant in the DFAm with higher m. Extensive numerical simulations (see Fig. 3 in [35]) show that the ratios of $s_\times^{(m)}/s_\times$ are 1.6, 2.6, 3.6, 4.5, and 5.4 for DFA1, DFA2, …, DFA5, with an error bar of approximately 0.1. Note, however, that the precise value of this ratio will depend on the method used for fitting the crossover times $s_\times^{(m)}$ (and the method used for generating the data if generated data is analyzed). If results for different orders of DFA shall be compared, an observed crossover $s_\times^{(m)}$ can be systematically corrected dividing by the ratio for the corresponding DFAm. If several orders of DFA are used in the procedure, several estimates for the real $s_\times$ will be obtained, which can be checked for consistency or used for an error approximation. A real crossover can thus be well distinguished from the effects of non-stationarities in the data, which lead to a different dependence of an apparent crossover on m.

The procedure is also required if the characteristic time scale of short-term correlations shall be studied with DFA. If consistent (corrected) $s_\times$ values are obtained based on DFAm with different m, the existence of a real characteristic correlation time scale is positively confirmed. Note that lower detrending orders are advantageous in this case, since the observed crossover time scale $s_\times^{(m)}$ might become quite large and nearly reach one forth of the total series length ($N/4$), where the results become statistically inaccurate.

We would like to note that studies showing scaling long-term correlations should not be based on DFA or variants of this method alone in most applications. In particular, if it is not clear whether a given time series is indeed long-term correlated or just short-term correlated with a fairly large crossover time scale, results of DFA should be compared with other methods. For example, one can employ wavelet methods (see, e.g., Subsect. "Discrete

Wavelet Transform (WT) Approach"). Another option is to remove short-term correlations by considering averaged series for comparison. For a time series with daily observations and possible short-term correlations up to two years, for example, one might consider the series of two-year averages and apply DFA together with FA, binned power spectra analysis, and/or wavelet analysis. Only if these methods still indicate long-term correlations, one can be sure that the data are indeed long-term correlated.

As discussed in Subsect. "Crossovers and Non-stationarities in Time Series", records from real measurements are often affected by non-stationarities, and in particular by trends. They have to be well distinguished from the intrinsic fluctuations of the system. To investigate the effect of trends on the DFAm fluctuation functions, one can generate artificial series $(\tilde{x}_i)$ with smooth monotonous trends by adding polynomials of different power p to the original record (x_i),

$$\tilde{x}_i = x_i + Ax^p \quad \text{with} \quad x = i/N \ . \tag{22}$$

For the DFAm, such trends in the data can lead to an artificial crossover in the scaling behavior of $F_2(s)$, i. e., the slope α is strongly increased for large time scales s. The position of this artificial crossover depends on the strength A and the power p of the trend. Evidently, no artificial crossover is observed, if the detrending order m is larger than p and p is integer. The order p of the trends in the data can be determined easily by applying the different DFAm. If p is larger than m or p is not an integer, an artificial crossover is observed, the slope α_{trend} in the large s regime strongly depends on m, and the position of the artificial crossover also depends strongly on m. The artificial crossover can thus be clearly distinguished from real crossovers in the correlation behavior, which result in identical slopes α and rather similar crossover positions for all detrending orders m. For more extensive studies of trends with non-integer powers we refer to [35,36]. The effects of periodic trends are also studied in [35].

If the functional form of the trend in given data is not known a priori, the fluctuation function $F_2(s)$ should be calculated for several orders m of the fitting polynomial. If m is too low, $F_2(s)$ will show a pronounced crossover to a regime with larger slope for large scales s [35,36]. The maximum slope of $\log F_2(s)$ versus $\log s$ is $m + 1$. The crossover will move to larger scales s or disappear when m is increased, unless it is a real crossover not due to trends. Hence, one can find m such that detrending is sufficient. However, m should not be larger than necessary, because shifts of the observed crossover time scales and deviations on short scales s increase with increasing m.

Sign and Magnitude (Volatility) DFA

To study the origin of long-term fractal correlations in a time series, the series can be split into two parts which are analyzed separately. It is particularly useful to split the series of increments, $\Delta x_i = x_i - x_{i-1}$, $i = 1, \ldots, N$, into a series of signs $\tilde{x}_i = s_i = \text{sign}\Delta x_i$ and a series of magnitudes $\tilde{x}_i = m_i = |\Delta x_i|$ [32,47,48]. There is an extensive interest in the magnitude time series in economics [49,50]. These data, usually called volatility, represent the absolute variations in stock (or commodity) prices and are used as a measure quantifying the risk of investments. While the actual prices are only short-term correlated, long-term correlations have been observed in volatility series [49,50].

Time series having identical distributions and long-range correlation properties can exhibit quite different temporal organizations of the magnitude and sign sub-series. The DFA method can be applied independently to both of these series. Since in particular the signs are often rather strongly anti-correlated and DFA will give incorrect results if α is too close to zero, one often studies integrated sign and magnitude series. As mentioned above, integration $\tilde{x}_i \rightarrow \sum_{j=1}^{i} \tilde{x}_j$ increases α by one.

Most published results report short-term anti-correlations and no long-term correlations in the sign series, i. e., $\alpha_{\text{sign}} < 1/2$ for the non-integrated signs s_i (or $\alpha_{\text{sign}} < 3/2$ for the integrated signs) on low time scales and $\alpha_{\text{sign}} \rightarrow 1/2$ asymptotically for large s. The magnitude series, on the other hand, are usually either uncorrelated $\alpha_{\text{magn}} = 1/2$ (or $3/2$) or positively long-term correlated $\alpha_{\text{magn}} > 1/2$ (or $3/2$). It has been suggested that findings of $\alpha_{\text{magn}} > 1/2$ are related with nonlinear properties of the data and in particular multifractality [32,47,48], if $\alpha < 1.5$ in standard DFA. Specifically, the results suggest that the correlation exponent of the magnitude series is a monotonically increasing function of the multifractal spectrum (i. e., the singularity spectrum) width of the original series (see Subsect. "The Structure Function Approach and Singularity Spectra"). On the other hand, the sign series mainly relates to linear properties of the original series. At small time scales $s < 16$ the standard α is approximately the average of α_{sign} and α_{magn}, if integrated sign and magnitude series are analyzed. For $\alpha > 1.5$ in the original series, the integrated magnitude and sign series have approximately the same two-point scaling exponents [47]. An analytical treatment is presented in [48].

Further Detrending Approaches

A possible drawback of the DFA method is the occurrence of abrupt jumps in the detrended profile $\tilde{Y}_s(j)$ (Eq. (20)) at the boundaries between the segments, since the fit-

ting polynomials in neighboring segments are not related. A possible way to avoid these jumps would be the calculation of $F_2(s)$ based on polynomial fits in overlapping windows. However, this is rather time consuming due to the polynomial fit in each segment and is consequently not done in most applications. To overcome the problem of jumps several modifications and extensions of the FA and DFA methods have been suggested in recent years. These methods include

- The detrended moving average technique [51,52,53], which we denote by the backward moving average (BMA) technique (following [54]),
- The centered moving average (CMA) method [54], an essentially improved version of BMA,
- The modified detrended fluctuation analysis (MDFA) [55], which is essentially a mixture of old FA and DFA,
- The continuous DFA (CDFA) technique [56,57], which is particularly useful for the detection of crossovers,
- The Fourier DFA [58],
- A variant of DFA based on empirical mode decomposition (EMD) [59],
- A variant of DFA based on singular value decomposition (SVD) [60,61], and
- A variant of DFA based on high-pass filtering [62].

Detrended moving average techniques will be thoroughly described and discussed in the next section. A study comparing DFA with CMA and MDFA can be found in [43]. For studies comparing DFA and BMA, see [63,64]; note that [64] also discusses CMA.

The method we denote as *modified detrended fluctuation analysis* (MDFA) [55], eliminates trends similar to the DFA method. A polynomial is fitted to the profile function $Y(j)$ in each segment v and the deviation between the profile function and the polynomial fit is calculated, $\tilde{Y}_s(j) = Y(j) - y_{v,s}^p(j)$ (Eq. (20)). To estimate correlations in the data, this method uses a derivative of $\tilde{Y}_s(j)$, obtained for each segment v, by $\Delta \tilde{Y}_s(j) = \tilde{Y}_s(j + s/2) - \tilde{Y}_s(j)$. Hence, the fluctuation function (compare with Eqs. (13) and (21)) is calculated as follows:

$$F_2(s) = \left[\frac{1}{N} \sum_{j=1}^{N} \left(\tilde{Y}_s(j + s/2) - \tilde{Y}_s(j) \right)^2 \right]^{1/2} . \quad (23)$$

As in case of DFA, MDFA can easily be generalized to remove higher order trends in the data. Since the fitting polynomials in adjacent segments are not related, $\tilde{Y}_s(j)$ shows abrupt jumps on their boundaries as well. This leads

to fluctuations of $F_2(s)$ for large segment sizes s and limits the maximum usable scale to $s < N/4$ as for DFA. The detection of crossovers in the data, however, is more exact with MDFA (compared with DFA), since no correction of the estimated crossover time scales seems to be needed [43].

The *Fourier-detrended fluctuation analysis* [58] aims to eliminate slow oscillatory trends which are found especially in weather and climate series due to seasonal influences. The character of these trends can be rather periodic and regular or irregular, and their influence on the detection of long-range correlations by means of DFA was systematically studied previously [35]. Among other things it has been shown that slowly varying periodic trends disturb the scaling behavior of the results much stronger than quickly oscillating trends and thus have to be removed prior to the analysis. In the case of periodic and regular oscillations, e. g., in temperature fluctuations, one simply removes the low frequency seasonal trend by subtracting the daily mean temperatures from the data. Another way, which the Fourier-detrended fluctuation analysis suggests, is to filter out the relevant frequencies in the signals' Fourier spectrum before applying DFA to the filtered signal. Nevertheless, this method faces several difficulties especially its limitation to periodic and regular trends and the need for a priori knowledge of the interfering frequency band.

To study correlations in data with quasi-periodic or irregular oscillating trends, *empirical mode decomposition* (EMD) was suggested [59]. The EMD algorithm breaks down the signal into its intrinsic mode functions (IMFs) which can be used to distinguish between fluctuations and background. The background, estimated by a quasi-periodic fit containing the dominating frequencies of a sufficiently large number of IMFs, is subtracted from the data, yielding a slightly better scaling behavior in the DFA curves. However, we believe that the method might be too complicated for wide-spread applications.

Another method which was shown to minimize the effect of periodic and quasi-periodic trends is based on *singular value decomposition* (SVD) [60,61]. In this approach, one first embeds the original signal in a matrix whose dimension has to be much larger than the number of frequency components of the periodic or quasi-periodic trends obtained in the power spectrum. Applying SVD yields a diagonal matrix which can be manipulated by setting the dominant eigenvalues (associated with the trends) to zero. The filtered matrix finally leads to the filtered data, and it has been shown that subsequent application of DFA determines the expected scaling behavior if the embedding dimension is sufficiently large. None the

less, the performance of this rather complex method seems to decrease for larger values of the scaling exponent. Furthermore SVD-DFA assumes that trends are deterministic and narrow banded.

The detrending procedure in DFA (Eq. (20)) can be regarded as a scale-dependent high-pass filter since (low-frequency) fluctuations exceeding a specific scale s are eliminated. Therefore, it has been suggested to obtain the detrended profile $\tilde{Y}_s(j)$ for each scale s directly by applying digital high-pass filters [62]. In particular, Butterworth, Chebyshev-I, Chebyshev-II, and an elliptical filter were suggested. While the elliptical filter showed the best performance in detecting long-range correlations in artificial data, the Chebyshev-II filter was found to be problematic. Additionally, in order to avoid a time shift between filtered and original profile, the average of the directly filtered signal and the time reversed filtered signal is considered. The effects of these complicated filters on the scaling behavior are, however, not fully understood.

Finally, a continuous DFA method has been suggested in the context of studying heartbeat data during sleep [56,57]. The method compares unnormalized fluctuation functions $F_2(s)$ for increasing length of the data. I. e., one starts with a very short recording and subsequently adds more points of data. The method is particularly suitable for the detection of change points in the data, e. g., physiological transitions between different activity or sleep stages. Since the main objective of the method is not the study of scaling behavior, we do not discuss it in detail here.

Centered Moving Average (CMA) Analysis

Particular attractive modifications of DFA are the *detrended moving average* (DMA) methods, where running averages replace the polynomial fits. The first suggested version, the *backward moving average* (BMA) method [51,52,53], however, suffers from severe problems, because an artificial time shift of s between the original signal and the moving average is introduced. This time shift leads to an additional contribution to the detrended profile $\tilde{Y}_s(j)$, which causes a larger fluctuation function $F_2(s)$ in particular for small scales in the case of long-term correlated data. Hence, the scaling exponent α is systematically underestimated [63]. In addition, the BMA method preforms even worse for data with trends [64], and its slope is limited by $\alpha < 1$ just as for the non-detrending method FA.

It was soon recognized that the intrinsic error of BMA can be overcome by eliminating the artificial time shift. This leads to the *centered moving average* (CMA)

method [54], where $\tilde{Y}_s(j)$ is calculated as

$$\tilde{Y}_s(j) = Y(j) - \frac{1}{s} \sum_{i=-(s-1)/2}^{(s-1)/2} Y(j+i), \qquad (24)$$

replacing Eq. (20) while Eq. (21) and the rest of the DFA procedure described in Subsect. "Detrended Fluctuation Analysis (DFA)" stay the same. Unlike DFA, the CMA method cannot easily be generalized to remove linear and higher order trends in the data.

It was recently proposed [43] that the scaling behavior of the CMA method is more stable than for DFA1 and MDFA1, suggesting that CMA could be used for reliable computation of α even for scales $s < 10$ (without correction of any systematic deviations needed in DFA for this regime) and up to $s_{\max} = N/2$. The standard errors in determining the scaling exponent α by fitting straight lines to the double logarithmic plots of $F_2(s)$ have been studied in [43]; they are comparable with DFA1 (see end of Subsect. "Detrended Fluctuation Analysis (DFA)").

Regarding the determination of crossovers, CMA is comparable to DFA1. Ultimately, the CMA seems to be a good alternative to DFA1 when analyzing the scaling properties in short data sets without trends. Nevertheless for data with possible unknown trends we recommend the application of standard DFA with several different detrending polynomial orders in order to distinguish real crossovers from artificial crossovers due to trends. In addition, an independent approach (e. g., wavelet analysis) should be used to confirm findings of long-term correlations (see also Subsect. "Detection of Trends and Crossovers with DFA").

Methods for Multifractal Time Series Analysis

This section describes the multifractal characterization of time series, for an introduction, see Subsect. "Multifractal Time Series". The simplest type of multifractal analysis is based upon the standard partition function multifractal formalism, which has been developed for the multifractal characterization of normalized, stationary measures [6,12,65,66]. Unfortunately, this standard formalism does not give correct results for non-stationary time series that are affected by trends or that cannot be normalized. Thus, in the early 1990s an improved multifractal formalism has been developed, the wavelet transform modulus maxima (WTMM) method [67,68,69,70,71], which is based on wavelet analysis and involves tracing the maxima lines in the continuous wavelet transform over all scales. An important alternative is the multifractal DFA (MFDFA) algorithm [33], which does not require the

modulus maxima procedure, and hence involves little more effort in programming than the conventional DFA. For studies comparing methods for detrending multifractal analysis (MFDFA and WTMM, see [33,72,73].

The Structure Function Approach and Singularity Spectra

In the general multifractal formalism, one considers a normalized measure $\mu(t)$, $t \in [0, 1]$, and defines the box probabilities $\tilde{\mu}_s(t) = \int_{t-s/2}^{t+s/2} \mu(t') \, dt'$ in neighborhoods of (scale) length $s \ll 1$ around t. The multifractal approach is then introduced by the partition function

$$Z_q(s) = \sum_{\nu=0}^{1/s-1} \tilde{\mu}_s^q[(\nu + 1/2)s] \sim s^{\tau(q)} \quad \text{for } s \ll 1 , \quad (25)$$

where $\tau(q)$ is the Renyi scaling exponent and q is a real parameter that can take positive as well as negative values. Note that $\tau(q)$ is sometimes defined with opposite sign (see, e.g., [6]). A record is called monofractal (or self-affine), when the Renyi scaling exponent $\tau(q)$ depends linearly on q; otherwise it is called multifractal. The generalized multifractal dimensions $D(q)$ (see also Subsect. "Multifractal Time Series") are related to $\tau(q)$ by $D(q) = \tau(q)/(q-1)$, such that the fractal dimension of the support is $D(0) = -\tau(0)$ and the correlation dimension is $D(2) = \tau(2)$.

In time series, a discrete version has to be used, and the considered data (x_i), $i = 1, \ldots, N$ may usually include negative values. Hence, setting $N_s = \text{int}(N/s)$ and $X(\nu, s) = \sum_{i=1}^{s} x_{\nu s+i}$ for $\nu = 0, \ldots, N_s - 1$ we can define [6,12],

$$Z_q(s) = \sum_{\nu=0}^{N_s-1} |X(\nu, s)|^q \sim s^{\tau(q)} \quad \text{for } s > 1 . \quad (26)$$

Inserting the profile $Y(j)$ and $F_{\text{FA}}(\nu, s)$ from Eqs. (10) and (11), respectively, we obtain

$$Z_q(s) = \sum_{\nu=0}^{N_s-1} \left\{ [Y((\nu + 1)s) - Y(\nu s)]^2 \right\}^{q/2}$$

$$= \sum_{\nu=1}^{N_s} |F_{\text{FA}}(\nu, s)| . \quad (27)$$

Comparing Eq. (27) with (13), we see that this multifractal approach can be considered as a generalized version of the fluctuation analysis (FA) method, where the exponent 2 is replaced by q. In particular we find (disregarding the summation over the second partition of the time series)

$$F_2(s) \sim \left[\frac{1}{N_s} Z_2(s) \right]^{1/2} \sim s^{[1+\tau(2)]/2}$$

$$\Rightarrow 2\alpha = 1 + \tau(2) = 1 + D(2). \quad (28)$$

We thus see that all methods for (mono-)fractal time analysis (discussed in Sect. "Methods for Stationary Fractal Time Series Analysis" and Sect. "Methods for Non-stationary Fractal Time Series Analysis") in fact study the correlation dimension $D(2) = 2\alpha - 1 = \beta = 1 - \gamma$ (see Eq. (15)).

It is straightforward to define a generalized (multifractal) Hurst exponent $h(q)$ for the scaling behavior of the qth moments of the fluctuations [65,66],

$$F_q(s) = \left[\frac{1}{N_s} Z_2(s) \right]^{1/q} \sim s^{[1+\tau(q)]/q} = s^{h(q)}$$

$$\Rightarrow h(q) = \frac{1 + \tau(q)}{q} \quad (29)$$

with $h(2) = \alpha \approx H$. In the following, we will use only $h(2)$ for the standard fluctuation exponent (denoted by α in the previous chapters), and reserve the letter α for the Hölder exponent.

Another way to characterize a multifractal series is the singularity spectrum $f(\alpha)$, that is related to $\tau(q)$ via a Legendre transform [6,12],

$$\alpha = \frac{\mathrm{d}}{\mathrm{d}q} \tau(q) \quad \text{and} \quad f(\alpha) = q\alpha - \tau(q) . \quad (30)$$

Here, α is the singularity strength or Hölder exponent (see also ▶ Fractals and Multifractals, Introduction to in the encyclopedia), while $f(\alpha)$ denotes the dimension of the subset of the series that is characterized by α. Note that α is *not* the fluctuation scaling exponent in this section, although the same letter is traditionally used for both. Using Eq. (29), we can directly relate α and $f(\alpha)$ to $h(q)$,

$$\alpha = h(q) + qh'(q) \quad \text{and} \quad f(\alpha) = q[\alpha - h(q)] + 1. \quad (31)$$

Wavelet Transform Modulus Maxima (WTMM) Method

The wavelet transform modulus maxima (WTMM) method [67,68,69,70,71] is a well-known method to investigate the multifractal scaling properties of fractal and self-affine objects in the presence of non-stationarities. For applications, see e.g. [74,75]. It is based upon the wavelet transform with continuous basis functions as defined in Subsect. "Wavelet Analysis", Eq. (16). Note that in this

case the series $\tilde{x}_i$ are analyzed directly instead of the profile $Y(j)$ defined in Eq. (10). Using wavelets orthogonal to mth order polynomials, the corresponding trends are eliminated.

Instead of averaging over all wavelet coefficients $L_\psi(\tau, s)$, one averages, within the modulo-maxima method, only the local maxima of $|L_\psi(\tau, s)|$. First, one determines for a given scale s, the positions τ_j of the local maxima of $|W(\tau, s)|$ as a function of τ, so that $|L_\psi(\tau_j - 1, s)| < |L_\psi(\tau_j, s)| \geq |L_\psi(\tau_j + 1, s)|$ for $j = 1, \ldots, j_{\max}$. This maxima procedure is demonstrated in Fig. 7. Then one sums up the qth power of the maxima,

$$Z(q, s) = \sum_{j=1}^{j_{\max}} |L_\psi(\tau_j, s)|^q . \tag{32}$$

The reason for the maxima procedure is that the absolute wavelet coefficients $|L_\psi(\tau, s)|$ can become arbitrarily small. The analyzing wavelet $\psi(x)$ must always have positive values for some x and negative values for other x, since it has to be orthogonal to possible constant trends. Hence there are always positive and negative terms in the sum (16), and these terms might cancel. If that happens, $|L_\psi(\tau, s)|$ can become close to zero. Since such small terms would spoil the calculation of negative moments in Eq. (32), they have to be eliminated by the maxima procedure.

In fluctuation analysis, on the other hand, the calculation of the variances $F^2(\nu, s)$, e. g. in Eq. (11), involves only positive terms under the summation. The variances cannot become arbitrarily small, and hence no maximum procedure is required for series with compact support. In addition, the variances will increase if the segment length s is increased, because the fit will usually be worse for a longer segment. In the WTMM method, in contrast, the absolute wavelet coefficients $|L_\psi(\tau, s)|$ need not increase with increasing scale s, even if only the local maxima are considered. The values $|L_\psi(\tau, s)|$ might become smaller for increasing s since just more (positive and negative) terms are included in the summation (16), and these might cancel even better. Thus, an additional supremum procedure has been introduced in the WTMM method in order to keep the dependence of $Z(q, s)$ on s monotonous. If, for a given scale s, a maximum at a certain position τ_j happens to be smaller than a maximum at $\tau'_j \approx \tau_j$ for a lower scale $s' < s$, then $L_\psi(\tau_j, s)$ is replaced by $L_\psi(\tau'_j, s')$ in Eq. (32).

Often, scaling behavior is observed for $Z(q, s)$, and scaling exponents $\hat{\tau}(q)$ can be defined that describe how $Z(q, s)$ scales with s,

$$Z(q, s) \sim s^{\hat{\tau}(q)} . \tag{33}$$
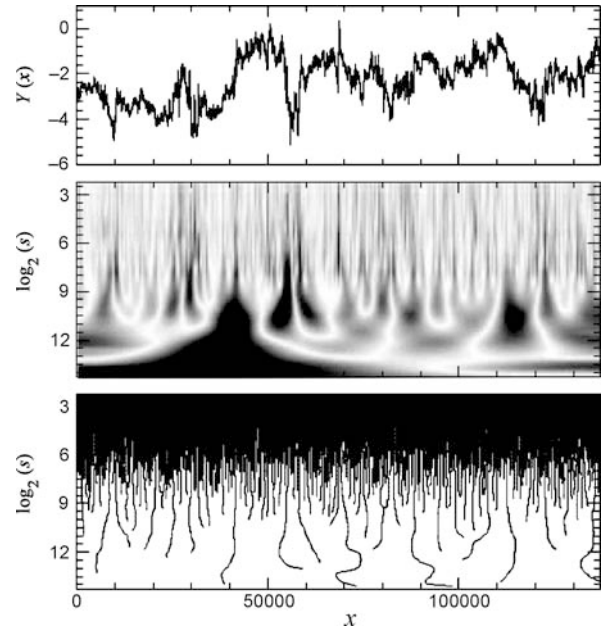
Fractal and Multifractal Time Series, Figure 7
Example of the wavelet transform modulus maxima (WTMM) method, showing the original data (*top*), its continuous wavelet transform (*gray scale* coded amplitude of wavelet coefficients, *middle*), and the extracted maxima lines (*bottom*) (figure taken from [68])

The exponents $\hat{\tau}(q)$ characterize the multifractal properties of the series under investigation, and theoretically they are identical with the $\tau(q)$ defined in Eq. (26) [67,68,69,71] and related to $h(q)$ by Eq. (29).

Multifractal Detrended Fluctuation Analysis (MFDFA)

The multifractal DFA (MFDFA) procedure consists of five steps [33]. The first three steps are essentially identical to the conventional DFA procedure (see Subsect. "Detrended Fluctuation Analysis (DFA)" and Fig. 4). Let us assume that $(\tilde{x}_i)$ is a series of length N, and that this series is of compact support. The support can be defined as the set of the indices j with nonzero values $\tilde{x}_j$, and it is compact if $\tilde{x}_j = 0$ for an insignificant fraction of the series only. The value of $\tilde{x}_j = 0$ is interpreted as having no value at this j. Note that we are *not* discussing the fractal or multifractal features of the plot of the time series in a two-dimensional graph (see also the discussion in Subsect. "Fractality, Self-Affinity, and Scaling"), but analyzing time series as one-dimensional structures with values assigned to each point. Since real time series always have finite length N, we explicitly want to determine the multifractality of finite series, and we are not discussing the limit for $N \to \infty$ here

(see also Subsect. "The Structure Function Approach and Singularity Spectra").

Step 1 Calculate the profile $Y(j)$, Eq. (10), by integrating the time series.

Step 2 Divide the profile $Y(j)$ into $N_s = \text{int}(N/s)$ non-overlapping segments of equal length s. Since the length N of the series is often not a multiple of the considered time scale s, the same procedure can be repeated starting from the opposite end. Thereby, $2N_s$ segments are obtained altogether.

Step 3 Calculate the local trend for each of the $2N_s$ segments by a least-square fit of the profile. Then determine the variance by Eqs. (20) and (21) for each segment $v = 1, \dots, 2N_s$. Again, linear, quadratic, cubic, or higher order polynomials can be used in the fitting procedure, and the corresponding methods are thus called MFDFA1, MFDFA2, MFDFA3, …) [33]. In (MF-)DFAm [mth order (MF-)DFA] trends of order m in the profile (or, equivalently, of order $m - 1$ in the original series) are eliminated. Thus a comparison of the results for different orders of DFA allows one to estimate the type of the polynomial trend in the time series [35,36].

Step 4 Average over all segments to obtain the qth order fluctuation function

$$F_q(s) = \left\{ \frac{1}{2N_s} \sum_{v=1}^{2N_s} \left[F_{\text{DFA}m}^2(v, s) \right]^{q/2} \right\}^{1/q} . \quad (34)$$

This is the generalization of Eq. (13) suggested by the relations derived in Subsect. "The Structure Function Approach and Singularity Spectra". For $q = 2$, the standard DFA procedure is retrieved. One is interested in how the generalized q dependent fluctuation functions $F_q(s)$ depend on the time scale s for different values of q. Hence, we must repeat steps 2 to 4 for several time scales s. It is apparent that $F_q(s)$ will increase with increasing s. Of course, $F_q(s)$ depends on the order m. By construction, $F_q(s)$ is only defined for $s \geq m + 2$.

Step 5 Determine the scaling behavior of the fluctuation functions by analyzing log-log plots $F_q(s)$ versus s for each value of q. If the series $\tilde{x}_i$ are long-range power-law correlated, $F_q(s)$ increases, for large values of s, as a power-law,

$$F_q(s) \sim s^{h(q)} \quad \text{with} \quad h(q) = \frac{1 + \tau(q)}{q} . \quad (35)$$

For very large scales, $s > N/4$, $F_q(s)$ becomes statistically unreliable because the number of segments N_s for the averaging procedure in step 4 becomes very small. Thus, scales $s > N/4$ should be excluded from the fitting procedure determining $h(q)$. Besides that, systematic deviations from the scaling behavior in Eq. (35), which can be corrected, occur for small scales $s \approx 10$.

The value of $h(0)$, which corresponds to the limit $h(q)$ for $q \to 0$, cannot be determined directly using the averaging procedure in Eq. (34) because of the diverging exponent. Instead, a logarithmic averaging procedure has to be employed,

$$F_0(s) = \exp \left\{ \frac{1}{4N_s} \sum_{v=1}^{2N_s} \ln \left[F^2(v, s) \right] \right\} \sim s^{h(0)} . \quad (36)$$

Note that $h(0)$ cannot be defined for time series with fractal support, where $h(q)$ diverges for $q \to 0$.

For monofractal time series with compact support, $h(q)$ is independent of q, since the scaling behavior of the variances $F_{\text{DFA}m}^2(v, s)$ is identical for all segments v, and the averaging procedure in Eq. (34) will give just this identical scaling behavior for all values of q. Only if small and large fluctuations scale differently, there will be a significant dependence of $h(q)$ on q. If we consider positive values of q, the segments v with large variance $F^2(v, s)$ (i.e. large deviations from the corresponding fit) will dominate the average $F_q(s)$. Thus, for positive values of q, $h(q)$ describes the scaling behavior of the segments with large fluctuations. On the contrary, for negative values of q, the segments v with small variance $F_{\text{DFA}m}^2(v, s)$ will dominate the average $F_q(s)$. Hence, for negative values of q, $h(q)$ describes the scaling behavior of the segments with small fluctuations. Figure 8 shows typical results obtained for $F_q(s)$ in the MFDFA procedure.

Usually the large fluctuations are characterized by a smaller scaling exponent $h(q)$ for multifractal series than the small fluctuations. This can be understood from the following arguments. For the maximum scale $s = N$ the fluctuation function $F_q(s)$ is independent of q, since the sum in Eq. (34) runs over only two identical segments. For smaller scales $s \ll N$ the averaging procedure runs over several segments, and the average value $F_q(s)$ will be dominated by the $F^2(v, s)$ from the segments with small (large) fluctuations if $q < 0$ ($q > 0$). Thus, for $s \ll N$, $F_q(s)$ with $q < 0$ will be smaller than $F_q(s)$ with $q > 0$, while both become equal for $s = N$. Hence, if we assume an homogeneous scaling behavior of $F_q(s)$ following Eq. (35), the slope $h(q)$ in a log-log plot of $F_q(s)$ with $q < 0$ versus s must be larger than the corresponding slope for $F_q(s)$ with $q > 0$. Thus, $h(q)$ for $q < 0$ will usually be larger than $h(q)$ for $q > 0$.
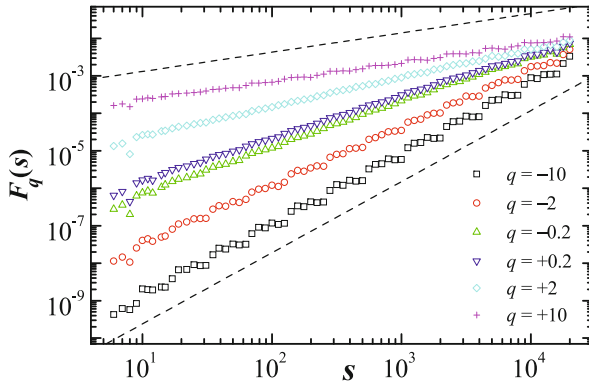
Fractal and Multifractal Time Series, Figure 8
Multifractal detrended fluctuation analysis (MFDFA) of data from the binomial multifractal model (see Subsect. "The Extended Binomial Multifractal Model") with $a = 0.75$. $F_q(s)$ is plotted versus s for the q values given in the legend; the slopes of the curves correspond to the values of $h(q)$. The *dashed lines* have the slopes of the theoretical slopes $h(\pm\infty)$ from Eq. (42). 100 configurations have been averaged

However, the MFDFA method can only determine *positive* generalized Hurst exponents $h(q)$, and it already becomes inaccurate for strongly anti-correlated signals when $h(q)$ is close to zero. In such cases, a modified (MF-)DFA technique has to be used. The most simple way to analyze such data is to integrate the time series before the MFDFA procedure. Following the MFDFA procedure as described above, we obtain a generalized fluctuation functions described by a scaling law with $\tilde{h}(q) = h(q) + 1$. The scaling behavior can thus be accurately determined even for $h(q)$ which are smaller than zero for some values of q.

The accuracy of $h(q)$ determined by MFDFA certainly depends on the length N of the data. For $q = \pm 10$ and data with $N = 10{,}000$ and $100{,}000$, systematic and statistical error bars (standard deviations) up to $\Delta h(q) \approx \pm 0.1$ and $\approx \pm 0.05$ should be expected, respectively [33]. A difference of $h(-10) - h(+10) = 0.2$, corresponding to an even larger width $\Delta\alpha$ of the singularity spectrum $f(\alpha)$ defined in Eq. (30) is thus not significant unless the data was longer than $N = 10{,}000$ points. Hence, one has to be very careful when concluding multifractal properties from differences in $h(q)$.

As already mentioned in the introduction, two types of multifractality in time series can be distinguished. Both of them require a multitude of scaling exponents for small and large fluctuations: (i) Multifractality of a time series can be due to a broad probability density function for the values of the time series, and (ii) multifractality can also be due to different long-range correlations for small and large fluctuations. The most easy way to distinguish be-

tween these two types is by analyzing also the corresponding randomly shuffled series [33]. In the shuffling procedure the values are put into random order, and thus all correlations are destroyed. Hence the shuffled series from multifractals of type (ii) will exhibit simple random behavior, $h_{\text{shuf}}(q) = 0.5$, i. e. non-multifractal scaling. For multifractals of type (i), on the contrary, the original $h(q)$ dependence is not changed, $h(q) = h_{\text{shuf}}(q)$, since the multifractality is due to the probability density, which is not affected by the shuffling procedure. If both kinds of multifractality are present in a given series, the shuffled series will show weaker multifractality than the original one.

Comparison of WTMM and MFDFA

The MFDFA results turn out to be slightly more reliable than the WTMM results [33,72,73]. In particular, the MFDFA has slight advantages for negative q values and short series. In the other cases the results of the two methods are rather equivalent. Besides that, the main advantage of the MFDFA method compared with the WTMM method lies in the simplicity of the MFDFA method. However, contrary to WTMM, MFDFA is restricted to studies of data with full one-dimensional support, while WTMM is not. Both WTMM and MFDFA have been generalized for higher dimensional data, see [34] for higher dimensional MFDFA and, e. g., [71] for higher dimensional WTMM. Studies of other generalizations of detrending methods like the discrete WT approach (see Subsect. "Discrete Wavelet Transform (WT) Approach") and the CMA method (see Subsect. "Centered Moving Average (CMA) Analysis") are currently under investigation [76].

Statistics of Extreme Events in Fractal Time Series

The statistics of return intervals between well defined extremal events is a powerful tool to characterize the temporal scaling properties of observed time series and to derive quantities for the estimation of the risk for hazardous events like floods, very high temperatures, or earthquakes. It was shown recently that long-term correlations represent a natural mechanism for the clustering of the hazardous events [77]. In this section we will discuss the most important consequences of long-term correlations and fractal scaling of time series upon the statistics of extreme events [77,78,79,80,81]. Corresponding work regarding multifractal data [82] is not discussed here.

Return Intervals Between Extreme Events

To study the statistics of return intervals we consider again a time series (x_i), $i = 1, \ldots, N$ with fractal scaling behav-

ior, sampled homogeneously and normalized to zero mean and unit variance. For describing the reoccurrence of rare events exceeding a certain threshold q, we investigate the return intervals r_q between these events, see Fig. 9. The average return interval $R_q = \langle r_q \rangle$ increases as a function of the threshold q (see, e. g. [83]). It is known that for uncorrelated records ('white noise'), the return intervals are also uncorrelated and distributed according to the Poisson distribution, $P_q(r) = \frac{1}{R_q} \exp(-r/R_q)$. For fractal (long-term correlated) data with auto-correlations following Eq. (5), we obtain a *stretched* exponential [77,78,79,80,84],

$$P_q(r) = \frac{a_\gamma}{R_q} \exp\left[-b_\gamma (r/R_q)^\gamma\right] . \tag{37}$$

This behavior is shown in Fig. 10. The exponent γ is the correlation exponent from $C(s)$, and the parameters a_γ and b_γ are independent of q. They can be determined from the normalization conditions for $P_q(r)$, i. e., $\int P_q(r)\,dr = 1$ and $\int r P_q(r)\,dr = R_q$. The form of the distribution (37) indicates that return intervals both well below and well above their average value R_q (which is independent of γ) are considerably more frequent for long-term correlated

than for uncorrelated data. It has to be noted that there are deviations from the stretched exponential law (37) for very small r (discretization effects and an additional power-law regime) and for very large r (finite size effects), see Fig. 10. The extent of the deviations from Eq. (37) depends on the distribution of the values x_i of the time series. For a discussion of these effects, see [80].

Equation (37) does not quantify, however, if the return intervals themselves are arranged in a correlated or in an uncorrelated fashion, and if clustering of rare events may be induced by long-term correlations. To study this question, one has to evaluate the auto-covariance function $C_r(s) = \langle r_q(l) r_q(l+s) \rangle - R_q^2$ of the return intervals. The results for model data suggests that also the return intervals are long-term power-law correlated, with the same exponent γ as the original record. Accordingly, large and small return intervals are not arranged in a random fashion but are expected to form clusters. As a consequence, the probability of finding a certain return interval r depends on the value of the preceding interval r_0, and this effect has to be taken into account in predictions and risk estimations [77,80].
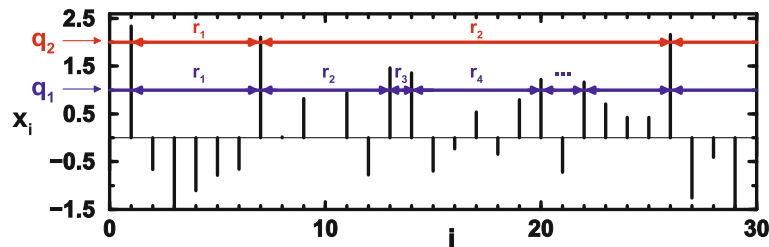
Fractal and Multifractal Time Series, Figure 9
Illustration for the definition of return intervals r_q between extreme events above two quantiles (thresholds) q_1 and q_2 (figure by Jan Eichner)
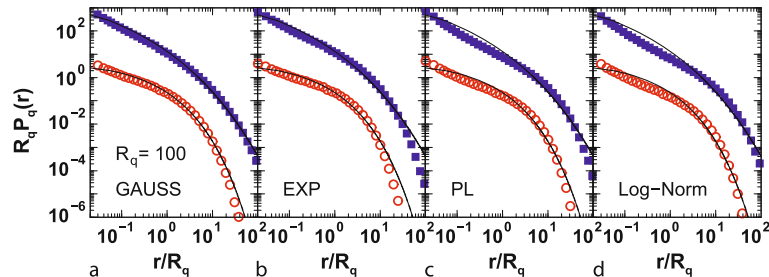
Fractal and Multifractal Time Series, Figure 10
Normalized rescaled distribution density functions $R_q P_q(r)$ of r values with $R_q = 100$ as a function of r/R_q for long-term correlated data with $\gamma = 0.4$ (*open symbols*) and $\gamma = 0.2$ (*filled symbols*; we multiplied the data for the filled symbols by a factor 100 to avoid overlapping curves). In **a** the original data were Gaussian distributed, in **b** exponentially distributed, in **c** power-law distributed with power -5.5, and in **d** log-normally distributed. All four figures follow quite well stretched exponential curves (*solid lines*) over several decades. For small r/R_q values a power-law regime seems to dominate, while on large scales deviations from the stretched exponential behavior are due to finite-size effects (figure by Jan Eichner)

The conditional distribution function $P_q(r|r_0)$ is a basic quantity, from which the relevant quantities in risk estimations can be derived [83]. For example, the first moment of $P_q(r|r_0)$ is the average value $R_q(r_0)$ of those return intervals that directly follow r_0. By definition, $R_q(r_0)$ is the expected waiting time to the next event, when the two events before were separated by r_0. The more general quantity is the expected waiting time $\tau_q(x|r_0)$ to the next event, when the time x has elapsed. For $x = 0$, $\tau_q(0|r_0)$ is identical to $R_q(r_0)$. In general, $\tau_q(x|r_0)$ is related to $P_q(r|r_0)$ by

$$\tau_q(x|r_0) = \int_x^\infty (r-x)P_q(r|r_0)\mathrm{d}r \Big/ \int_x^\infty P_q(r|r_0)\mathrm{d}r. \quad (38)$$

For uncorrelated records, $\tau_q(x|r_0)/R_q = 1$ (except for discreteness effects that lead to $\tau_q(x|r_0)/R_q > 1$ for $x > 0$, see [85]). Due to the scaling of $P_q(r|r_0)$, also $\tau_q(x|r_0)/R_q$ scales with r_0/R_q and x/R_q. Small and large return intervals are more likely to be followed by small and large ones, respectively, and hence $\tau_q(0|r_0)/R_q = R_q(r_0)/R_q$ is well below (above) one for r_0/R_q well below (above) one. With increasing x, the expected residual time to the next event increases. Note that only for an infinite long-term correlated record, the value of $\tau_q(x|r_0)$ will increase indefinitely with x and r_0. For real (finite) records, there exists a maximum return interval which limits the values of x, r_0 and $\tau_q(x|r_0)$.

Distribution of Extreme Events

In this section we describe how the presence of fractal long-term correlations affects the statistics of the extreme events, i. e., maxima within time segments of fixed duration R, see Fig. 11 for illustration. By definition, extreme events are rare occurrences of extraordinary nature, such as, e. g. floods, very high temperatures, or earthquakes. In hydrological engineering such conventional extreme value statistics are commonly applied to decide what building projects are required to protect river-side areas against typical floods that occur, for example, once in 100 years. Most of these results are based on statistically independent values x_i and hold only in the limit $R \to \infty$. However, both of these assumptions are not strictly fulfilled for correlated fractal scaling data.

In classical extreme value statistics one assumes that records (x_i) consist of i.i.d. data, described by density distributions $P(x)$, which can be, e. g., a Gaussian or an exponential distribution. One is interested in the distribution density function $P_R(m)$ of the maxima (m_j) determined in segments of length R in the original series (x_i), see Fig. 11. Note that all maxima are also elements of the original data.
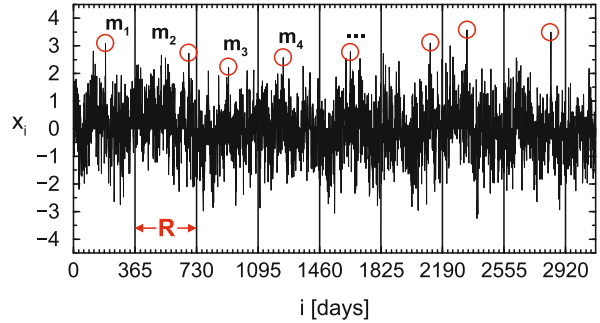
Fractal and Multifractal Time Series, Figure 11
Illustration for the definition of maxima m_R within periods of $R = 365$ values (figure by Jan Eichner)

The corresponding integrated maxima distribution $G_R(m)$ is defined as

$$G_R(m) = 1 - E_R(m) = \int_{-\infty}^m P_R(m')\mathrm{d}m'. \quad (39)$$

Since $G_R(m)$ is the probability of finding a maximum smaller than m, $E_R(m)$ denotes the probability of finding a maximum that exceeds m. One of the main results of traditional extreme value statistics states that for independently and identically distributed (i.i.d.) data (x_i) with Gaussian or exponential distribution density function $P(x)$ the integrated distribution $G_R(m)$ converges to a double exponential (Fisher–Tippet–Gumbel) distribution (often labeled as Type I) [86,87,88,89,90], i. e.,

$$G_R(m) \to G\left(\frac{m-u}{\alpha}\right) = \exp\left[-\exp\left(-\frac{m-u}{\alpha}\right)\right] \quad (40)$$

for $R \to \infty$, where α is the scale parameter and u the location parameter. By the method of moments those parameters are given by $\alpha = \sqrt{6}/\pi \, \sigma_R$ and $u = m_R - n_e\alpha$ with the Euler constant $n_e = 0.577216$ [89,91,92,93]. Here m_R and σ_R denote the (R-dependent) mean maximum and the standard deviation, respectively. Note that different asymptotics will be reached for broader distributions of data (x_i) that belong to other domains of attraction [89]. For example, for data following a power-law distribution (or Pareto distribution), $P(x) = (x/x_0)^{-k}$, $G_R(m)$ converges to a Fréchet distribution, often labeled as Type II. For data following a distribution with finite upper endpoint, for example the uniform distribution $P(x) = 1$ for $0 \le x \le 1$, $G_R(m)$ converges to a Weibull distribution, often labeled as Type III. We do not consider the latter two types of asymptotics here.

Numerical studies of fractal model data have recently shown that the distribution $P(x)$ of the original data has a much stronger effect upon the convergence towards the

Gumbel distribution than the long-term correlations in the data. Long-term correlations just slightly delay the convergence of $G_R(m)$ towards the Gumbel distribution (40). This can be observed very clearly in a plot of the integrated and scaled distribution $G_R(m)$ on logarithmic scale [81].

Furthermore, it was found numerically that (i) the maxima series (m_j) exhibit long-term correlations similar to those of the original data (x_i), and most notably (ii) the maxima distribution as well as the mean maxima significantly depend on the history, in particular on the previous maximum [81]. The last item implies that conditional mean maxima and conditional maxima distributions should be considered for improved extreme event predictions.

Simple Models for Fractal and Multifractal Time Series

Fourier Filtering

Fractal scaling with long-term correlations can be introduced most easily into time series by the Fourier-filtering technique, see, e. g., [94,95,96]. The Fourier filtering technique is not limited to the generation of long-term correlated data characterized by a power-law auto-correlation function $C(s) \sim x^{-\gamma}$ with $0 < \gamma < 1$. All values of the scaling exponents $\alpha = h(2) \approx H$ or $\beta = 2\alpha - 1$ can be obtained, even those that cannot be found directly by the fractal analysis techniques described in Sect. "Methods for Stationary Fractal Time Series Analysis" and Sect. "Methods for Non-stationary Fractal Time Series Analysis" (e. g. $\alpha < 0$). Note, however, that Fourier filtering will always yield Gaussian distributed data values and that no nonlinear or multifractal properties can be achieved (see also Subsect. "Multifractal Time Series", Subsect. "Sign and Magnitude (Volatility) DFA", and Sect. "Methods for Multifractal Time Series Analysis"). In Subsect. "Detection of Trends and Crossovers with DFA", we have briefly described a modification of Fourier filtering for obtaining reliable short-term correlated data.

For the generation of data characterized by fractal scaling with $\beta = 2\alpha - 1$ [94,95] we start with uncorrelated Gaussian distributed random numbers x_i from an i.i.d. generator. Transforming a series of such numbers into frequency space with discrete Fourier transform or FFT (fast Fourier transform, for suitable series lengths N) yields a flat power spectrum, since random numbers correspond to white noise. Multiplying the (complex) Fourier coefficients by $f^{-\beta/2}$, where $f \propto 1/s$ is the frequency, will rescale the power spectrum $S(f)$ to follow Eq. (6), as expected for time series with fractal scaling. After transform-

ing back to the time domain (using inverse Fourier transform or inverse FFT) we will thus obtain the desired long-term correlated data $\tilde{x}_i$. The final step is the normalization of this data.

The Fourier filtering method can be improved using modified Bessel functions instead of the simple factors $f^{-\beta/2}$ in modifying the Fourier coefficients [96]. This way problems with the divergence of the autocorrelation function $C(s)$ at $s = 0$ can be avoided.

An alternative method to the Fourier filtering technique, the random midpoint displacement method, is based on the construction of self-affine surfaces by an iterative procedure, see, e. g. [6]. Starting with one interval with constant values, the intervals are iterative split in the middle and the midpoint is displaced by a random offset. The amplitude of this offset is scaled according to the length of the interval. Since the method generates a self-affine surface x_i characterized by a Hurst exponent H, the differentiated series Δx_i can be used as long-term correlated or anti-correlated random numbers. Note, however, that the correlations do not persist for the whole length of the data generated this way. Another option is the use of wavelet synthesis, the reverse of wavelet analysis described in Subsect. "Wavelet Analysis". In that method, the scaling law is introduced by setting the magnitudes of the wavelet coefficients according to the corresponding time scale s.

The Schmitz–Schreiber Method

When long-term correlations in random numbers are introduced by the Fourier-filtering technique (see previous section), the original distribution $P(x)$ of the time series values x_i is always modified such that it becomes closer to a Gaussian. Hence, no series (x_i) with broad distributions of the values *and* fractal scaling can be generated. In these cases an iterative algorithm introduced by Schreiber and Schmitz [98,99] must be applied.

The algorithm consists of the following steps: First one creates a Gaussian distributed long-term correlated data set with the desired correlation exponent γ by standard Fourier-filtering [96]. The power spectrum $S_G(f) = F_G(f)F_G^*(f)$ of this data set is considered as reference spectrum (where f denotes the frequency in Fourier space and the $F_G(f)$ are the complex Fourier coefficients). Next one creates an uncorrelated sequence of random numbers (x_i^{ref}), following a desired distribution $P(x)$. The (complex) Fourier transform $F(f)$ of the (x_i^{ref}) is now divided by its absolute value and multiplied by the square root of the reference spectrum,

$$F_{\text{new}}(f) = \frac{F(f)\sqrt{S_G(f)}}{|F(f)|} \, . \tag{41}$$

After the Fourier back-transformation of $F_{new}(f)$, the new sequence (x_i^{new}) has the desired correlations (i. e. the desired γ), but the shape of the distribution has changed towards a (more or less) Gaussian distribution. In order to enforce the desired distribution, we exchange the (x_i^{new}) by the (x_i^{ref}), such that the largest value of the new set is replaced by the largest value of the reference set, the second largest of the new set by the second largest of the reference set and so on. After this the new sequence has the desired distribution and is clearly correlated. However, due to the exchange algorithm the perfect long-term correlations of the new data sequence were slightly altered again. So the procedure is repeated: the new sequence is Fourier transformed followed by spectrum adjustment, and the exchange algorithm is applied to the Fourier back-transformed data set. These steps are repeated several times, until the desired quality (or the best possible quality) of the spectrum of the new data series is achieved.

The Extended Binomial Multifractal Model

The multifractal cascade model [6,33,65] is a standard model for multifractal data, which is often applied, e. g., in hydrology [97]. In the model, a record x_i of length $N = 2^{n_{max}}$ is constructed recursively as follows. In generation $n = 0$, the record elements are constant, i. e. $x_i = 1$ for all $i = 1, \ldots, N$. In the first step of the cascade (generation $n = 1$), the first half of the series is multiplied by a factor a and the second half of the series is multiplied by a factor b. This yields $x_i = a$ for $i = 1, \ldots, N/2$ and $x_i = b$ for $i = N/2 + 1, \ldots, N$. The parameters a and b are between zero and one, $0 < a < b < 1$. One need not restrict the model to $b = 1 - a$ as is often done in the literature [6]. In the second step (generation $n = 2$), we apply the process of step 1 to the two subseries, yielding $x_i = a^2$ for $i = 1, \ldots, N/4$, $x_i = ab$ for $i = N/4 + 1, \ldots, N/2$, $x_i = ba = ab$ for $i = N/2 + 1, \ldots, 3N/4$, and $x_i = b^2$ for $i = 3N/4 + 1, \ldots, N$. In general, in step $n + 1$, each subseries of step n is divided into two subseries of equal length, and the first half of the x_i is multiplied by a while the second half is multiplied by b. For example, in generation $n = 3$ the values in the eight subseries are $a^3, a^2 b, a^2 b, ab^2, a^2 b, ab^2, ab^2, b^3$. After n_{max} steps, the final generation has been reached, where all subseries have length 1 and no more splitting is possible. We note that the final record can be written as $x_i = a^{n_{max} - n(i-1)} b^{n(i-1)}$, where $n(i)$ is the number of digits 1 in the binary representation of the index i, e. g. $n(13) = 3$, since 13 corresponds to binary 1101.

For this multiplicative cascade model, the formula for $\tau(q)$ has been derived earlier [6,33,65]. The result is

$\tau(q) = [-\ln(a^q + b^q) + q \ln(a + b)]/\ln 2$ or

$$h(q) = \frac{1}{q} - \frac{\ln(a^q + b^q)}{q \ln 2} + \frac{\ln(a + b)}{\ln 2} . \quad (42)$$

It is easy to see that $h(1) = 1$ for all values of a and b. Thus, in this form the model is limited to cases where $h(1)$, which is the exponent Hurst defined originally in the R/S method, is equal to one.

In order to generalize this multifractal cascade process such that any value of $h(1)$ is possible, one can subtract the offset $\Delta h = \ln(a + b)/\ln(2)$ from $h(q)$ [100]. The constant offset Δh corresponds to additional long-term correlations incorporated in the multiplicative cascade model. For generating records without this offset, we rescale the power spectrum. First, we transform (FFT) the simple multiplicative cascade data into the frequency domain. Then, we multiply all Fourier coefficients by $f^{-\Delta h}$, where f is the frequency. This way, the slope β of the power spectra $S(f) \sim f^{-\beta}$ is decreased from $\beta = 2h(2) - 1 = [2 \ln(a + b) - \ln(a^2 + b^2)]/\ln 2$ into $\beta' = 2[h(2) - \Delta h] - 1 = -\ln(a^2 + b^2)/\ln 2$. Finally, backward FFT is employed to transform the signal back into the time domain.

The Bi-fractal Model

In some cases a simple bi-fractal model is already sufficient for modeling apparently multifractal data [101]. For bi-fractal records the Renyi exponents $\tau(q)$ are characterized by two distinct slopes α_1 and α_2,

$$\tau(q) = \begin{cases} q\alpha_1 - 1 & q \le q_\times \\ q\alpha_2 + q_\times(\alpha_1 - \alpha_2) - 1 & q > q_\times \end{cases} \quad (43)$$

or

$$\tau(q) = \begin{cases} q\alpha_1 + q_\times(\alpha_2 - \alpha_1) - 1 & q \le q_\times \\ q\alpha_2 - 1 & q > q_\times \end{cases} . \quad (44)$$

If this behavior is translated into the $h(q)$ picture using Eq. (29), we obtain that $h(q)$ exhibits a plateau from $q = -\infty$ up to a certain $q_\times$ and decays hyperbolically for $q > q_\times$,

$$h(q) = \begin{cases} \alpha_1 & q \le q_\times \\ q_\times(\alpha_1 - \alpha_2)\frac{1}{q} + \alpha_2 & q > q_\times \end{cases} , \quad (45)$$

or vice versa,

$$h(q) = \begin{cases} q_\times(\alpha_2 - \alpha_1)\frac{1}{q} + \alpha_1 & q \le q_\times \\ \alpha_2 & q > q_\times \end{cases} . \quad (46)$$

Both versions of this bi-fractal model require three parameters. The multifractal spectrum is degenerated to two single points, thus its width can be defined as $\Delta\alpha = \alpha_1 - \alpha_2$.

Future Directions

The most straightforward future direction is to analyze more types of time series from other complex systems than those listed in Sect. "Introduction" to check for the presence of fractal scaling and in particular long-term correlations. Such applications may include (i) data that are not recorded as a function of time but as a function of another parameter and (ii) higher dimensional data. In particular, the inter-relationship between fractal time series and spatially fractal structures can be studied. Studies of *fields* with fractal scaling in time and space have already been performed in Geophysics. In some cases studying new types of data will require dealing with more difficult types of non-stationarities and transient behavior, making further development of the methods necessary. In many studies, detrending methods have not been applied yet. However, discovering fractal scaling in more and more systems cannot be an aim on its own.

Up to now, the reasons for observed fractal or multifractal scaling are not clear in most applications. It is thus highly desirable to study causes for fractal and multifractal correlations in time series, which is a difficult task, of course. One approach might be based on modeling and comparing the fractal aspects of real and modeled time series by applying the methods described in this article. The fractal or multifractal characterization can thus be helpful in improving the models. For many applications, practically usable models which display fractal or transient fractal scaling still have to be developed. One example for a model explaining fractal scaling might be a precipitation, storage and runoff model, in which the fractal scaling of runoff time series could be explained by fractional integration of rainfall in soil, groundwater reservoirs, or river networks characterized by a fractal structure. Also studies regarding the inter-relationship between fractal scaling and complex networks, representing the structure of a complex system, are desirable. This way one could gain an interpretation of the causes for fractal behavior.

Another direction of future research is regarding the linear and especially non-linear inter-relationships between several time series. There is great need for improved methods characterizing cross-correlations and similar statistical inter-relationships between several non-stationary time series. Most methods available so far are reserved to stationary data, which is, however, hardly found in natural recordings. An even more ambitious aim is the (time-dependent) characterization of a larger network of signals. In such a network, the signals themselves would represent the nodes, while the (possibly directed) inter-relationships between each pair represent the links (or bonds) between the nodes. The properties of both nodes and links can vary with time or change abruptly, when the represented complex system goes through a phase transition.

Finally, more work will have to be invested in studying the practical consequences of fractal scaling in time series. Studies should particularly focus on predictions of future values and behavior of time series and whole complex systems. This is very relevant, not only in hydrology and climate research, where a clear distinguishing of trends and natural fluctuations is crucial, but also for predicting dangerous medical events on-line in patients based on the continuous recording of time series.

Acknowledgment

Bibliography

1. Mandelbrot BB, van Ness JW (1968) Fractional Brownian motions, fractional noises and applications. SIAM Review 10: 422
2. Mandelbrot BB, Wallis JR (1969) Some long-run properties of geophysical records. Water Resour Res 5:321–340
3. Mandelbrot BB (1999) Multifractals and 1/f noise: wild self-affinity in physics. Springer, Berlin
4. Hurst HE (1951) Long-term storage capacity of reservoirs. Tran Amer Soc Civ Eng 116:770
5. Hurst HE, Black RP, Simaika YM (1965) Long-term storage: an experimental study. Constable, London
6. Feder J (1988) Fractals. Plenum Press, New York
7. Barnsley MF (1993) Fractals everywhere. Academic Press, San Diego
8. Bunde A, Havlin S (1994) Fractals in science. Springer, Berlin
9. Jorgenssen PET (2000) Analysis and probability: Wavelets, signals, fractals. Springer, Berlin
10. Bunde A, Kropp J, Schellnhuber HJ (2002) The science of disasters – climate disruptions, heart attacks, and market crashes. Springer, Berlin
11. Kantz H, Schreiber T (2003) Nonlinear time series analysis. Cambridge University Press, Cambridge
12. Peitgen HO, Jürgens H, Saupe D (2004) Chaos and fractals. Springer, Berlin
13. Sornette D (2004) Critical phenomena in natural sciences. Springer, Berlin
14. Peng CK, Mietus J, Hausdorff JM, Havlin S, Stanley HE, Goldberger AL (1993) Long-range anti-correlations and non-Gaussian behaviour of the heartbeat. Phys Rev Lett 70: 1343

15. Bunde A, Havlin S, Kantelhardt JW, Penzel T, Peter JH, Voigt K (2000) Correlated and uncorrelated regions in heart-rate fluctuations during sleep. Phys Rev Lett 85:3736

16. Vyushin D, Zhidkov I, Havlin S, Bunde A, Brenner S (2004) Volcanic forcing improves atmosphere-ocean coupled general circulation model scaling performance. Geophys Res Lett 31:L10206

17. Koscielny-Bunde E, Bunde A, Havlin S, Roman HE, Goldreich Y, Schellnhuber HJ (1998) Indication of a universal persistence law governing atmospheric variability. Phys Rev Lett 81:729

18. Box GEP, Jenkins GM, Reinsel GC (1994) Time-series analysis. Prentice Hall, New Jersey

19. Chatfield C (2003) The analysis of time series. An introduction. Taylor & Francis, Boca Raton

20. Schmitt DT, Schulz M (2006) Analyzing memory effects of complex systems from time series. Phys Rev E 73:056204

21. Taqqu MS, Teverovsky V, Willinger W (1995) Estimators for long-range dependence: An empirical study. Fractals 3:785

22. Delignieresa D, Ramdania S, Lemoinea L, Torrea K, Fortesb M, Ninot G (2006) Fractal analyses for 'short' time series: A reassessment of classical methods. J Math Psychol 50:525

23. Mielniczuk J, Wojdyllo P (2007) Estimation of Hurst exponent revisited. Comp Stat Data Anal 51:4510

24. Hunt GA (1951) Random Fourier transforms. Trans Amer Math Soc 71:38

25. Rangarajan G, Ding M (2000) Integrated approach to the assessment of long range correlation in time series data. Phys Rev E 61:4991

26. Peng CK, Buldyrev SV, Goldberger AL, Havlin S, Sciortino F, Simons M, Stanley HE (1992) Long-range correlations in nucleotide sequences. Nature 356:168

27. Goupillaud P, Grossmann A, Morlet J (1984) Cycle-octave and related transforms in seismic signal analysis. Geoexploration 23:85

28. Daubechies I (1988) Orthogonal bases of compactly supported wavelets. Commun Pure Appl Math 41:909

29. Bogachev M, Schumann AY, Kantelhardt JW, Bunde A (2009) On distinguishing long-term and short-term memory in finite data. Physica A, to be published

30. Kantelhardt JW, Roman HE, Greiner M (1995) Discrete wavelet approach to multifractality. Physica A 220:219

31. Peng C-K, Buldyrev SV, Havlin S, Simons M, Stanley HE, Goldberger AL (1994) Mosaic organization of DNA nucleotides. Phys Rev E 49:1685

32. Ashkenazy Y, Ivanov PC, Havlin S, Peng CK, Goldberger AL, Stanley HE (2001) Magnitude and sign correlations in heartbeat fluctuations. Phys Rev Lett 86:1900

33. Kantelhardt JW, Zschiegner SA, Bunde A, Havlin S, Koscielny-Bunde E, Stanley HE (2002) Multifractal detrended fluctuation analysis of non-stationary time series. Physica A 316:87

34. Gu GF, Zhou WX (2006) Detrended fluctuation analysis for fractals and multifractals in higher dimensions. Phys Rev E 74:061104

35. Kantelhardt JW, Koscielny-Bunde E, Rego HHA, Havlin S, Bunde A (2001) Detecting long-range correlations with detrended fluctuation analysis. Physica A 295:441

36. Hu K, Ivanov PC, Chen Z, Carpena P, Stanley HE (2001) Effect of trends on detrended fluctuation analysis. Phys Rev E 64:011114

37. Chen Z, Ivanov PC, Hu K, Stanley HE (2002) Effect of nonstationarities on detrended fluctuation analysis. Phys Rev E 65:041107

38. Chen Z, Hu K, Carpena P, Bernaola-Galvan P, Stanley HE, Ivanov PC (2005) Effect of nonlinear filters on detrended fluctuation analysis. Phys Rev E 71:011104

39. Grau-Carles P (2006) Bootstrap testing for detrended fluctuation analysis. Physics A 360:89

40. Nagarajan R (2006) Effect of coarse-graining on detrended fluctuation analysis. Physica A 363:226

41. Heneghan C, McDarby G (2000) Establishing the relation between detrended fluctuation analysis and power spectral density analysis for stochastic processes. Phys Rev E 62:6103

42. Weron R (2002) Estimating long-range dependence: finite sample properties and confidence intervals. Physica A 312:285

43. Bashan A, Bartsch R, Kantelhardt JW, Havlin S (2008) Comparison of detrending methods for fluctuation analysis. Physica A 387:580

44. Bahar S, Kantelhardt JW, Neiman A, Rego HHA, Russell DF, Wilkens L, Bunde A, Moss F (2001) Long range temporal anticorrelations in paddlefish electro-receptors. Europhys Lett 56:454

45. Bartsch R, Henning T, Heinen A, Heinrichs S, Maass P (2005) Statistical analysis of fluctuations in the ECG morphology. Physica A 354:415

46. Santhanam MS, Bandyopadhyay JN, Angom D (2006) Quantum spectrum as a time series: fluctuation measures. Phys Rev E 73:015201

47. Ashkenazy Y, Havlin S, Ivanov PC, Peng CK, Schulte-Frohlinde V, Stanley HE (2003) Magnitude and sign scaling in power-law correlated time series. Physica A 323:19

48. Kalisky T, Ashkenazy Y, Havlin S (2005) Volatility of linear and nonlinear time series. Phys Rev E 72:011913

49. Mantegna RN, Stanley HE (2000) An introduction to econophysics – correlations and complexity in finance. Cambridge Univ Press, Cambridge

50. Bouchaud JP, Potters M (2003) Theory of financial risks: from statistical physics to risk management. Cambridge Univ Press, Cambridge

51. Alessio E, Carbone A, Castelli G, Frappietro V (2002) Second-order moving average and scaling of stochastic time series. Europ Phys J B 27:197

52. Carbone A, Castelli G, Stanley HE (2004) Analysis of clusters formed by the moving average of a long-range correlated time series. Phys Rev E 69:026105

53. Carbone A, Castelli G, Stanley HE (2004) Time-dependent Hurst exponent in financial time series. Physica A 344:267

54. Alvarez-Ramirez J, Rodriguez E, Echeverría JC (2005) Detrending fluctuation analysis based on moving average filtering. Physica A 354:199

55. Kiyono K, Struzik ZR, Aoyagi N, Togo F, Yamamoto Y (2005) Phase transition in a healthy human heart rate. Phys Rev Lett 95:058101

56. Staudacher M, Telser S, Amann A, Hinterhuber H, Ritsch-Marte M (2005) A new method for change-point detection developed for on-line analysis of the heart beat variability during sleep. Physica A349:582

57. Telser S, Staudacher M, Hennig B, Ploner Y, Amann A, Hinter-

huber H, Ritsch-Marte M (2007) Temporally resolved fluctuation analysis of sleep-ECG. J Biol Phys 33:190

58. Chianca CV, Ticona A, Penna TJP (2005) Fourier-detrended fluctuation analysis. Physica A 357:447

59. Jánosi IM, Müller R (2005) Empirical mode decomposition and correlation properties of long daily ozone records. Phys Rev E 71:056126

60. Nagarajan R, Kavasseri RG (2005) Minimizing the effect of trends on detrended fluctuation analysis of long-range correlated noise. Physica A 354:182

61. Nagarajan R (2006) Reliable scaling exponent estimation of long-range correlated noise in the presence of random spikes. Physica A 366:1

62. Rodriguez E, Echeverria JC, Alvarez-Ramirez J (2007) Detrending fluctuation analysis based on high-pass filtering. Physica A 375:699

63. Grech D, Mazur Z (2005) Statistical properties of old and new techniques in detrended analysis of time series. Acta Phys Pol B 36:2403

64. Xu L, Ivanov PC, Hu K, Chen Z, Carbone A, Stanley HE (2005) Quantifying signals with power-law correlations: a comparative study of detrended fluctuation analysis and detrended moving average techniques. Phys Rev E 71:051101

65. Barabási AL, Vicsek T (1991) Multifractality of self-affine fractals. Phys Rev A 44:2730

66. Bacry E, Delour J, Muzy JF (2001) Multifractal random walk. Phys Rev E 64:026103

67. Muzy JF, Bacry E, Arneodo A (1991) Wavelets and multifractal formalism for singular signals: Application to turbulence data. Phys Rev Lett 67:3515

68. Muzy JF, Bacry E, Arneodo A (1994) The multifractal formalism revisited with wavelets. Int J Bifurcat Chaos 4:245

69. Arneodo A, Bacry E, Graves PV, Muzy JF (1995) Characterizing long-range correlations in DNA sequences from wavelet analysis. Phys Rev Lett 74:3293

70. Arneodo A, Manneville S, Muzy JF (1998) Towards log-normal statistics in high Reynolds number turbulence. Eur Phys J B 1:129

71. Arneodo A, Audit B, Decoster N, Muzy JF, Vaillant C (2002) Wavelet based multifractal formalism: applications to DNA sequences, satellite images of the cloud structure, and stock market data. In: Bunde A, Kropp J, Schellnhuber HJ (eds) The science of disaster: climate disruptions, market crashes, and heart attacks. Springer, Berlin

72. Kantelhardt JW, Rybski D, Zschiegner SA, Braun P, Koscielny-Bunde E, Livina V, Havlin S, Bunde A (2003) Multifractality of river runoff and precipitation: comparison of fluctuation analysis and wavelet methods. Physica A 330:240

73. Oswiecimka P, Kwapien J, Drozdz S (2006) Wavelet versus detrended fluctuation analysis of multifractal structures. Phys Rev E 74:016103

74. Ivanov PC, Amaral LAN, Goldberger AL, Havlin S, Rosenblum MG, Struzik ZR, Stanley HE (1999) Multifractality in human heartbeat dynamics. Nature 399:461

75. Amaral LAN, Ivanov PC, Aoyagi N, Hidaka I, Tomono S, Goldberger AL, Stanley HE, Yamamoto Y (2001) Behavioral-independence features of complex heartbeat dynamics. Phys Rev Lett 86:6026

76. Bogachev M, Schumann AY, Kantelhardt JW (2008) (in preparation)

77. Bunde A, Eichner JF, Kantelhardt JW, Havlin S (2005) Long-term memory: A natural mechanism for the clustering of extreme events and anomalous residual times in climate records. Phys Rev Lett 94:048701

78. Bunde A, Eichner JF, Kantelhardt JW, Havlin S (2003) The effect of long-term correlations on the return periods of rare events. Physica A 330:1

79. Altmann EG, Kantz H (2005) Recurrence time analysis, long-term correlations, and extreme events. Phys Rev E 71:056106

80. Eichner JF, Kantelhardt JW, Bunde A, Havlin S (2007) Statistics of return intervals in long-term correlated records. Phys Rev E 75:011128

81. Eichner JF, Kantelhardt JW, Bunde A, Havlin S (2006) Extreme value statistics in records with long-term persistence. Phys Rev E 73:016130

82. Bogachev MI, Eichner JF, Bunde A (2007) Effect of nonlinear correlations on the statistics of return intervals in multifractal data sets. Phys Rev Lett 99:240601

83. Storch HV, Zwiers FW (2001) Statistical analysis in climate research. Cambridge Univ Press, Cambridge

84. Newell GF, Rosenblatt M (1962) Ann Math Statist 33:1306

85. Sornette D, Knopoff L (1997) The paradox of the expected time until the next earthquake. Bull Seism Soc Am 87:789

86. Fisher RA, Tippett LHC (1928) Limiting forms of the frequency distribution of the largest or smallest member of a sample. Proc Camb Phi Soc 24:180

87. Gumbel EJ (1958) Statistics of extremes. Columbia University Press, New York

88. Galambos J (1978) The asymptotic theory of extreme order statistics. Wiley, New York

89. Leadbetter MR, Lindgren G, Rootzen H (1983) Extremes and related properties of random sequences and processes. Springer, New York

90. Galambos J, Lechner J, Simin E (1994) Extreme value theory and applications. Kluwer, Dordrecht

91. te Chow V (1964) Handbook of applied hydrology. McGraw-Hill, New York

92. Raudkivi AJ (1979) Hydrology. Pergamon Press, Oxford

93. Rasmussen PF, Gautam N (2003) Alternative PWM-estimators of the Gumbel distribution. J Hydrol 280:265

94. Mandelbrot BB (1971) A fast fractional Gaussian noise generator. Water Resour Res 7:543

95. Voss RF (1985) In: Earnshaw RA (ed) Fundamental algorithms in computer graphics. Springer, Berlin

96. Makse HA, Havlin S, Schwartz M, Stanley HE (1996) Method for generating long-range correlations for large systems. Phys Rev E 53:5445

97. Rodriguez-Iturbe I, Rinaldo A (1997) Fractal river basins – change and self-organization. Cambridge Univ Press, Cambridge

98. Schreiber T, Schmitz A (1996) Improved surrogate data for nonlinearity tests. Phys Rev Lett 77:635

99. Schreiber T, Schmitz A (2000) Surrogate time series. Physica D 142:346

100. Koscielny-Bunde E, Kantelhardt JW, Braun P, Bunde A, Havlin S (2006) Long-term persistence and multifractality of river runoff records. J Hydrol 322:120

101. Kantelhardt JW, Koscielny-Bunde E, Rybski D, Braun P, Bunde A, Havlin S (2006) Long-term persistence and multifractality of precipitation and river runoff records. J Geophys Res Atmosph 111:D01106

Fractals in Biology

SERGEY V. BULDYREV
Department of Physics, Yeshiva University,
New York, USA

Article Outline

Glossary

Allometric laws An allometric law describes the relationship between two attributes of living organisms y and x, and is usually expressed as a power-law: $y \sim x^\alpha$, where α is the scaling exponent of the law. For example, x can represent total body mass M and y can represent the mass of a brain m_b. In this case $m_b \sim M^{3/4}$. Another example of an allometric law: $B \sim M^{3/4}$ where B is metabolic rate and M is body mass. Allometric laws can be also found in ecology: the number of different species N found in a habitat of area A scales as $N \sim A^{1/4}$.

Radial distribution function Radial distribution function $g(r)$ describes how the average density of points of a set behaves as function of distance r from a point of this set. For an empirical set of N data points, the distances between all pair of points are computed and the number of pairs $N_p(r)$ such that their distance is less then r is found. Then $M(r) = 2N_p(r)/N$ gives the average number of the neighbors (mass) of the set within a distance r. For a certain distance bin $r_1 < r < r_2$, we define $g[(r_2 + r_1)/2] = [M(r_2) - M(r_1)]/[V_d(r_2) - V_d(r_1)]$, where

$V_d(r) = 2\pi^{d/2} r^d/[d\,\Gamma(d/2)]$ is the volume/area/length of a d-dimensional sphere/circle/interval of radius r.

Fractal set We define a fractal set with the fractal dimension $0 < d_f < d$ as a set for which $M(r) \sim r^{d_f}$ for $r \to \infty$. Accordingly, for such a set $g(r)$ decreases as a power law of the distance $g(r) \sim r^{-\chi}$, where $\chi = d - d_f$.

Correlation function For a superposition of a fractal set and a set with a finite density defined as $\rho = \lim_{r \to \infty} M(r)/V_d(r)$, the correlation function is defined as $h(r) \equiv g(r) - \rho$.

Long-range power law correlations The set of points has long-range power law correlations (LRPLC) if $h(r) \sim r^{-\chi}$ for $r \to \infty$ with $0 < \chi < d$. LRPLC indicate the presence of a fractal set with fractal dimension $d_f = d - \chi$ superposed with a uniform set.

Critical point Critical point is defined as a point in the system parameter space (e. g. temperature, $T = T_c$, and pressure $P = P_c$), near which the system acquires LRPLC

$$h(r) \sim \frac{1}{r^{d-2+\eta}} \exp(r/\xi)\,, \tag{1}$$

where ξ is the correlation length which diverges near the critical point as $\sim |T - T_c|^{-\nu}$. Here $\eta > 0$ and $\nu > 0$ are critical exponents which depend on the few system characteristics such as dimensionality of space. Accordingly, the system is characterized by fractal density fluctuations with $d_f = 2 - \eta$.

Self-organized criticality Self-organized criticality (SOC) is a term which describes a system for which the critical behavior characterized by a large correlation length is achieved for a wide range of parameters and thus does not require special tuning. This usually occurs when a critical point corresponds to an infinite value of a system parameter, such as a ratio of the characteristic time of the stress build up and a characteristic time of the stress release.

Morphogenesis Morphogenesis is a branch of developmental biology concerned with the shapes of organs and the entire organisms. Several types of molecules are particularly important during morphogenesis. Morphogens are soluble molecules that can diffuse and carry signals that control cell differentiation decisions in a concentration-dependent fashion. Morphogens typically act through binding to specific protein receptors. An important class of molecules involved in morphogenesis are transcription factor proteins that determine the fate of cells by interacting with DNA. The morphogenesis of the branching fractal-like structures such as lungs involves a dozen of morpho-

genes. The mechanism for keeping self-similarity of the branches at different levels of branching hierarchy is not yet fully understood. The experiments with transgenic mice with certain genes knocked-out produce mice without limbs and lungs or without terminal buds.

Definition of the Subject

Fractals occur in a wide range of biological applications:

1) In morphology when the shape of an organism (tree) or an organ (vertebrate lung) has a self-similar brunching structure which can be approximated by a fractal set (Sect. "Self-Similar Branching Structures").
2) In allometry when the allometric power laws can be deduced from the fractal nature of the circulatory system (Sect. "Fractal Metabolic Rates").
3) In ecology when a colony or a habitat acquire fractal shapes due to some SOC processes such as diffusion limited aggregation (DLA) or percolation which describes forest fires (Sects. "Physical Models of Biological Fractals"–"Percolation and Forest Fires").
4) In epidemiology when some of the features of the epidemics is described by percolation which in turn leads to fractal behavior (Sect. "Percolation and Forest Fires").
5) In behavioral sciences, when a trajectory of foraging animal acquires fractal features (Sect. "Lévy Flight Foraging").
6) In population dynamics, when the population size fluctuates chaotically (Sect. "Dynamic Fractals").
7) In physiology, when time series have LRPLC (Sect. "Fractals and Time Series").
8) In evolution theory, which may have some features described by SOC (Sect. "SOC and Biological Evolution").
9) In bioinformatics when a DNA sequence has a LRPLC or a network describing protein interactions has a self-similar fractal behavior (Sect. "Fractal Features of DNA Sequences").

Fractal geometry along with Euclidian geometry became a part of general culture which any scientist must be familiar with. Fractals often originate in the theory of complex systems describing the behavior of many interacting elements and therefore have a great number of biological applications. Complex systems have a general tendency for self-organization and complex pattern formation. Some of these patterns have certain nontrivial symmetries, for example fractals are characterized by scale invariance i. e. they look similarly on different magnification. Fractals are characterized by their fractal dimension, which is specific

for each model and therefore may shed light on the origin of a particular biological phenomenon. In Sect. "Diffusion Limited Aggregation and Bacterial Colonies", we discuss the techniques for measuring fractal dimension and their limitations.

Introduction

The fact that simple objects of Euclidian geometry such as straight lines, circles, cubes, and spheres are not sufficient to describe complex biological shapes has been known for centuries. Physicists were always accused by biologists for introducing a "spherical cow". Nevertheless people from antiquity to our days were fascinated by finding simple mathematical regularities which can describe the anatomy and physiology of leaving creatures. Five centuries ago, Leonardo da Vinci observed that "the branches of a tree at every stage of its height when put together are equal in thickness to the trunk" [107]. Another famous but still poorly understood phenomenon is the emergence of the Fibonacci numbers in certain types of pine cones and composite flowers [31,134].

In the middle of the seventies a new concept of fractal geometry was introduced by Mandelbrot [79]. This concept was readily accepted for analysis of the complex shapes in the biological world. However, after initial splash of enthusiasm, [14,40,41,74,75,76,88,122] the application of fractals in biology significantly dwindled and today the general consensus is that a "fractal cow" is often not much better than a "spherical cow". Nature is always more complex than mathematical abstractions.

Strictly speaking the fractal is an object whose mass, M, grows as a fractional power law of its linear dimension, L,

$$M \sim L^{d_f} , \tag{2}$$

where d_f is a non-integer quantity called fractal dimension. Simple examples of fractal objects of various fractal dimensions are given by iterative self-similar constructs such as a Cantor set, a Koch curve, a Sierpinski gasket, and a Menger sponge [94]. In all this constructs a next iteration of the object is created by arranging p exact copies of the previous iteration of the object (Fig. 1) in such a way that the linear size of the next iteration is q times larger than the linear size of the previous iteration. Thus the mass, M_n, and the length, L_n, of the nth iteration scale as

$$\begin{aligned} M_n &= p^n M_0 \\ L_n &= q^n L_0 , \end{aligned} \tag{3}$$

where M_0 and L_0 are mass and length of the zero order
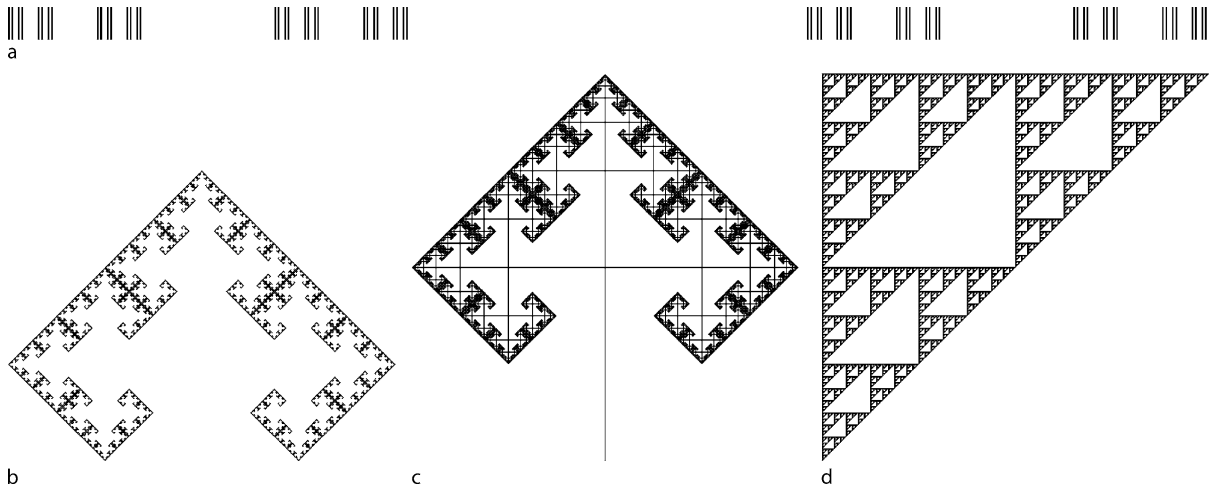
Fractals in Biology, Figure 1
a Cantor set ($p = 2$, $q = 3$, $d_f = \ln 2/\ln 3 \approx 0.63$) is an example of fractal "dust" with fractal dimension less then 1; **b** Fractal tree ($p = 3$, $q = 2$, $d_f = \ln 3/\ln 2 \approx 1.58$) with branches removed so that only the terminal points (leaves) can be seen. **c** The same tree with the trunks added. Both trees are produced by recursive combining of the three smaller trees: one serving as the top of the tree and the other two rotated by 90° clockwise and counter-clockwise serving as two brunches joined with the top at the middle. In **c** a vertical segment representing a trunk of the length equal to the diagonal of the branches is added at each recursive step. Mathematically, the fractal dimensions of sets **b** and **c** are the same, because the mass of the tree with trunks (number of black pixels) for the system of linear size 2^n grows as $3^{n+1} - 2^{n+1}$, while for the tree without trunks mass scales simply as 3^n. In the limit $n \to \infty$ this leads to the same fractal dimension $\ln 3/\ln 2$. However, visual inspection suggests that the tree with branches has a larger fractal dimension. This is in accord with the box counting method, which produces a higher value of the slope for the finite tree with the trunks. The slope slowly converges to the theoretical value as the number of recursive steps increases. **d** Sierpinski gasket has the same fractal dimension as the fractal tree ($p = 3$, $q = 2$) but has totally different topology and visual appearance than the tree

iteration. Excluding n from Eq. (3), we get

$$M_n = M_0 \left(\frac{L_n}{L_0} \right)^{d_f} , \qquad (4)$$

where

$$d_f = \ln p/\ln q \qquad (5)$$

can be identified as fractal dimension. The above-described objects have the property of self-similarity, the previous iteration magnified q times looks exactly like the next iteration once we neglect coarse-graining on the lowest iteration level, which can be assumed to be infinitely small. An interesting feature of such fractals is the power law distribution of their parts. For example, the cumulative distribution of distances L between the points of Cantor set and the branch lengths of the fractal tree (Fig. 1) follows a power law:

$$P(L > x) \sim x^{-d_f} . \qquad (6)$$

Thus the emergence of power law distributions and other power law dependencies is often associated with fractals and often these power law regularities not necessarily related to geometrical fractals are loosely referred as fractal properties.

Self-similar Branching Structures

Perfect deterministic fractals described above are never observed in nature, where all shapes are subjects to random variations. The natural examples closest to the deterministic fractals are structures of trees, certain plants such as a cauliflower, lungs and a cardiovascular system. The control of branching morphogenesis [85,114,138] involves determining when and where a branch will occur, how long the tube grows before branching again, and at what angle the branch will form. The development of different organs (such as salivary gland, mammary gland, kidney, and lung) creates branching patterns easily distinguished from each other. Moreover, during the development of a particular organ, the form of branching often changes, depending on the place or time when the branching occurs. Morphogenesis is controlled by complex molecular interactions of morphogenes.

Let us illustrate the idea of calculating fractal dimension of a branching object such as a cauliflower [51,64]. If we assume (which is not quite correct, see Fig. 2) that each branch of a cauliflower give rise to exactly p branches of the next generation exactly q times smaller than the original branch applying Eq. (5), we get $d_f = \ln q/\ln p$. Note
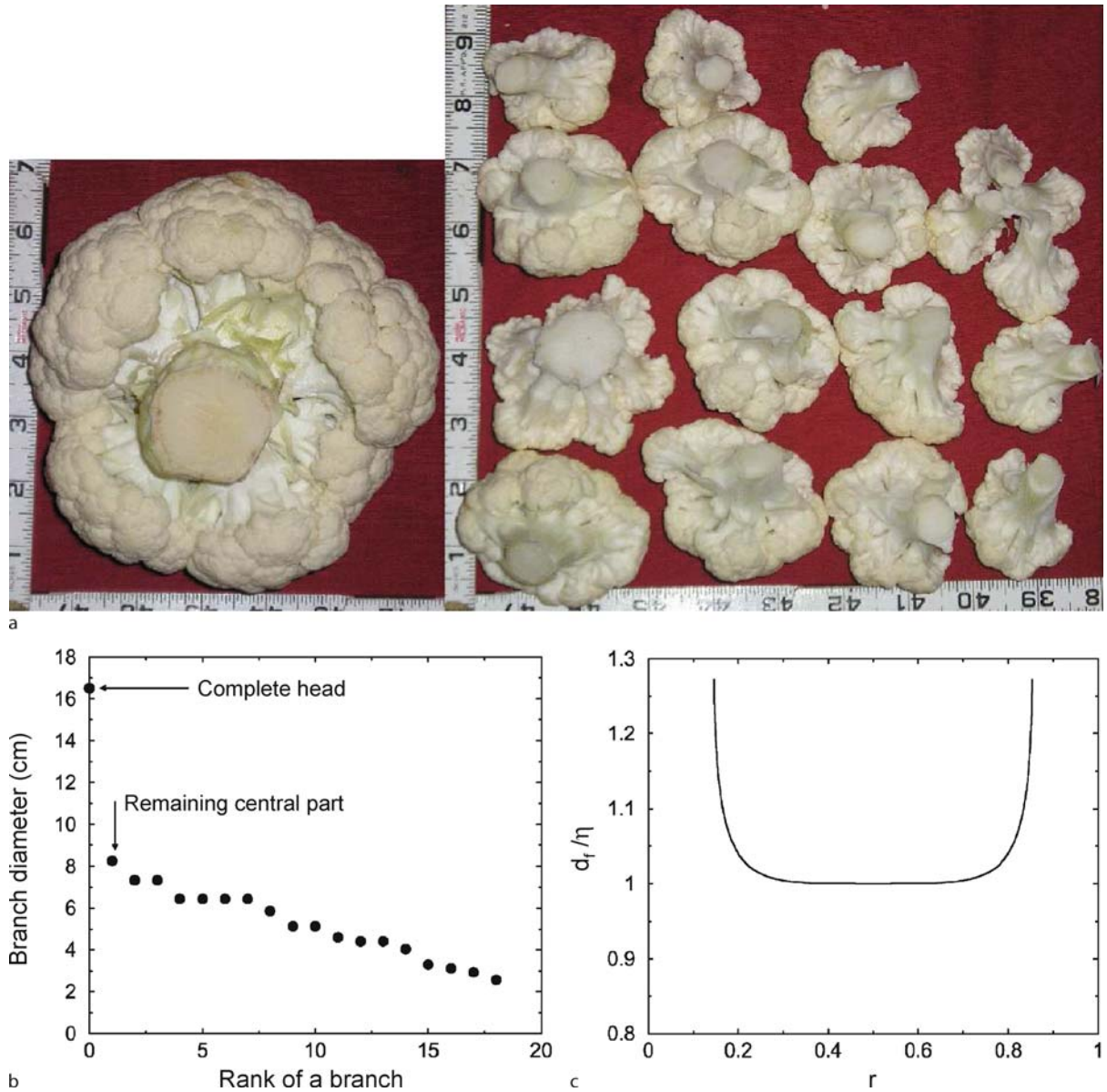
Fractals in Biology, Figure 2

a Cauliflower anatomy. A complete head of a cauliflower (*left*) and after it is taken apart (*right*). We remove 17 branches until the diameter of the remaining part becomes half of the diameter of the original head. **b** Diameters of the branches presented in **a**. **c** Estimation of the fractal dimension for asymmetric lungs ($\eta = 3$) and trees ($\eta = 2$) with branching parameter r using Eq. (7). For $r > 0.855$ or $r < 0.145$ the fractal dimension is not defined due to the "infrared catastrophe": the number of large branches diverge in the limit of an infinite tree

that this is not the fractal dimension of the cauliflower itself, but of its skeleton in which each brunch is represented by the elements of its daughter branches, with an addition of a straight line connecting the daughter branches as in the example of a fractal tree (Fig. 1b, c). Because this addition does not change the fractal dimension formula, the fractal dimension of the skeleton is equal to the fractal dimension of the surface of a cauliflower, which can be represented as the set of the terminal branches. As a physical object, the cauliflower is not a fractal but a three-dimensional body so that the mass of a branch of length L scales as L^3. This is because the diameter of each branch is proportional to the length of the branch.

The distribution of lengths of the collection of all the branches of a cauliflower is also a power law given by a simple idea that there are $N_n = \sum_{k=0}^{n-1} p^k \approx p^n/(p-1)$ branches of length larger than $L_n = L_0 q^{-n}$. Excluding n gives $N_n(L > L_n) \sim L_n^{-d_f}$. In reality, however (Fig. 2a, b), the branching structure of a cauliflower is more complex. Simple measurements show that there are about $p = 18$ branches of sizes $L_k = L_0 \approx L_0[0.5 - 0.2(k-1)]$ for $k = 1, 2, \ldots, p$, where L_0 is the diameter of the complete head of the cauliflower. (We keep removing branches until the diameter of the central remaining part is equal to the half diameter of the complete head and we assume that it is similar to the rest of the branches). Subsequent generations (we count at least eight) obey similar rules, however p has a tendency to decrease with the generation number. To find fractal dimension of a cauliflower we will count the number of branches $N(L)$ larger than certain size L. As in case of equal branches, this number scales as L^{-d_f}. Calculations, similar to those presented in [77] show that in this case the fractal dimension is equal to

$$d_f = -\frac{\mu}{\sigma^2}\left(\sqrt{1 - \ln p\,\frac{2\sigma^2}{\mu^2}} - 1\right),\tag{7}$$

where μ and σ are the mean and the standard deviation of $\ln q_k \equiv \ln(L_0/L_k)$. For the particular cauliflower shown in Fig. 2a the measurements presented in Fig. 2b give $d_f = 2.75$, which is very close to the estimate $d_f = 2.8$, of [64]. The physiological reason for such a peculiar branching pattern of cauliflower is not known. It is probably designed to store energy for a quick production of a dense inflorescence.

For a tree, [142] the sum of the cross-section areas of the two daughter branches according to Leonardo is equal to the cross-section area of the mother branch. This can be understood because the number of capillary bundles going from the mother to the daughters is conserved. Accordingly, we have a relation between the daughter branch diameters d_1 and d_2 and the mother branch diameter d_0,

$$d_1^\eta + d_2^\eta = d_0^\eta,\tag{8}$$

where $\eta = 2$. We can assume that the branch diameter of the largest daughter branch scales as $d_1^\eta = r d_0^\eta$, where r is asymmetry ratio maintaining for each generation of branches. In case of a symmetric tree $d_1 = d_2$, $r = 1/2$. If we assume that the branch length $L = sd$, where s is a constant which is the same for any branch of the tree, then we can relate our tree to a fractal model with $p = 2$ and $q = 2^{1/\eta}$. Using Eq. (5) we get $d_f = \eta = 2$, i.e. the tree skeleton or the surface of the terminal branches is an ob-

ject of fractal dimension two embedded in the three-dimensional space. This is quite natural because all the leaves whose number is proportional to the number of terminal branches must be exposed to the sunlight and the easiest way to achieve this is to place all the leaves on the surface of a sphere which is a two-dimensional object.

For an asymmetric tree the fractal dimension can be computed using Eq. (7) with

$$\mu = |\ln[r(1-r)]/2\eta|\tag{9}$$

and

$$\sigma = |\ln[r/(1-r)]/2\eta|.\tag{10}$$

This value is slightly larger than 2 for a wide range of r (Fig. 2c). This property may be related to a tendency of a tree to maximize the surface of it leaves, which must not be necessarily exposed to the direct sunlight but can suffice on the light reflected by the outer leaves.

For a lung [56,66,77,102,117], the flow is not a capillary but can be assumed viscous. According to flow conservation, the sum of the air flows of the daughter branches must be equal to the flow of the mother branch:

$$Q_1 + Q_2 = Q_0.\tag{11}$$

For the Poiseuille flow, $Q \sim \Delta P d^4/L$, where ΔP is the pressure drop, which is supposed to be the same for the airways of all sizes. Assuming that the lung maintains the ratio $s = L/d$ in all generations of branches, we conclude that the diameters of the daughter and mother branches must satisfy Eq. (8) with $\eta = 4 - 1 = 3$. Accordingly, for a symmetrically branching lung $d_f = \eta = 3$, which means that the surface of the alveoli which is proportional to the number of terminal airways scales as L^3, i.e. it is a space filling object. Again this prediction is quite reasonable because nature tends to maximize the gas exchange area so that it completely fills the volume of the lung.

In reality, the flow in the large airways is turbulent, and the parameter η of lungs of different species varies between 2 and 3 [77]. Also the lungs are known to be asymmetric and the ratio $r = Q_1/Q_0 \neq 1/2$ changes from one generation of the airways to the next [77]. However, Eq. (7) with μ and σ given by Eqs. (9) and (10) shows that the fractal dimension of an asymmetric tree remains very close to 3 for a wide range of $0.146 < r < 0.854$ (Fig. 2c). The fact that the estimated fractal dimension is slightly larger than 3 does not contradict common sense because the branching stops when the airway diameter becomes smaller than some critical cutoff. Other implications of the lung asymmetry are discussed in [77]. An interesting idea

was proposed in [117], according to which the dependence of the branching pattern on the generation of the airways can be derived from optimization principles and give rise to the complex value of the fractal dimension.

An interesting property of the crackle sound produced by diseased lungs is that during inflation the resistance to airflow of the small airways decreases in discrete jumps. Airways do not open individually, but in a sequence of bursts or avalanches involving many airways; both the size of these jumps and the time intervals between jumps follow power-law distributions [128]. These avalanches are not related to the SOC as one might expect, but their power law distributions directly follow from the branching structure of lungs (see [3] and references therein).

Fractal Metabolic Rates

An important question in biology is the scaling of the metabolic rate with respect to the body mass (Kleiber's Law). It turns out that for almost all species of animals, which differ in terms of mass by 21 orders of magnitudes, the metabolic rate B scales as $B \sim M^{3/4}$ [28,112,140]. The scatter plot $\ln B$ vs $\ln M$ is a narrow cloud concentrated around the straight line with a slope 3/4 (Fig. 3). This is one of the many examples of the allometric laws which describe the dependence of various biological parameters on the body mass or population size [1,9,87,105,106].

A simple argument based on the idea that the thermal energy loss and hence the metabolic rate should be proportional to the surface area predicts however that

Fractals in Biology, Figure 3
Dependence of metabolic rate on body mass for different types of organisms (Kleiber's Law, after [112]). The least square fit lines have slopes 0.76 ± 0.01

$B \sim L^2 \sim M^{2/3}$. Postulating that the metabolic rate is proportional to some effective metabolic area leads to a strange prediction that this area should have a fractal dimension of 9/4.

Many attempts has been made to explain this interesting fact [10,32,142,143]. One particular attempt [142] was based on the ideas of energy optimization and the fractal organization of the cardiovascular system. However, the arguments were crucially dependent on such details as turbulent vs laminar flow, the elasticity of the blood vessels and the pulsatory nature of the cardiovascular system. The fact that this derivation does not work for species with different types of circulatory systems suggests that there might be a completely different and quite general explanation of this phenomenon.

Recently [10], as the theory of networks has became a hot subject, a general explanation of the metabolic rates was provided using a mathematical theorem that the total flow Q (e. g. of the blood) in the most efficient supply network must scale as $Q \sim BL/u$, where L is the linear size of the network, B is the total consumption rate (e. g. metabolic rate), and u is the linear size of each consumption unit (e. g. cell). The authors argue that the total flow is proportional to the total amount of the liquid in the network, which must scale as the body mass M. On the other hand, $L \sim M^{1/3}$. If one assumes that u is independent of the body mass, then $B \sim M^{2/3}$, which is identical to the simple but incorrect prediction based on the surface area. In order to make ends meet, the authors postulate that some combination of parameters must be independent of the body size, from where it follows that $u \sim M^{1/12}$. If one identifies a consumption unit with the cell, it seems that the cell size must scale as $L^{1/4}$. Thus the cells of a whale ($L = 30$ m) must be about 12 times larger than those of a C. elegance ($L = 1$ mm), which more or less consistent with the empirical data [110] for slowly dividing cells such as neurons in mammals. However, the issue of the metabolic scaling is still not fully resolved [11,33,144]. It is likely that the universal explanation of Kleiber's law is impossible. Moreover, recently it was observed that Kleiber's law does not hold for plants [105].

An interesting example of an allometric law which may have some fractal implications is the scaling of the brain size with the body mass $M_b \sim M^{3/4}$ [1]. Assuming that the mass of the brain is proportional to the mass of the neurons in the body, we can conclude that if the average mass of a neuron does not depend on the body mass, neurons must form a fractal set with the fractal dimension 9/4. However, if we assume that the mass of a neuron must scale with the body mass as $M^{1/12}$, as it follows from [10,110], we can conclude that the number of neu-

rons in the body scales simply as $M^{2/3} \sim L^2$, which means that the number of neurons is proportional to the surface area of the body. The latter conclusion is physiologically more plausible than the former, since it is obvious that the neurons are more likely to be located near the surface of an organism.
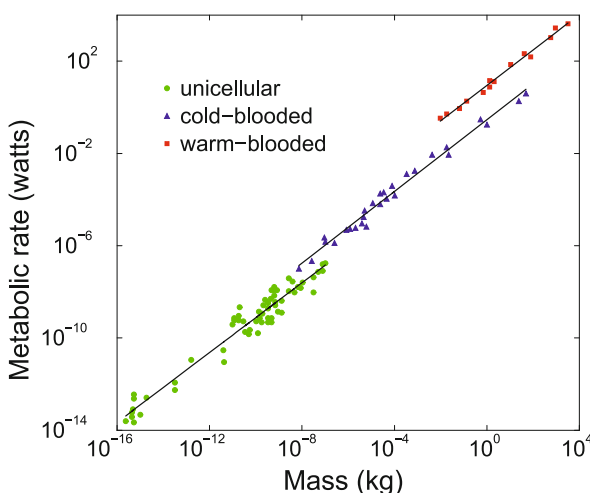
From the point of view of comparative zoology, the universal scaling of the brain mass is not as important as the deviations from it. It is useful [1] to characterize an organism by the ratio of its actual brain mass to its expected brain mass which is defined as $E_b = AM^{3/4}$, where A is a constant measured by the intercept of the log–log graph of brain mass versus body mass. Homo sapiens have the largest $M_b/E_b = 8$, several times larger than those of gorillas and chimpanzees.

Physical Models of Biological Fractals

The fractals discussed in Sect. "Self-Similar Branching Structures", which are based on the explicit rules of construction and self-similarity, are useful concepts for analyzing branching structures in the living organisms whose development and growth are programmed according to these rules. However, in nature, there are many instances when fractal shapes emerge just from general physical principles without a special fractal "blueprint" [26,34, 58,124,125,131]. One of such examples particularly relevant to biology is the diffusion limited aggregation (DLA) [145]. Another phenomenon, described by the same equations as DLA and thus producing similar shapes is viscous fingering [17,26,131]. Other examples are a random walk (RW) [42,58,104,141], a self-avoiding walk (SAW) [34,58], and a percolation cluster [26,58,127].

DLA explains the growth of ramified inorganic aggregates which sometimes can be found in rock cracks. The aggregates grow because certain ions or molecules deposit on the surface of the aggregate. These ions come to the surface due to diffusion which is equivalent on the microscopic level to Brownian motion of individual ions. The Brownian trajectories can be modeled as random walks in which the direction of each next step is independent of the previous steps.

The Brownian trajectory itself has fractal properties expressed by the Einstein formula:

$$r^2 = 2dtD,\tag{12}$$

where r^2 is the average square displacement of the Brownian particle during time t, D is the diffusion coefficient and d is the dimensionality of the embedding space. The number of steps, n, of the random walk is proportional to time of the Brownian motion $t = n\tau$, where τ is the
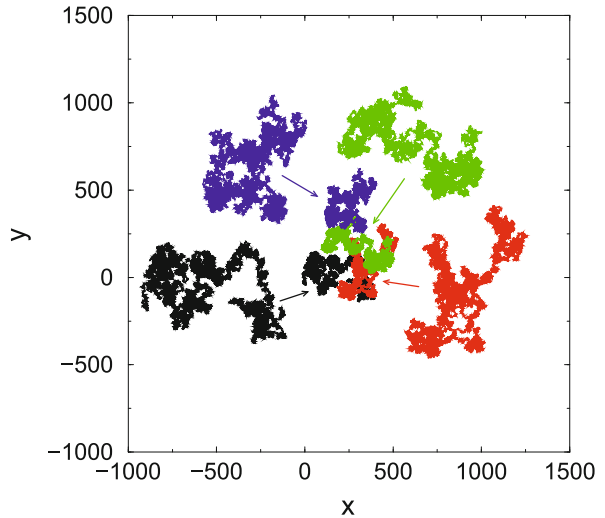
Fractals in Biology, Figure 4
A random walk of $n = 2^{16}$ steps is surrounded by its $p = 4$ parts of $m = n/p = 2^{14}$ steps magnified by factor of $q = \sqrt{4} = 2$. One can see that the shapes and sizes of each of the magnified parts are similar to the shape and size of the whole

average duration of the random walk step. The diffusion coefficient can be expressed as $D = r^2(m)/(2dm\tau)$, where $r^2(m)$ is the average square displacement during m time steps. Accordingly,

$$r(n) = \sqrt{p}\,r(m),\tag{13}$$

which shows that the random walk of n steps consists of $p = n/m$ copies of the one-step walks arranged in space in such a way that the average linear size of this arrangement is $q = \sqrt{n/m}$ times larger than the size of the m-step walk (Fig. 4). Applying Eq. (5), we get $d_f = 2$ which does not depend on the embedding space. It is the same in one-dimensional, two-dimensional, and three-dimensional space. Note that Brownian trajectory is self-similar only in a statistical sense: each of the n concatenated copies of a m-step trajectory are different from each other, but their average square displacements are the same. Our eye can easily catch this statistical self-similarity. A random walk is a cloudy object of an elongated shape, its inertia ellipsoid is characterized by the average ratios of the squares of its axis: 12.1/2.71/1 [18].

The Brownian trajectory itself has a relevance in biology, since it may describe the changing of the electrical potential of a neuron [46], spreading of a colony [68], the foraging trajectory of a bacteria or an animal as well as the motion of proteins and other molecules in the cell. The self-avoiding walk (SAW) which has a smaller fractal dimension ($d_{f,SAW} = 4/3$ in $d = 2$ and $d_{f,SAW} \approx 1.7$ in
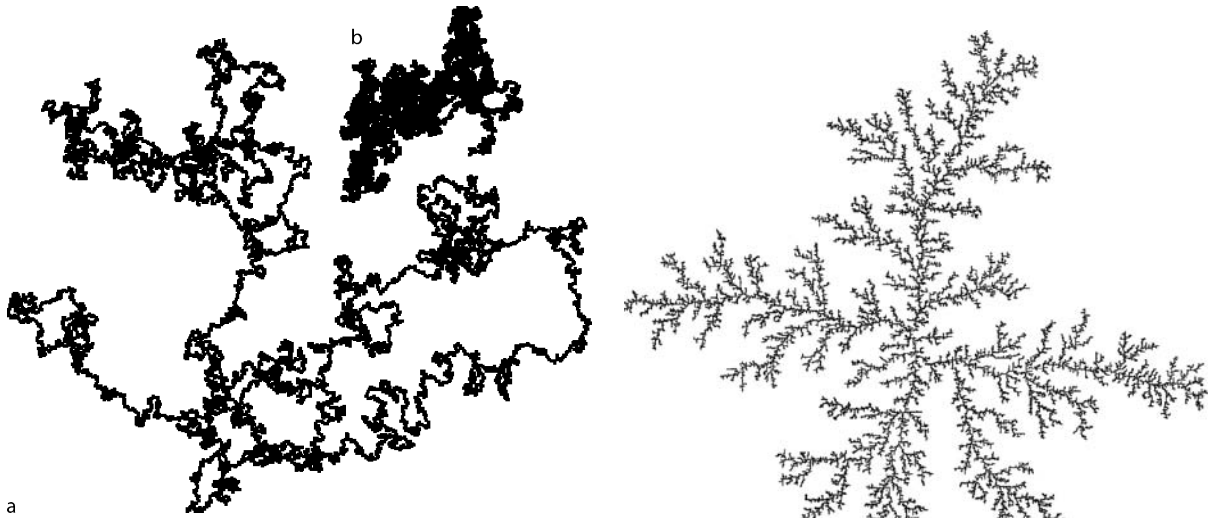
Fractals in Biology, Figure 5
A self-avoiding walk (a) of $n = 10^4$ in comparison with a random walk (b) of the same number of steps, both in $d = 3$. Both walks are produced by molecular dynamic simulations of the bead-on-a-string model of polymers. The hard core diameter of monomers in the SAW is equal to the bond length ℓ, while for the RW it is zero. Comparing their fractal dimensions $d_{f,SAW} \approx 1.7$, and $d_{f,RW} = 2$, one can predict that the average size (radius of inertia) of SAW must be $n^{1/d_{f,SAW} - 1/d_{f,RW}} \approx 2.3$ times larger than that of RW. Indeed the average radius of inertia of SAW and RW are 98ℓ and 40ℓ, respectively. The fact that their complex shapes resemble leaving creatures was noticed by P. G. de Gennes in a cartoon published in his book "Scaling Concepts in Polymer Physics" [34]

$d = 3$) is a model of a polymer (Fig. 5) in a good solvent, such as for example a random coil conformation of a protein. Thus a SAW provides another example of a fractal object which has certain relevance in molecular biology. The fractal properties of SAWs are well established in the works of Nobel Laureates Flory and de Gennes [34,44,63].

Diffusion Limited Aggregation and Bacterial Colonies

The properties of a Brownian trajectory explain the ramified structure of the DLA cluster. Since the Brownian trajectory is not straight, it is very difficult for it to penetrate deep into the fiords of a DLA cluster. With a much higher probability it hits the tips of the branches. Thus DLA cluster (Fig. 6a) has a tree-like structure with usually 5 main branches in 2 dimensions. The analytical determination of its fractal dimension is one of the most challenging questions in modern mathematics. In computer simulations it is defined by measuring the number of aggregated particles versus the radius of gyration of the aggregate (Fig. 6c). The fractal dimension thus found is approximately 1.71
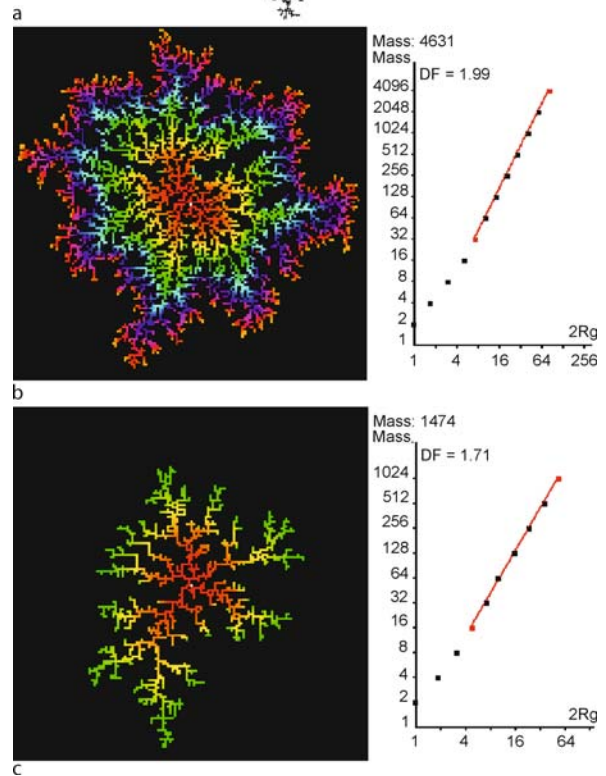
Fractals in Biology, Figure 6
a A DLA cluster of $n = 2^{14}$ particles produced by the aggregation of the random walks on the square lattice (*top*). A comparison of a small BD cluster (**b**) and a DLA cluster (**c**). The color in **b** and **c** indicates the deposition time of a particle. The slopes of the log-log graphs of the mass of the growing aggregates versus their gyration radius give the values of their fractal dimensions in the limit of large mass
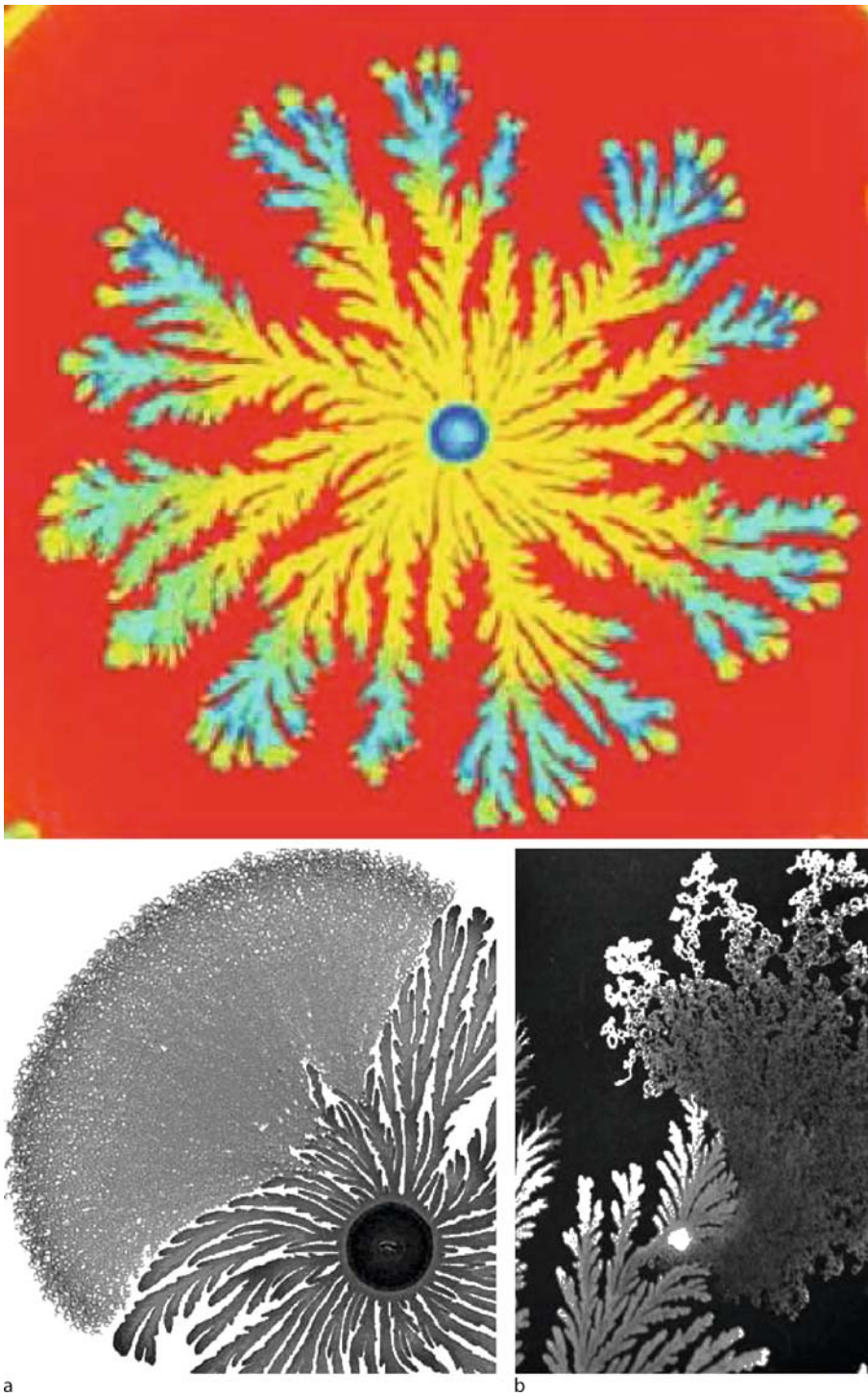
Fractals in Biology, Figure 7
A typical DLA-like colony grown in a Petri dish with a highly viscous substrate (*top*). A change in morphology from DLA-like colony growth to a swimming chiral pattern presumably due to a cell morphotype transition (*bottom*) (from [15])

in two dimensions and 2.50 in three dimensions [34]. It seems that as the aggregate grows larger the fractal dimension slightly increases.

Note that if the deposition is made by particles moving along straight lines (ballistic deposition) the aggregate changes its morphology and becomes almost compact and circular with small fluctuation on the boundaries and small holes in the interior (Fig. 6b). The fractal dimension of the ballistic aggregate coincides with the dimension of embedding space. Ballistic deposition (BD) belongs to the same universality class as the Eden model, the first model to describe the growth of cancer. In this model, each cell on the surface of the cluster can produce an offspring with equal probability. This cell does not diffuse but occupies one of the empty spaces neighboring to the parent cell [13,83].

DLA is supposed to be a good model for growth of bacterial colonies in the regime when nutrients are coming via diffusion in a viscous media from the exterior. Under different conditions bacteria colonies observe various transition in their morphology. If the nutrient supply is greater, the colonies increase their fractal dimension and start to resemble BD.

An interesting phenomenon is observed upon changing the viscosity of the substrate [15]. If the viscosity is small it is profitable for bacteria to swim in order to get to the regions with high nutrient concentrations. If the viscosity is large it is more profitable to grow in a static colony which is supplied by the diffusion of the nutrient from the outside. It seems that the way how the colony grows is inherited in the bacteria gene expression. When a bacteria grown at high viscosity is planted into a low viscosity substrate, its descendants continue to grow in a DLA pattern until a transition in a gene-expression happens in a sufficiently large number of neighboring bacteria which change their morphotype to a swimming one (Fig. 7). All the descendants of these bacteria start to swim and thus quickly take over the slower growing morphotype. Conversely, when a swimming bacteria is planted into a viscous media its descendants continue to swim until a reverse transition of a morphotype happens and the descendants of the bacteria with this new morphotype start a DLA-like growing colony. The morphotype transition can be also induced by fungi. Thus, it is likely that although bacteria are unicellular organisms, they exchange chemical signals similarly to the cells in the multicellular organisms, which undergo a complex process of cell differentiation during organism development (morphogenesis).

There is evidence that the tree roots also follow the DLA pattern, growing in the direction of diffusing nutrients [43]. Coral reefs [80] whose life depends on the diffusion of oxygen and nutrients also may to a certain degree follow DLA or BD patterns. Another interesting conjecture is that neuronal dendrites grow in vivo obeying the same mechanism [29]. In this case, they follow signaling chemicals released by other cells. It was also conjectured that even fingers of vertebrate organisms may branch in a DLA controlled fashion, resembling the pattern of viscous fingering which are observed when a liquid of low viscosity is pushed into a liquid of higher viscosity.

Measuring Fractal Dimension of Real Biological Fractals

There were attempts to measure fractal dimension of growing neurons [29] and other biological objects from photographs using a box-counting method and a circle method developed for empirical determination of fractal dimension. The circle method is the simplified version of the radial distribution function analysis described in the definition section. It consists of placing a center of a circle or a sphere of a variable radius R at each point of an object and counting the average number of other points of this object $M(R)$ found inside these circles or spheres. The slope of $\ln M(R)$ vs. $\ln R$ gives an estimate of the fractal dimension.

The box counting method consists of placing a square grid of a given spacing ℓ on a photograph and counting number of boxes $n(\ell)$ needed to cover all the points belonging to a fractal set under investigation. Each box containing at least one point of a fractal set is likely to contain on average $N(\ell) = \ell^{d_f}$ other points of this set, and $n(\ell)N(\ell) = N$, where N is the total number of the points. Thus for a fractal image $n(\ell) \sim N/\ell^{d_f} \sim \ell^{-d_f}$. Accordingly the fractal dimension of the image can be estimated by the slope of a graph of $\ln n(\ell)$ vs. $\ln \ell$.

The problem is that real biological objects do not have many orders of self-similarity, and also, as we see above, only a skeleton or a perimeter of a real object has fractal properties. The box counting method usually produces curvy lines on a log–log paper with at best one decade of approximately constant slope (Fig. 8). What is most disappointing is that almost any image analyzed in this fashion produces a graph of similar quality. For example, an Einstein cartoon presented in [93] has the same fractal dimension as some of the growing neurons. Therefore, experimental determination of the fractal dimension from the photographs is no longer in use in scientific publications. However, in the 1980s and early 1990s when the computer scanning of images became popular and the fractals were at the highest point of their career, these exercises were frequent.
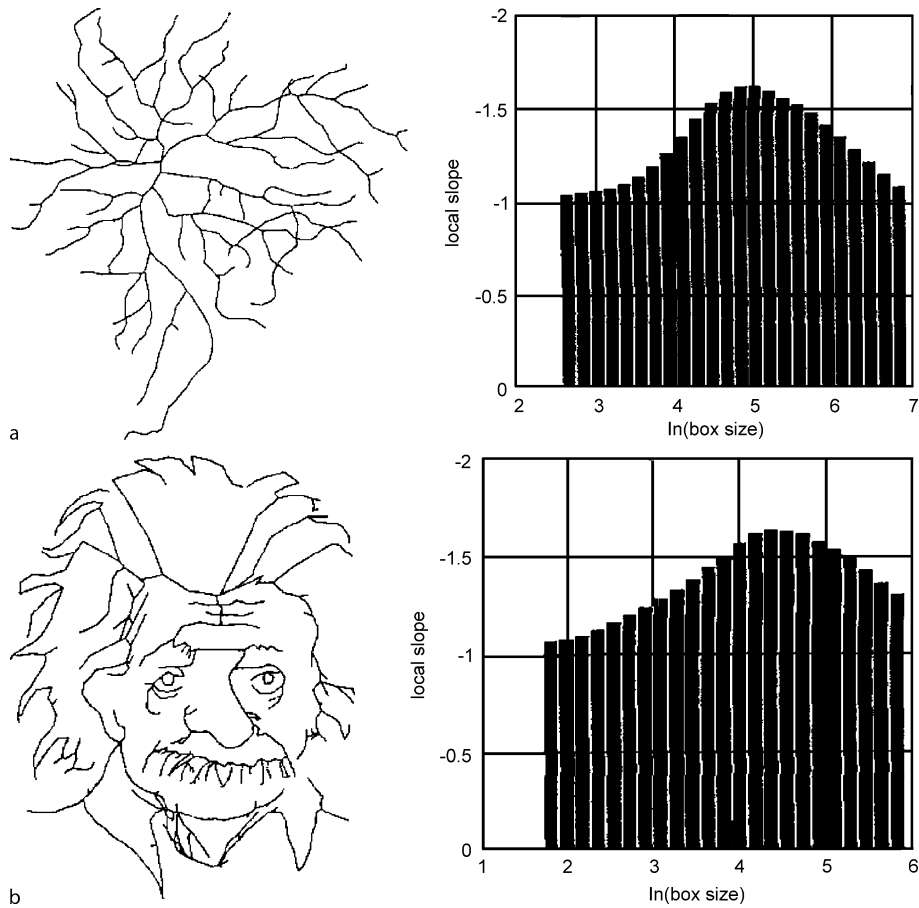
Fractals in Biology, Figure 8
An image of a neuron and its box-counting analysis (**a**) together with the analogous analysis of an Einstein cartoon (**b**). The log–log plot of the box-counting data has a shape of a curve with the changing slope. *Bars* represent the local slope of this graph, which may be interpreted as some effective fractal dimension. However, in both graphs the slopes change dramatically from 1 to almost 1.7 and it is obvious that any comparison to the DLA growth mechanism based on these graphs is invalid (from [93])

For example, R. Voss [137] analyzed fractal dimensions of Chinese graphics of various historical periods. He found that the fractal dimension of the drawings fluctuated from century to century and was about 1.3 at the time of the highest achievements of this technique. It was argued that the perimeters of many natural objects such as perimeters of the percolation clusters and random walks have similar fractal dimension; hence the works of the artists who were able to reproduce this feature were the most pleasant for our eyes.

Recently, the box counting method was used to analyze complex networks such as protein interaction networks (PIN). In this case the size of the box was defined as the maximal number of edges needed to connect any two nodes belonging to the box. The box-counting method appears to be a powerful tool for analyzing self-similarity of the network structure [120,121].

Percolation and Forest Fires

It was also conjectured that biological habitats of certain species may have fractal properties. However it is not clear whether we have a true self-similarity or just an apparent mosaic pattern which is due to complex topography and geology of the area. If there is no reasonable theoretical model of an ecological process, which displays true fractal properties the assertions of the fractality of a habitat are pointless. One such model, which produces fractal clusters is percolation [26,127]. Suppose a lightning hits a random tree in the forest. If the forest is sufficiently dry, and the inflammable trees are close enough the fire will spread from a tree to a tree and can burn the entire forest. If the inflammable trees form a connected cluster, defined as such that its any two trees are connected with a path along which the fire can spread, all the trees in this cluster will

be burnt down. Percolation theory guarantees that there is a critical density p_c of inflammable trees below which all the connected clusters are finite and the fire will naturally extinguish burning only an infinitesimally small portion of the forest. In contrast, above this threshold there exists a giant cluster which constitutes a finite portion of the forest so that the fire started by a random lightning will on average destroy a finite portion of the forest. Exactly at the critical threshold, the giant cluster is a fractal with a fractal dimension $d_f = 91/48 \approx 1.89$. The probability to burn a finite cluster of mass S follows a power law $P(S) \sim S^{-d/d_f}$. Above and below the percolation threshold, this distribution is truncated by an exponential factor which specifies that clusters of linear dimensions exceeding a characteristic size are exponentially rare. The structure of the clusters which do not exceed this characteristic scale is also fractal, so if their mass S is plotted versus their linear size R (e. g. radius of inertia) it follows a power law $S \sim R^{d_f}$

In the natural environment the thunderstorms happen regularly and they produce fractal patterns of burned down patches. However, if the density of trees is small, the patches are finite and the fires are confined. As the density the trees reaches the critical threshold, the next thunderstorm is likely to destroy the giant cluster of the forest, which will produce a bare patch of a fractal shape spreading over the entire forest. No major forest fire will happen until the new forest covers this patch creating a new giant cluster, because the remaining disconnected groves are of finite size.

There is an evidence that the forest is the system which drives itself to a critical point [30,78,111] (Fig. 9). Suppose that each year there is certain number of lightning strokes per unit area n_l. The average number of trees in a cluster is $\langle S \rangle = a|p - p_c|^{-\gamma}$, where $\gamma = 43/18$ is one of the critical exponents describing percolation, a is a constant, and p is the number of inflammable trees per unit area. Thus the number of trees destroyed annually per unit area is $n_l a|p_c - p|^{-\gamma}$. On the other hand, since the trees are growing, the number of trees per unit area increases annually by n_t, which is the parameter of the ecosystem. At equilibrium, $n_l a|p_c - p|^{-\gamma} = n_t$. Accordingly, the equilibrium density of trees must reach $p_e = p_c - (an_l/n_t)^{1/\gamma}$. For very low n_l, p_e will be almost equal to p_c, and the chance that in the next forest fire a giant cluster which spans the entire ecosystem will be burned is very high. It can be shown that for a given chance c of hitting such a cluster in a random lightning stroke, the density of trees must reach $p = p_c - f(c)L^{-1/\nu}$, where L is the linear size of the forest, $\nu = 4/3$ is the correlation critical exponent and $f(c) > 0$ is a logarithmically growing function of c.

Accordingly, if $n_t/n_l > bL^{\gamma/\nu}$, where b is some constant, the chance of getting a devastating forest fire is close to 100%. We have here a paradoxical situation: the more frequent are the forest fires, the least dangerous they are. This implies that the fire fighters should not extinguish small forest fires which will be contained by themselves. Rather they should annually cut a certain fraction of trees to decrease n_t.

As we see, the forest fires can be regarded as a self-organized critical (SOC) system which drive themselves towards criticality. As in many SOC systems, here there are two processes one is by several orders of magnitudes faster than the other. In this case they are the tree growing process and the lightning striking process. The model reaches the critical point if a tuning parameter $n_t/n_l \to \infty$ and in addition $n_t \to 0$ and $L \to \infty$, which are quite reasonable assumptions. In a regular critical system a tuning parameter (e. g. temperature) must be in the vicinity of a specific finite value. In a SOC system the tuning parameter must be just large enough.

There is an evidence, that the forest fires follow a power law distribution [109]. One can also speculate that the areas burned down by the previous fires shape fractal habitats for light loving and fire-resistant trees such as pines. The attempts to measure a fractal dimension of such habitats from aerial photographs are dubious due to the limitations discussed above and also because by adjusting a color threshold one can produce fractal-like clusters in almost any image. This artifact is itself a trivial consequence of the percolation theory.

The frequently reported Zipf's law for the sizes of colonies of various species including the areas and populations of the cities [146] usually arises not from fractality of the habitats but from preferential attachment growth in which the old colonies grow in proportion to their present population, while the new colonies may form with a small probability. The preferential attachment model is a very simple mechanism which can create a power-law distribution of colony sizes $P(S) \sim S^{-2-\phi}$, where ϕ is a small correction which is proportional to the probability of formation of new colonies [25]. Other examples of power laws in biology [67], such as distributions of clusters in the metabolic networks, the distribution of families of proteins, etc. are also most likely come from the preferential attachment or also, in some cases, can arise as artifacts of specifically selected similarity threshold, which brings a network under consideration to a critical point of the percolation theory.

The epidemic spreading can also be described by a percolation model. In this case the contagious disease spreads from a person to a person as the forest fire spreads from
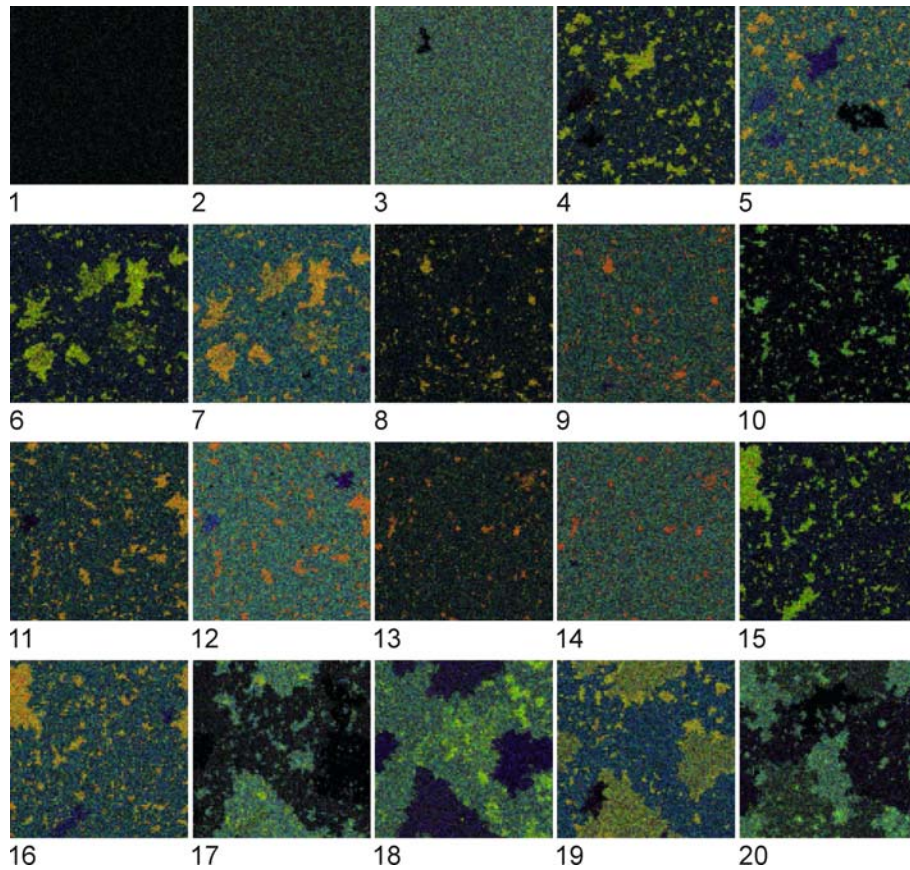
Fractals in Biology, Figure 9
A sequence of frames of a forest fire model. Each tree occupies a site on 500 × 500 square lattice. At each time step a tree (a colored site) is planted at a randomly chosen empty site (*black*). Each 10,000 time steps a lightning strikes a randomly chosen tree and the forest fire eliminates a connected cluster of trees. The frames are separated by 60,000 time steps. The color code indicates the age of the trees from *blue* (young) to *red* (old). The initial state of the system is an empty lattice. As the concentration of trees reaches percolation threshold (frame 3) a small finite cluster is burned. However it does not sufficiently decrease the concentration of trees and it continues to build up until a devastating forest fire occurs between frames 3 and 4, with only few green groves left. Between frames 4 and 5 several lightnings hit these groves and they are burned down, while surrounding patch of the old fire continues to be populated by the new trees. Between frame 5 and 6 a new devastating forest fire occurs. At the end of the movie huge intermittent forest fires produce gigantic patches of dense and rare groves of various ages

a tree to a tree. Analogously to the forest fire model, a person who caught the disease dies or recovers and becomes immune to the disease, so he or she cannot catch it again. This model is called susceptible-infective-removed (SIR) [62]. The main difference is that the epidemics spread not on a two-dimensional plane but on the network describing contacts among people [61,86]. This network is usually supposed to be scale free, i. e. the number of contacts different people have (degree) is distributed according to a an inverse power law [2,71]. As the epidemic spreads the number of connections in the susceptible population depletes [115] so the susceptible population comes to a percolation threshold after which the epidemic stops.

This model explains for example why the Black Death epidemic stopped in the late 14th century after killing about one third of the total European population.

Critical Point and Long-Range Correlations

Percolation critical threshold [26,127] is an example of critical phenomena the most well known of which is the liquid-gas critical point [124]. As one heats a sample of any liquid occupying a certain fraction of a closed rigid container, part of the liquid evaporates and the pressure in the container increases so that the sample of liquid remains at equilibrium with its vapor. However, at certain

temperature the visible boundary between the liquid at the bottom and the gas on the top becomes fuzzy and eventually the system becomes completely nontransparent as milk. This phenomena is called critical opalescence and the temperature, pressure, and density at which it happens is called a critical point of this liquid. For water, the critical point is $T_c = 374\,\mathrm{C}^\circ$, $P_c = 220\,\mathrm{atm}$ and $\rho_c = 330\,\mathrm{kg/m^3}$. As the temperature goes above the critical point the system becomes transparent again, but the phase boundary disappears: liquid and gas cannot coexist above the critical temperature. They form a single phase, supercritical fluid, which has certain properties of both gas (high compressibility) and liquid (high density and slow diffusivity).

At the critical point the system consists of regions of high density (liquid like) and low density (gas like) of all sizes from the size of a single molecule to several microns across which are larger than the wave length of visible light $\approx 0.5\,\mu\mathrm{m}$. These giant density fluctuations scatter visible light causing the critical opalescence. The characteristic linear size of the density fluctuations is called the correlation length ξ which diverges at critical temperature as $\xi \sim |T - T_c|^{-\nu}$. The shapes of these density fluctuations are self-similar, fractal-like. The system can be represented as a superposition of a uniform set with the density corresponding to the average density of the system and a fractal set with the fractal dimension d_f. The density correlation function $h(r)$ decreases for $r \to \infty$ as $r^{-\chi}\exp(-r/\xi)$, where $\chi \equiv d - 2 + \eta = d - d_f$ and $d = 1, 2, 3, \ldots$ is the dimension of space in which the critical behavior is observed. The exponential cutoff sets the upper limit of the fractal density fluctuations to be equal to the correlation length. The lower limit of the fractal behavior of the density fluctuations is one molecule. As the system approaches the critical point, the range of fractal behavior increases and can reach at least three orders of magnitude in the narrow vicinity of the critical point. The intensity of light scattered by density fluctuations at certain angle can be expressed via the Fourier transform of the density correlation function $S(\mathbf{f}) \sim \int h(\mathbf{r})\exp(i\mathbf{r}\cdot\mathbf{f})d\mathbf{r}$, where the integral is taken over d-dimensional space. This experimentally observed quantity $S(f)$ is called the structure factor. It is a powerful tool to study the fractal properties of matter since it can uncover the presence of a fractal set even if it is superposed with a set of uniform density. If the density correlation function has a power law behavior $h(r) \sim r^{-\chi}$, the structure factor also has a power law behavior $S(f) \sim f^{\chi-d}$. For the true fractals with $d_f < d$, the average density over the entire embedding space is zero, so the correlation function coincides with the average density of points of the set at certain distance from a given point, $h(r) \sim r^{-d+1}dM(r)/dr \sim r^{d_f - d}$, where $M(r) \sim r^{d_f}$ is the

mass of the fractal within the radius r from a given point. Thus, for a fractal set of points with fractal dimension d_f, the structure factor has a simple power law form $S(f) \sim f^{-d_f}$. Therefore whenever $S(f) \sim f^{-\beta}$, $\beta < d$ is identified with the fractal dimension and the system is said to have fractal density fluctuations even if the average density of the system is not zero.

There are several methods of detecting spatial correlations in ecology [45,108] including box-counting [101]. In addition one can study the density correlation function $h(r)$ of certain species on the surface of the Earth defined as $h(r) = N(r)/2\pi r\Delta r - N/A$, where $N(r)$ is the average number of the representatives of a species within a circular rim of radius r and width Δr from a given representative, N is the total population, and A is the total area. For certain species, the correlation function may follow about one order of magnitude of a fractal (power-law) behavior [9,54,101]. One can speculate that there is some effective attraction between the individuals such as cooperation (mimicking the van der Walls forces between molecules) and a tendency to spread over the larger territory (mimicking the thermal motion of the particles). The interplay of these two tendencies may produce a fractal behavior like in the vicinity of the critical point.

However, there is no reason of why the ecological system although complex and chaotic [81] should drive itself to a critical point. Whether or not there is a fractal behavior, the density correlation function is a useful way to describe the spatial distribution of the population. It can be applied not only in ecology but also in physiology and anatomy to describe the distribution of cells, for example, neurons in the cortex [22].

Lévy Flight Foraging

A different set of mathematical models which produce fractal patterns and may be relevant in ecology is the Lévy flight and Lévy walk models [118,147]. The Lévy flight model is a generalization of a random walk, in which the distribution of the steps (flights) of length ℓ follows a power law $P(\ell) \sim \ell^{-\mu}$ with $\mu < 3$. Such distributions do not have a finite variance. The probability density of the landing points of the Lévy flight converges not to a Gaussian as for a normal random walk with a finite step variance but to a Lévy stable distribution with the parameter $\alpha \equiv \mu - 1$. The landing points of a Lévy flight form fractal dust similar to a Cantor set with the fractal dimension $d_f = \alpha$.

It was conjectured that certain animals may follow Lévy flights during foraging [103,116,132,133]. It is a mathematical theorem [23,24] that in case when targets

are scarce but there is a high probability that a new target could be discovered in the vicinity of the previously found target, the optimal strategy for a forager is to perform a Lévy flight with $\mu = 2 + \phi$, where ϕ is a small correction which depends on the density of the target sites, the radius of sight of the forager and the probability to find a new target in the vicinity of the old one. In this type of "inverse square" foraging strategy a balance is reached between finding new rich random target areas and returning back to the previously visited area which although depleted, may still provide some food necessary for survival. The original report [132] of the distribution of the flight times of wandering albatross was in agreement with the theory.

However, recent work [39] using much longer flight time records and more reliable analysis showed that the distribution of flight times is better described by a Poisson distribution corresponding to regular random walk rather then by a power law. Several other reports of the Lévy flight foraging are also found dubious. The theory [23] however predicts that the inverse square law of foraging is optimal only in case of scarce sources distribution. Thus if the harvest is good and the food is plenty there is no reason to discover the new target locations and the regular random walk strategy becomes the most efficient. Subsequent observations show that power laws exist in some other marine predator search behavior [119]. In order to find a definitive answer, the study must be repeated for the course of many successive years characterized by different harvests.

Once the miniature electronic devices for tracing animals are becoming cheaper and more efficient, a new branch of electronic ecology is emerging with the goal of quantifying migratory patterns and foraging strategies of various species in various environments. Whether or not the foraging patterns are fractal, this novel approach will help to establish better conservational policy with scientifically sound borders of wild-life reservations. Recent observations indicate that human mobility patterns might also possess Lévy flight properties [49].

Dynamic Fractals

It is not necessary that a fractal is a real geometrical object embedded in the regular three-dimensional space. Fractals can be found in the properties of the time series describing the behavior of the biological objects. One classical example of a biological phenomenon which under certain condition may display fractal properties is the famous logistic map [38,82,95,123] based on the ideas of a great British economist and demographer Robert Malthus. Sup-

pose that there is a population N_t of a certain species enclosed in a finite habitat (e. g. island) at time t. Here t is an integer index which may denote year. Suppose that at time $t + 1$ the population becomes

$$ N_{t+1} = b N_t - d N_t^2 \,, $$

where b is the natural birth rate and d is the death rate caused by the competition for limited resources. In the most primitive model, the animals are treated as particles randomly distributed over area A annihilating at each time step if the distance between them is less than r. In this case $d = \pi r^2 / A$. The normalized population $x_t = dN_t / b$ obeys a recursive relation with a single parameter b:

$$ x_{t+1} = b x_t (1 - x_t) \,. $$

The behavior of the population is quite different for different b. For $b \leq 1$ the population dies out as $t \to \infty$. For $1 < b \leq b_0 = 3$ the population converges to a stable size. If $b_{n-1} < b \leq b_n$, the population repetitively visits 2^n values $0 < x_1, \ldots, x_{2^n} < 1$ called attractors as $t \to \infty$. The bifurcation points $3 = b_0 < b_1 < b_2 < \cdots < b_n < b_\infty \approx 3.569945672$ converge to a critical value b_∞, at which the set of population sizes becomes a fractal with fractal dimension $d_f \approx 0.52$ [50]. This fractal set resembles a Cantor set confined between 0 and 1. For $b_\infty < b < 4$, the behavior is extremely complex. At certain values b chaos emerges and the behavior of the population become unpredictable, i. e. exponentially sensitive to the initial conditions. At some intervals of parameter b, the predictable behavior with a finite attractor set is restored. For $b > 4$, the population inevitably dies out. Although the set of attractors becomes truly fractal only at certain values of the birth rate, and the particular value of the fractal dimension does not have any biological meaning, the logistic map has a paradigmatic value in the studies of population dynamics with an essentially Malthusian take home massage: excessive birth rate leads to disastrous consequences such as famines coming at unpredictable times, and in the case of Homo sapiens to devastating wars and revolutions.

Fractals and Time Series

As we can see in the previous section, even simple systems characterized by nonlinear feedbacks may display complex temporal behavior, which often becomes chaotic and sometimes to fractal. Obviously, such features must be present in the behavior of the nervous system, in particular in the human brain which is probably the most complex system known to contemporary science. Nevertheless, the source of fractal behavior can be sometimes trivial with-

out evidence of any cognitive ability. An a example of such a trivial fractal behavior is the random firing of a neuron [46] which integrates through its dendritic synapses inhibitory and excitatory signals from its neighbors. The action potential of such a neuron can be viewed as performing a one-dimensional random walk going down if an inhibitory signal comes from a synapse or going up if an excitatory signal comes form a different synapse. As soon as the action potential reaches a firing threshold the neuron fires, its action potential drops to an original value and the random walk starts again. Thus the time intervals between the firing spikes of such a neuron are distributed as the returning times of a one-dimensional random walk to an origin. It is well known [42,58,104,141], that the probability density $P(\Delta t)$ of the random walk returns scales as $(\Delta t)^{-\mu}$ with $\mu = 3/2$. Accordingly, the spikes on the time axis form a fractal dust with the fractal dimension $d_f = \mu - 1 = 1/2$.

A useful way to study the correlations in the time series is to compute its autocorrelation function and its Fourier transform which is called power spectrum, analogous to the structure factor for the spatial correlation analysis. Due to the property of the Fourier transform to convert a convolution into a product, the power spectrum is also equal to the square of the Fourier transform of the original time series. Accordingly it has a simple physical meaning telling how much energy is carried in a certain frequency range. In case of a completely uncorrelated signal, the power spectrum is completely flat, which means that all frequencies carry the same energy as in white light which is the mixture of all the rainbow colors of different frequencies. Accordingly a signal which has a flat power spectrum is called white noise. If the autocorrelation function $C(t)$ decays for $t \to \infty$ as $C(t) \sim t^{-\chi}$, where $0 < \chi < 1$, the power spectrum $S(f)$ of this time series diverges as $f^{-\beta}$, with $\beta = 1 - \chi$. Thus the LRPLC in the time series can be detected by studying the power spectrum. There are alternative ways of detecting LRPLC in time series, such as Hurst analysis [41], detrended fluctuation analysis (DFA) [19,98] and wavelet analysis [4]. These methods are useful for studying short time series for which the power spectrum is too noisy. They measure the Hurst exponent $\alpha = (1 + \beta)/2 = 1 - \chi/2$.

It can be shown [123] that for a time series which is equal to zero everywhere except at points $t_1, t_2, \ldots, t_n, \ldots$, at which it is equal to unity and these points form a fractal set with fractal dimension d_f, the time autocorrelation function $C(t)$ decreases for $t \to \infty$ as $C(t) \sim t^{-\chi}$, where $\chi = 1 - d_f$. Therefore the power spectrum $S(f)$ of this time series diverges for $f \to 0$ as $f^{-\beta}$, with $\beta = 1 - \chi = d_f$. Accordingly, for the random walk model

of neuron firing, the power spectrum is characterized by $\beta = 1/2$.

In the more general case, the distribution of the intervals $\Delta t_n = t_n - t_{n-1}$ can decay as a power law $P(\Delta t) \sim (\Delta t)^{-\mu}$, with $1 < \mu < 3$. For $\mu < 2$, the set $t_1, t_2, \ldots, t_n$ is a fractal, and the power spectrum decreases as power law $S(f) \sim f^{-d_f} = f^{-\mu+1}$. For $2 < \mu < 3$, the set $t_1, t_2, \ldots, t_n$ has a finite density, with $d_f = D = 1$. However, the power spectrum and the correlation function maintain their power law behavior for $\mu < 3$. This behavior indicates that although the time series itself is uniform, the temporal fluctuations remain fractal. In this case, the exponent β characterizing the low frequency limit is given by $\beta = 3 - \mu$. The maximal value of $\beta = 1$ is achieved when $\mu = 2$. This type of signal is called $1/f$-noise or "red" noise. If $\mu \geq 3$, $\beta = 0$ in the limit of low frequencies and we again recover white noise.

The physical meaning of $1/f$ noise is that the temporal correlations are of infinite range. For the majority of the processes in nature, temporal correlations decay exponentially $C(t) \sim \exp(-t/\tau)$, where the characteristic memory span τ is called relaxation or correlation time. For a time series with a finite correlation time τ, the power spectrum has a Lorentzian shape, namely it stays constant for $f < 1/\tau$ and decreases as $1/f^2$ for $f > 1/\tau$. The signal in which $S(f) \sim 1/f^2$ is called brown noise, because it describes the time behavior of the one-dimensional Brownian motion. Thus for the majority of the natural processes, the time spectrum has a crossover from a white noise at low frequencies to a brown noise at high frequencies. The relatively unusual case when $S(f) \sim f^{-\beta}$, where $0 < \beta < 1$ is called fractal noise because as we see above it describes the behavior of the fractal time series. $1/f$-noise is the special type of the fractal noise corresponding to the maximal value of β, which can be achieved in the limit of low frequencies.

R. Voss and J. Clarke [135,136] had analyzed the music written by different composers and found that it follows $1/f$ noise over at least three orders of magnitude. It means that music does not have a characteristic time-scale. There is an evidence that physiological process such as heart-beat, gate, breath and sleeping patterns as well as certain types of human activity such as sending e-mails has certain fractal features [6,16,20,27,47,55,57, 72,73,97,99,100,113,126,129]. A. Goldberger [99,126] suggested that music is pleasant for us because it mimics the fractal features of our physiology. It has to be pointed out that in all these physiological time series there is no clear power law behavior expanding over many orders of magnitude. There is also no simple clear explanation of the origins of fractality. One possible mechanism could be due

the distribution of the return times of a random walk, which has been used to explain the sleeping patterns and response to the e-mails.

SOC and Biological Evolution

Self-organized criticality (SOC) [8,92] describes the behavior of the systems far form equilibrium, the general feature of which is a slow increase of strain which is interrupted by an avalanche-like stress release. These avalanches are distributed in a power law fashion. The power spectrum of an activity at a given spatial site is described by a fractal noise. One of the most successful application of SOC is the explanation of the power-law distribution of the magnitudes of the earthquakes (Gutenberg–Richter's law) [89,130].

The simple physical models of self-organized criticality are invasion percolation [26,127], sand-pile model [8], and Bak–Sneppen model of biological evolution [7]. In the one-dimensional Bak–Sneppen model, an ecosystem is represented by a linear chain of the pray-predator relationships in with each species is represented by a site on a straight line surrounded by its predator (a site to the right) and a pray (a site to the left). Each site is characterized by its fitness f which at the beginning is uniformly distributed between 0 and 1. At each time step, a site with the lowest fitness becomes extinct and is replaced by a mutated species with a new fitness randomly taken from a uniform distribution between 0 and 1. The fitnesses of its two neighbors (predator and prey) are also changed at random. After certain equilibration time, the fitness of almost all the species except a few becomes larger than a certain critical value f_c. These few active species with low fitness which can spontaneously mutate form a fractal set on the pray-predator line. The activity of each site can be represented by a time series of mutations shown as spikes corresponding to the times of individual mutations at this site. The power spectrum of this time series indicates the presence of the fractal noise. At a steady state the minimal fitness value which spontaneously mutates fluctuates below f_c with a small probability $P(\epsilon)$ comes into the interval between $f_c - \epsilon$ and f_c. The distribution of the first return times Δt_ϵ to a given ϵ-vicinity of f_c follows a power law with an ϵ-dependent exponential cutoff. Since each time step corresponds to a mutation, the time interval for which f stays below $f_c - \epsilon$ corresponds to an avalanche of mutations caused by a mutation of a very stable species with $f > f_c - \epsilon$. Accordingly one can speculate, that evolution goes as a punctuated equilibrium so that an extinction of a stable species causes a gigantic extinction of many other species which hitherto have been well adopted. The problem with this SOC model is the def-

inition of a time step. In order to develop a realistic model of evolution one needs to assume that the real time needed for a spontaneous mutation of species with different fitness dramatically increases with f, going for example as $\exp(fA)$, where A is a large value. Unfortunately, it is impossible to verify the predictions of this model because paleontological records do not provide us with a sufficient statistics.

Fractal Features of DNA Sequences

DNA molecules are probably the largest molecules in nature [139]. Each strand of DNA in large human chromosomes consist of about 10^8 monomers or base-pairs (bp) which are adenine (A), cytosine (C), guanine (G), and thymine (T). The length of this molecule if stretched would reach several centimeters. The geometrical packing of DNA in a cell resembles a self-similar structure with at least 6 levels of packing: a turn of the double helix (10 bp), a nucleosome (200 bp), a unit of 30 nm fiber (6 nucleosomes), a loop domain ($\approx$ 100 units of 30 nm fiber), a turn of a metaphase chromosome ($\approx$ 100 loop domains), a metaphase chromosome ($\approx$ 100 turns). The packing principle is quite similar to the organization of the information in the library: letters form lines, lines form pages, pages form books, books are placed on the shelves, shelves form bookcases, bookcases form rows, and rows are placed in different rooms. This structure however is not a rigorous fractal, because packing of the units on different levels follows different organization principles.

The DNA sequence treated as a sequence of letters also has certain fractal properties [4,5,19,69,70,96]. This sequence can be transformed into a numerical sequence by several mapping rules: for example A rule, in which A is replaced by 1 and C, T, G are replaced by 0, or SW rule in which strongly bonded bp (C and G) are replaced by 1 and weakly bonded bp (A and T) are replaced by -1. Purine-Pyrimidine (RY) mapping rule (A,G: +1; C,T:-1) and KM mapping rule (A,C:+1; G,T -1) are also possible. Power spectra of such sequences display large regions of approximate power law behavior in the range from $f = 10^{-2}$ to $f = 10^{-8}$. For SW mapping rule we has almost perfect $1/f$ noise in the region of low frequencies (Fig. 10). This is not surprising because chromosomes are organized in a large CG rich patches followed by AT rich patches called isochors which extend over millions of bp. The changing slope of the power spectra for different frequency ranges clearly indicates that DNA sequences are also not rigorous fractals but rather a mosaic structures with different organization principles on different length-scales [60,98]. A possible relation between the frac-
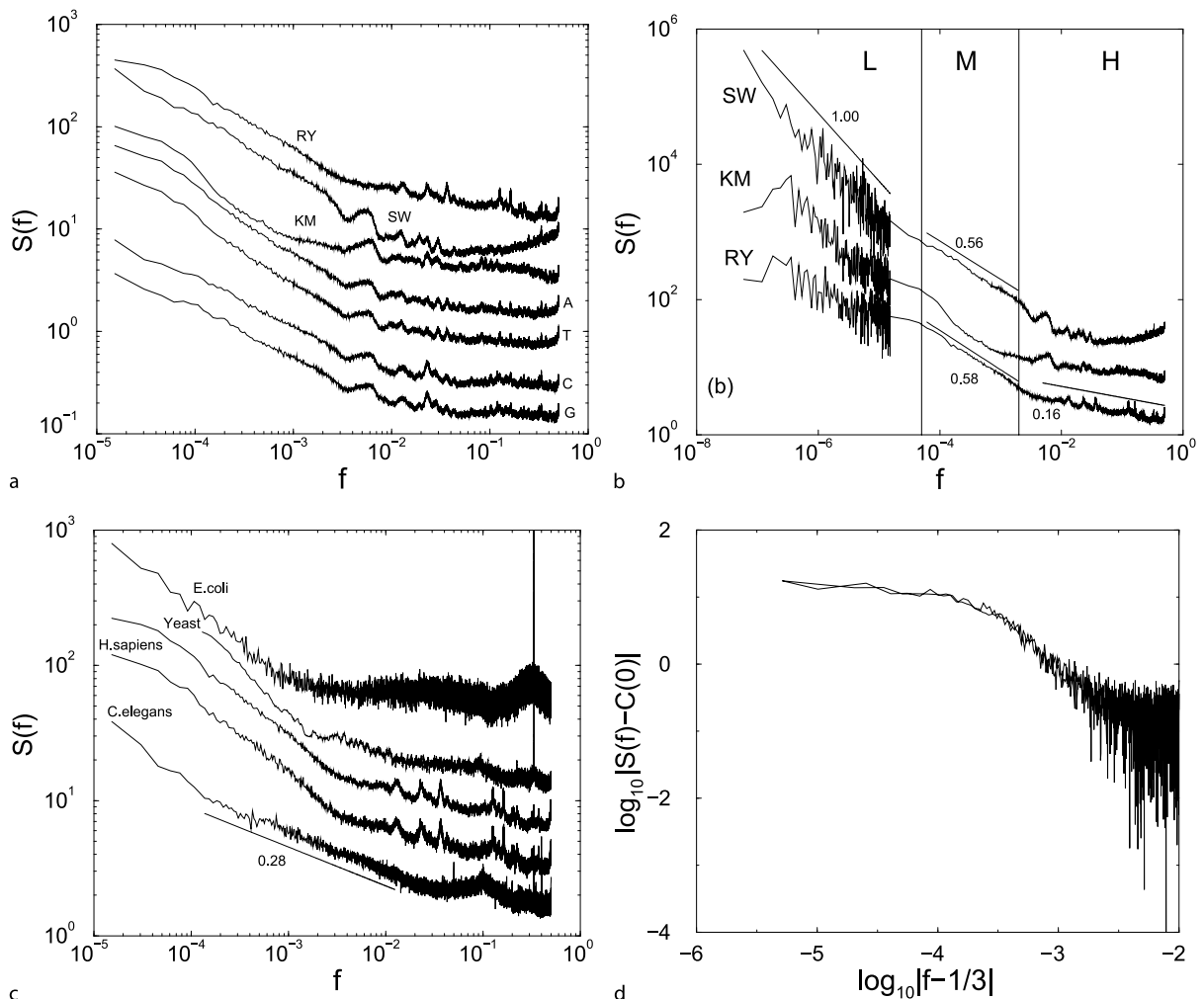
Fractals in Biology, Figure 10
a Power spectra for seven different mapping rules computed for the Homo sapiens chromosome XIV, genomic contig NT_026437. The result is obtained by averaging 1330 power spectra computed by fast Fourier transform for non-overlapping segments of length $N = 2^{16} = 65536$. **b** Power spectra for SW, RY, and KM mapping rules for the same contig extended to the low frequency region characterizing extremely long range correlations. The extension is obtained by extracting low frequencies from the power spectra computed by FFT with $N = 2^{24} \approx 16 \times 10^6$ bp. Three distinct correlation regimes can be identified. High frequency regime ($f < 0.003$) is characterized by small sharp peaks. Medium frequency regime ($0.5 \cdot 10^{-5} < f < 0.003$) is characterized by approximate power-law behavior for RY and SW mapping rules with exponent $\beta_M = 0.57$. Low frequency regime ($f < 0.5 \cdot 10^{-5}$) is characterized by $\beta = 1.00$ for SW rule. The high frequency regime for RY rule can be approximated by $\beta_H = 0.16$ in agreement with the data of Fig. 11. **c** RY Power spectra for the entire genome of *E. coli* (bacteria), *S. cerevisae* (yeast) chromosome IV, *H. sapiens* (human) chromosome XIV and the largest contig (NT_032977.6) on the chromosome I; and *C. elegans* (worm) chromosome X. It can be clearly seen that the high frequency peaks for the two different human chromosomes are exactly the same, while they are totally different from the high frequency peaks for other organisms. These high frequency peaks are associated with the interspersed repeats. One can also notice the presence of enormous peaks for $f = 1/3$ in *E. coli* and yeast, indicating that their genomes do not have introns, so that the lengths of coding segments are very large. The *C. elegans* data can be very well approximated by power law correlations $S(f) \sim f^{-0.28}$ for $10^{-4} < f < 10^{-2}$. **d** Log–log plot of the RY power spectrum for *E. coli* with subtracted white noise level versus $|f - 1/3|$. It shows a typical behavior for a signal with finite correlation length, indicating that the distribution of the coding segments in *E. coli* has finite average square length of approximately 3×10^3 bp
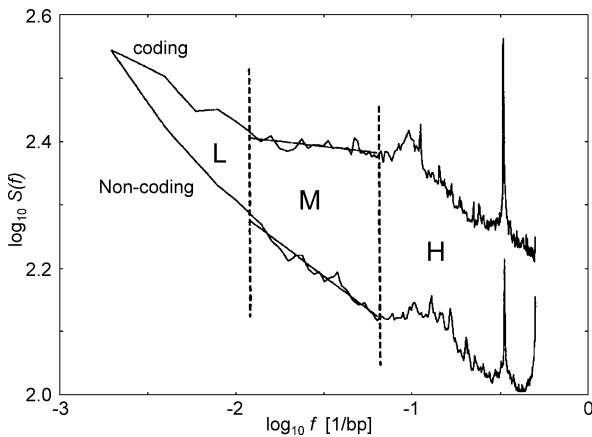
Fractals in Biology, Figure 11

RY Power spectra averaged over all eukaryotic sequences longer than 512 bp, obtained by FFT with window size 512. *Upper curve is average over 29,453 coding sequences; lower curve is average over 33,301 noncoding sequences. For clarity, the power spectra are shifted vertically by arbitrary quantities. The straight lines are* least squares fits for second decade (Region M). The values of β_M for coding and noncoding DNA obtained from the slopes of the fits are 0.03 and 0.21, respectively (from [21])

Fractals in Biology, Figure 12

Comparison of the correlation properties of coding and non-coding DNA of different mammals. Shown RY power spectra averaged over all complimentary DNA sequences of *Homo sapiens* (HS-C) and *Mus musculus* (mouse) (MM-C). The complimentary DNA sequences are obtained from messenger RNA by reverse transcriptase and thus lack non-coding elements. They are characterized by huge peaks at $f = 1/3$, corresponding to the inverse codon length 3 bp. The power spectra for human and mouse are almost indistinguishable. Also shown RY power spectra of large continuously sequenced segments of chromosomes (contigs) of about 10^7 bp long for mouse (MM), human (HS) and chimpanzee (PT, Pan troglodytes). Their power spectra have different high frequency peaks absent in the coding DNA power spectra: a peak at $f = 1/2$, corresponding to simple repeats and several large peaks in the range from $f = 1/3$ to $f = 1/100$ corresponding to interspersed repeats. Notice that the magnitudes of these peaks are similar for humans and chimpanzees (although for humans they are slightly larger, especially the peak at 80 bp corresponding to the long interspersed repeats) and much larger than those of mouse. This means that mouse has much smaller number of the interspersed repeats than primates. On the other hand, mouse has much larger fraction of dimeric simple repeats indicated by the peak at $f = 1/2$

tal correlations of DNA sequences and packing of DNA molecules was suggested in [52,53].

An intriguing property of the DNA of multicellular organisms is that an overwhelming portion of it (97% in case of humans) is not used for coding proteins. It is interesting that the percent of non-coding DNA increases with the complexity of an organism. Bacteria practically do not have non-coding DNA, and Yeast has only 30% of it. The coding sequences form genes ($\approx 10^4$ bp) which carry information for one protein. The genes of multicellular organisms are broken by many noncoding intervening sequences (introns). Only exons which are short ($\approx 10^2$ bp) coding sequences located between introns ($\approx 10^3$ bp) are eventually translated into a protein. The genes themselves are separated by very long intergenic sequences $\approx 10^5$ bp. Thus the coding structure of DNA resembles a Cantor set.

The purpose and properties of coding DNA are well understood. Each three consequent bp form a codon, which is translated into one amino acid of the protein sequence. Accordingly, the power spectrum computed for the coding DNA has a characteristic peak at $f = 1/3$ corresponding to the inverse codon length (Figs. 11 and 12). Coding DNA is highly conserved and the power spectra of coding DNA of different organisms are very similar and in case of mammals are indistinguishable (Fig. 12).

The properties of noncoding DNA are very different. Non coding DNA contains a lot of useful information in-

cluding protein binding sites controlling gene transcription and expression and other regulatory sequences. However its overwhelming fraction lacks any known purpose. This "junk" DNA is full of simple repeats such as CA-CACACA... as well as the interspersed repeats or retroposons which are virus-like sequences inserting themselves in great number of copies into the intergenic DNA. It is a great challenge of molecular biology and genetics to understand the meaning of non-coding DNA and even to learn how to manipulate it. It would be very interesting to create transgenic animals without non-coding DNA and test if their phenotype will differ from that of the wild type species. The non-coding DNA even for very closely

related species can significantly differ (Fig. 12). The length of simple repeats varies even for close relatives. That is why simple repeats are used in forensic studies.

The power spectra of non-coding DNA significantly differ from those of coding DNA. Non-coding DNA does not have a peak at $f = 1/3$. The presence of simple repeats make non-coding DNA more correlated than coding DNA on a scale from 10 to 100 bp [21] (Fig. 11). This difference observed in 1992 [70,96] lead to the hypothesis that noncoding DNA is governed by some mutation-duplication stochastic process which creates long-range (fractal) correlations, while the coding DNA lacks long-range correlations because it is highly conserved. Almost any mutation which happens in coding DNA alter the sequence of the corresponding protein and thus may negatively affect its function and lead to a non-viable or less fit organism. Researchers have proposed using the difference in the long-range correlations to find the coding sequences in the sea of non-coding DNA [91]. However this method appears to be unreliable and today the non-coding sequences are found with much greater accuracy by the bioinformatics methods based on sequence similarity to the known proteins.

Bioinformatics has developed powerful methods for comparing DNA of different species. Even a few hundred bp stretch of the mitochondrial DNA of a Neanderthal man can tell that Neanderthals diverged from humans about 500 000 years ago [65]. The power spectra and other correlation methods such as DFA or wavelet analysis does not have such an accuracy. Never the less, power spectra of the large stretches of DNA carry important information on the evolutionary processes in the DNA. They are similar for different chromosomes of the same organism, but differ even for closely related species (Fig. 12). In this sense they can play a role similar to the role played by X-ray or Raman spectra for chemical substances. A quick look at them can tell a human and a mouse and even a human and a monkey apart. Especially interesting is the difference in peak heights produced by different interspersed repeats. The height of these peaks is proportional to the number of the copies of the interspersed repeats. According to this simple criterion, the main difference between humans and chimps is the insertion of hundreds of thousands of extra copies of interspersed repeats into human DNA [59].

Future Directions

All the above examples show that there are no rigorous fractals in biology. First of all, there are always a lower and an upper cutoff of the fractal behavior. For example, a polymer in a good solvent which is probably the most rigorous of all real fractals in biology has the lower cutoff corresponding to the persistence length comprised of a few monomers and the upper cutoff corresponding to the length of the entire polymer or to the size of the compartment in which it is confined.

For the hierarchical structures such as trees and lungs, in addition to the obvious lower and upper cutoffs, the branching pattern changes from one generation to the next. In the DNA, the packing principles employ different mechanisms on the different levels of packing. In bacterial colonies, the lower cutoff is due to some tendency of bacteria to clump together analogous to surface tension, and the upper cutoff is due to the finite concentration of the nutrients, which originally are uniformly distributed on the Petri dish. In the ideal DLA case, the concentration of the diffusing particles is infinitesimally small, so at any given moment of time there is only one particle in the vicinity of the aggregate.

The temporal physiological series are also not exactly self-similar, but are strongly affected by the daily schedule with a characteristic frequency of 24 h and shorter overtones. Some of these signals can be described with help of multifractals.

Fractals in ecology are limited by the topographical features of the land which may be also fractal due to complex geological processes, so it is very difficult to distinguish whether certain features are caused by biology or geology. The measurements of the fractal dimension are hampered by the lack of statistics, the noise in the image or signal, and by the crossovers due to intrinsic features of the system which is differently organized on different length scales. So very often the mosaic organization of the system with patches of several fixed length scales can be mistakenly identified with the fractal behavior. Thus the use of fractal dimension or Hurst exponent for diagnostics or distinguishing some parts of the system from one another has a limited value. After all, the fractal dimension is only one number which is usually obtained from a slope of a least square linear fit of a log–log graph over a subjectively identified range of values. Other features of the power spectrum, such as peaks at certain characteristic frequencies may have more biological information than fractal dimension. Moreover, the presence of certain fractal behavior may originate from simple physical principles, while the deviation from it may indicate the presence of a nontrivial biological phenomenon.

On the other hand, fractal geometry is an important concept which can be used for qualitative understanding of the mechanisms behind certain biological processes. The use of similar organization principles on different length scales is a remarkable feature, which is certainly em-

ployed by nature to design the shape and behavior of living organisms.

Though fractals themselves may have a limited value in biology, the theory of complex systems in which they often emerge continues to be a leading approach in understanding life. One of the most impelling challenges of modern interdisciplinary science which involves biology, chemistry, physics, mathematics, computer science, and bioinformatics is to build a comprehensive theory of morphogenesis. Such a great mind as de Gennes turned to this subject in his late years [35,36,37]. In these studies the theory of complex networks [2,12,48,67,90] which describe interactions of biomolecules with many complex positive and negative feedbacks will take a leading part.

Challenges of the same magnitude face researches in neuroscience, behavioral science, and ecology. Only complex interdisciplinary approach involving the specialists in the theory of complex systems may lead to new breakthroughs in this field.

Bibliography

1. Aiello L, Dean C (1990) An Introduction to Human Evolutionary Anatomy. Academic Press, London
2. Albert R, Barabási A-L (2002) Statistical mechanics of complex networks. Rev Mod Phys 74:47–97
3. Alencar AM, Buldyrev SV, Majumdar A, Stanley HE, Suki B (2003) Perimeter growth of a branched structure: application to crackle sounds in the lung. Phys Rev E 68:11909
4. Arneodo A, Bacry E, Graves PV, Muzy JF (1995) Characterizing long-range correlations in DNA sequences from wavelet analysis. Phys Rev Lett 74:3293–3296
5. Arneodo A, D'Aubenton-Carafa Y, Audit B, Bacry E, Muzy JF, Thermes C (1998) What can we learn with wavelets about DNA sequences. Physica A 249:439–448
6. Ashkenazy Y, Hausdorff JM, Ivanov PC, Stanley HE (2002) A stochastic model of human gait dynamics. Physica A 316:662–670
7. Bak P, Sneppen K (1993) Punctuated equilibrium and criticality in a simple model of evolution. Phys Rev Lett 71:4083–4086
8. Bak P, Tang C, Wiesenfeld K (1987) Self-organized criticality: an explanation of $1/f$ noise. Phys Rev Lett 59:381–384
9. Banavar JR, Green JL, Harte J, Maritan A (1999) Finite size scaling in ecology. Phys Rev Lett 83:4212–4214
10. Banavar JR, Damuth J, Maritan A, Rinaldo A (2002) Supply-demand balance and metabolic scaling. Proc Natl Acad Sci USA 99:10506–10509 (2002)
11. Banavar JR, Damuth J, Maritan A, Rinaldo A (2003) Allometric cascades. Nature 421:713–714
12. Barabási A-L (2005) The origin of bursts and heavy tails in human dynamics. Nature 435:207–211
13. Barabási A-L, Stanley HS (1995) Fractal Concepts in Surface Growth. Cambridge University Press, Cambridge
14. Bassingthwaighte JB, Liebovitch L, West BJ (1994) Fractal Physiology. Oxford University Press, Oxford
15. Ben Jacob E, Aharonov Y, Shapira Y (2005) Bacteria harnessing complexity. Biofilms 1:239–263
16. Bernaola-Galvan P, Ivanov PC, Amaral LAN, Stanley HE (2001) Scale invariance in the nonstationarity of human heart rate. Phys Rev Lett 87:168105
17. Bhaskar KR, Turner BS, Garik P, Bradley JD, Bansil R, Stanley HE, LaMont JT (1992) Viscous fingering of HCl through gastric mucin. Nature 360:458–461
18. Bishop M, Michels JPJ (1985) The shape of ring polymers. J Chem Phys 82:1059–1061
19. Buldyrev SV (2006) Power Law Correlations in DNA Sequences. In: Koonin EV, Karev G, Wolf Yu (eds) Power Laws, Scale-Free Networks and Genome Biology. Springer, Berlin, pp 123–164
20. Buldyrev SV, Goldberger AL, Havlin S, Peng C-K, Stanley HE (1994) Fractals in Biology and Medicine: From DNA to the Heartbeat. In: Bunde A, Havlin S (eds) Fractals in Science. Springer, Berlin, pp 48–87
21. Buldyrev SV, Goldberger AL, Havlin S, Mantegna RN, Matsa ME, Peng C-K, Simons M, Stanley HE (1995) Long-range correlation properties of coding and noncoding DNA sequences: GenBank analysis. Phys Rev E 51:5084–5091
22. Buldyrev SV, Cruz L, Gomez-Isla T, Havlin S, Stanley HE, Urbanc B, Hyman BT (2000) Description of microcolumnar ensembles in association cortex and their disruption in alzheimer and lewy body dementia. Proc Natl Acad Sci USA 97:5039–5043
23. Buldyrev SV, Havlin S, Kazakov AY, da Luz MGE, Raposo EP, Stanley HE, Viswanathan GM (2001) Average time spent by levy flights and walks on an interval with absorbing boundaries. Phys Rev E 64:041108
24. Buldyrev SV, Gitterman M, Havlin S, Kazakov AY, da Luz MGE, Raposo EP, Stanley HE, Viswanathan GM (2001) Properties of Levy fights on an interval with absorbing boundaries. Physica A 302:148–161
25. Buldyrev SV, Pammolli F, Riccaboni M, Yamasaki K, Fu D-F, Matia K, Stanley HE (2007) Generalized preferential attachment model for business firms growth rates II. Eur Phys J B 57:131–138
26. Bunde A, Havlin S (eds) (1996) Fractals and Disordered Systems, 2nd edn. Springer, New York
27. Bunde A, Havlin S, Kantelhardt J, Penzel T, Peter J-H, Voigt K (2000) Correlated and uncorrelated regions in heart-rate fluctuations during sleep. Phys Rev Lett 85:3736–3739
28. Calder WA 3rd (1984) Size, Function and Life History. Harvard University Press, Cambridge
29. Caserta F, Eldred WD, Fernández E, Hausman RE, Stanford LR, Bulderev SV, Schwarzer S, Stanley HE (1995) Determination of fractal dimension of physiologically characterized neurons in two and three dimensions. J Neurosci Meth 56:133–144
30. Clar S, Drossel B, Schwabl F (1996) Self-organized criticality in forest-fire models and elsewhere. Review article. J Phys C 8:6803
31. D'Arcy WT, John TB (ed) (1992) On Growth and Form. Cambridge University Press, Cambridge
32. Darveau C-A, Suarez RK, Andrews RD, Hochachka PW (2002) Allometric cascade as a unifying principle of body mass effects on metabolism. Nature 417:166–170
33. Darveau C-A, Suarez RK, Andrews RD, Hochachka PW (2003) Allometric cascades – Reply. Nature 421:714–714
34. De Gennes PG (1979) Scaling Concepts in Polymer Physics. Cornell University Press, Ithaca

35. De Gennes PG (2004) Organization of a primitive memory: Olfaction. Proc Natl Acad Sci USA 101:15778–15781
36. De Gennes PG (2007) Collective neuronal growth and self organization of axons. Proc Natl Acad Sci 104:4904–4906
37. De Gennes PG, Puech PH, Brochard-Wyart F (2003) Adhesion induced by mobile stickers: A list of scenarios. Langmuir 19:7112–7119
38. Devaney RL (1989) An Introduction to Chaotic Dynamical Systems, 2nd edn. Addison-Wesley, Redwood City
39. Edwards AM, Phillips RA, Watkins NW, Freeman MP, Murphy EJ, Afanasyev V, Buldyrev SV, da Luz MGE, Raposo EP, Stanley HE, Viswanathan GM (2007) Revisiting Levy flight search patterns of wandering albatrosses, bumblebees and deer. Nature 449:1044–1047
40. Falconer K (2003) Fractal Geometry: Mathematical Foundations and Applications. Wiley, Hoboken
41. Feder J (1988) Fractals. Plenum, New York
42. Feller W (1970) An introduction to ptobability theory and its applications, vol 1–2. Wiley, New York
43. Fleury V, Gouyet J-F, Leonetti M (eds) (2001) Branching in Nature: Dynamics and Morphogenesis. Springer, Berlin
44. Flory PJ (1955) Principles of Polymer Chemistry. Cornell University Press, Ithaca
45. Fortin MJ, Dale MRT (2005) Spatial analysis: a guide for ecologists. Cambridge Univ Press, Cambridge
46. Gerstein GL, Mandelbrot B (1964) Random walk models for the spike activity of a single neuron. Biophys J 4:41–68
47. Goldberger AL, Amaral LAN, Hausdorff JM, Ivanov PC, Peng C-K, Stanley HE (2002) Fractal dynamics in physiology: Alterations with disease and aging. Proc Natl Acad Sci USA 99:2466–2472
48. Gonzalez MC, Barabási A-L (2007) Complex networks – From data to models. Nature Phys 3:224–225
49. Gonzalez MC, Hidalgo CA, Barabási A-L (2008) Understanding individual human mobility patterns. Nature 453:779–782
50. Grassberger P, Procaccia I (1983) Measuring the strangeness of strange attractors. Physica D 9:189–208
51. Grey F, Kjems JK (1989) Aggregates, broccoli and cauliflower. Physica D 38:154–159
52. Grosberg A, Rabin Y, Havlin S, Neer A (1993) Crumpled globule model of three-dimensional structure of DNA. Europhys Lett 23:373–378
53. Grosberg A, Rabin Y, Havlin S, Neer A (1993) Self-similarity in the structure of DNA: why are introns needed? Biofizika 38:75–83
54. Halley JM, Hartley S, Kallimanis AS, Kunin WE, Lennon JJ, Sgardelis SP (2004) Uses and abuses of fractal methodology in ecology. Ecol Lett 7:254–271
55. Hausdorff JM, Ashkenazy Y, Peng CK, Ivanov PC, Stanley HE, Goldberger AL (2001) When human walking becomes random walking: Fractal analysis and modeling of gait rhythm fluctuations. Physica A 302:138–147
56. Horsfield K, Thurlbeck A (1981) Relation between diameter and flow in branches of the bronchial tree. Bull Math Biol 43:681–691
57. Hu K, Ivanov PC, Hilton MF, Chen Z, Ayers RT, Stanley HE, Shea SA (2004) Endogenous circadian rhythm in an index of cardiac vulnerability independent of changes in behavior. Proc Natl Acad Sci USA 101:18223–18227
58. Hughes BD (1995) Random Walks and Random Environments, vol 1: Random Walks. Clarendon Press, Oxford
59. Hwu RH, Roberts JW, Dawidson EH, et al. (1986) Insertion and/or deletion of many repeated DNA sequences in human and higher apes evolution. Proc Natl Acad Sci 83:3875–3879
60. Karlin S, Brandel V (1993) Patchiness and correlations in DNA sequences. Science 259:677–680
61. Kenah E, Robins MJ (2007) Second look at the spread of epidemics on networks. Phys Rev E 76:036113
62. Kermack WO, McKendrick AG (1927) Contribution to the Mathematical Theory of Epidemics. Proc R Soc A 115:700–721
63. Khokhlov AR, Grosberg AY (2002) Statistical physics of macromolecules. AIP, Woodbury
64. Kim S-H (2005) Fractal structure of a white cauliower. J Korean Phys Soc 46:474–477
65. Kings M, Geisert H, Schmitz RW, Krainitzki H, Pääbo S (1999) DNA sequence of the mitochondrial hypervariable region II from the neandertal type specimen. Proc Natl Acad Sci USA 96:5581–5585
66. Kitaoka H, Suki B (1997) Branching design of the bronchial tree based on a diameter-flow relationship. J Appl Physiol 82:968–976
67. Koonin EV, Karev G, Wolf Yu (eds) (2006) Power Laws, Scale-Free Networks and Genome Biology. Springer, Berlin
68. Larralde H, Trunfio PA, Havlin S, Stanley HS, Weiss GH (1992) Territory covered by N diffusing particles. Nature 355:423–426
69. Li W (1997) The study of correlation structures of DNA sequences: a critical review. Comput Chem 21:257–271
70. Li W, Kaneko K (1992) Long-range correlation and partial 1/F-alpha spectrum in a noncoding DNA-sequence. Europhys Lett 17:655–660
71. Liljeros F, Edling CR, Amaral LAN, Stanley HE, Aberg Y (2001) The web of human sexual contacts. Nature 411:907–908
72. Lo C-C, Amaral LAN, Havlin S, Ivanov PC, Penzel T, Peter J-H, Stanley HE (2002) Dynamics of sleep-wake transitions during sleep. Europhys Lett 57:625–631
73. Lo C-C, Chou T, Penzel T, Scammell T, Strecker RE, Stanley HE, Ivanov PC (2004) Common scale-invariant pattern of sleep-wake transitions across mammalian species. Proc Natl Acad Sci USA 101:17545–17548
74. Losa GA, Merlini D, Nonnenmacher TF, Weibel ER (eds) (1998) Fractals in Biology and Medicine, vol II. Birkhäuser Publishing, Berlin
75. Losa GA, Merlini D, Nonnenmacher TF, Weibel ER (eds) (2002) Fractals in Biology and Medicine, vol III. Birkhäuser Publishing, Basel
76. Losa GA, Merlini D, Nonnenmacher TF, Weibel ER (eds) (2005) Fractals in Biology and Medicine, vol IV. Birkhäuser Publishing, Basel
77. Majumdar A, Alencar AM, Buldyrev SV, Hantos Z, Lutchen KR, Stanley HE, Suki B (2005) Relating airway diameter distributions to regular branching asymmetry in the lung. Phys Rev Lett 95:168101
78. Malamud BD, Morein G, Turcotte DL (1998) Forest fires: an example of self-organized critical behavior. Science 281:1840–1841
79. Mandelbrot BB (1982) The Fractal Geometry of Nature. Freeman WH and Co., New York
80. Mark DM (1984) Fractal dimension of a coral-reef at ecological scales – a discussion. Mar Ecol Prog Ser 14:293–294
81. May RM (1975) In: Cody ML, Diamond JM (eds) Ecology

and Evolution of Communities. Belknap Press, Cambridge, pp 81–120

82. May RM (1976) Simple mathematical models with very complicated dynamics. Nature 261:459–467

83. Meakin P (1998) Fractals, Scaling and Growth Far from Equilibrium. Cambridge University Press, Cambridge

84. Menshutin AY, Shchur LN, Vinokur VM (2007) Probing surface characteristics of diffusion-limited-aggregation clusters with particles of variable size. Phys Rev E 75:010401

85. Metzger RJ, Krasnow MA (1999) Genetic Control of Branching Morphogenesis Science 284:1635–1639

86. Newman MEJ (2002) Spread of epidemic disease on networks. Phys Rev E 66:016128

87. Niklas KJ (2007) Sizing up life and death. PNAS 104:15589–15590

88. Nonnenmacher TF, Losa GA, Weibel ER (eds) (1994) Fractals in Biology and Medicine, vol I. Birkhäuser Publishing, Basel

89. Olami Z, Feder HJS, Christensen K (1992) Self-organized criticality in a continuous, nonconservative cellular automaton modeling earthquakes. Phys Rev Lett 68:1244–1247

90. Oliveira JG, Barabási A-L (2005) Human dynamics: Darwin and Einstein correspondence patterns. Nature 437:1251–1251

91. Ossadnik SM, Buldyrev SV, Goldberger AL, Havlin S, Mantegna RN, Peng C-K, Simons M, Stanley HE (1994) Correlation approach to identify coding regions in DNA sequences. Biophys J 67:64–70

92. Paczuski M, Maslov S, Bak P (1996) Avalanche dynamics in evolution, growth, and depinning models. Phys Rev E 53:414–343

93. Panico J, Sterling P (1995) Retinal neurons and vessels are not fractal but space-filling. J Comparat Neurol 361:479–490

94. Peitgen H-O, Saupe D (ed) (1988) The Science of Fractal Images. Springer, Berlin

95. Peitgen H-O, Jiirgens H, Saupe D (1992) Chaos and Fractals. Springer, New York

96. Peng C-K, Buldyrev SV, Goldberger A, Havlin S, Sciortino F, Simons M, Stanley HE (1992) Long-range correlations in nucleotide sequences. Nature 356:168–171

97. Peng C-K, Mietus J, Hausdorff JM, Havlin S, Stanley HE, Goldberger AL (1993) Long-range anticorrelations and non-gaussian behavior of the heartbeat. Phys Rev Lett 70:1343–1347

98. Peng C-K, Buldyrev SV, Havlin S, Simons M, Stanley HE, Goldberger AL (1994) Mosaic organization of DNA nucleotides. Phys Rev E 49:1685–1689

99. Peng C-K, Buldyrev SV, Hausdorff JM, Havlin S, Mietus JE, Simons M, Stanley HE, Goldberger AL (1994) Non-equilibrium dynamics as an indispensable characteristic of a healthy biological system. Integr Physiol Behav Sci 29:283–293

100. Peng C-K, Havlin S, Stanley HE, Goldberger AL (1995) Quantification of scaling exponents and crossover phenomena in nonstationary heartbeat time series. Chaos 5:82–87

101. Pocock MJO, Hartley S, Telfer MG, Preston CD, Kunin WE (2006) Ecological correlates of range structure in rare and scarce British plants. J Ecol 94:581–596

102. Ramchandani R, Bates JHT, Shen X, Suki B, Tepper RS (2001) Airway branching morphology of mature and immature rabbit lungs. J Appl Physiol 90:1584–1592

103. Ramos-Fernandez G, Meteos JL, Miramontes O, Cocho G, Larralde H, Ayala-Orozco B (2004) Lévy walk patterns in the foraging movements of spider monkeys (Ateles geoffroyi). Behav Ecol Sociobiol 55:223–230

104. Redner S (2001) A Guide to First-Passage Processes. Cambridge University Press, Cambridge

105. Reich PB, Tjoelker MG, Machado J-L, Oleksyn J (2006) Universal scaling of respiratory metabolism, size and nitrogen in plants. Nature 439:457–461

106. Reiss MJ (2006) Allometry of Growth and Reproduction. Cambridge University Press, Cambridge

107. Richter JP (ed) (1970) The Notebooks of Leonardo da Vinci. Dover Publications, New York

108. Rosenzweig ML (1995) Species Diversity in Space and Time. Cambridge University Press, Cambridge

109. Santullia A, Telescaa L (2005) Time-clustering analysis of forest-fire sequences in southern Italy Rosa Lasaponaraa. Chaos Solitons Fractals 24:139–149

110. Savage VM, Allen AP, Brown JH, Gillooly JF, Herman AB, Woodruff WH, West GB (2007) Scaling of number, size, and metabolic rate of cells with body size in mammals. Proc Natl Acad Sci 104:4718–4723

111. Schenk K, Drossel B, Schwabl F (2002) The self-organized critical forest-fire model on large scales. Phys Rev E 65:026135

112. Schmidt-Nielsen K (1984) Scaling: Why is Animal Size so Important? Cambridge University Press, Cambridge

113. Schulte-Frohlinde V, Ashkenazy Y, Ivanov PC, Glass L, Goldberger AL, Stanley HE (2001) Noise effects on the complex patterns of abnormal heartbeats. Phys Rev Lett 87:068104

114. Scott FG (2006) Developmental Biology, 8th edn. Sinauer Associates, Sunderland

115. Shao J, Buldyrev SV, Cohen R, Kitsak M, Havlin S, Stanley HE (2008) Fractal boundaries of complex networks. Preprint

116. Shlesinger MF (1986) In: Stanley HE, Ostrowsky N (eds) On Growth and Form. Nijhoff, Dordrecht

117. Shlesinger MF, West BJ (1991) Complex fractal dimension of the bronchial tree. Phys Rev Lett 67:2106–2109

118. Shlesinger MF, Zaslavsky G, Frisch U (eds) (1995) Lévy Flights and Related Topics in Physics. Springer, Berlin

119. Sims DW et al (2008) Scaling laws in marine predator search behavior. Nature 451:1098–1102

120. Song C, Havlin H, Makse H (2005) Self-similarity of complex networks. Nature 433:392–395

121. Song C, Havlin S, Makse HA (2006) Origins of fractality in the growth of complex networks. Nature Phys 2:275–281

122. Sornette D (2003) Critical Phenomena in Natural Sciences. Chaos, Fractals, Selforganization and Disorder: Concepts and Tools, 2nd edn. Springer, Berlin

123. Sprott JC (2003) Chaos and Time-Series Analysis. Oxford University Press

124. Stanley HE (1971) Introduction to Phase Transitions and Critical Phenomena. Oxford University Press, New York

125. Stanley HE, Ostrowsky N (eds) (1986) On Growth and Form. Nijhoff, Dordrecht

126. Stanley HE, Buldyrev SV, Goldberger AL, Goldberger ZD, Havlin S, Mantegna RN, Ossadnik SM, Peng C-K, Simons M (1994) Statistical mechanics in biology – how ubiquitous are long-range correlations. Physica A 205:214–253

127. Stauffer D, Aharony A (1992) Introduction to percolation theory. Taylor & Francis, Philadelphia

128. Suki B, Barabási A-L, Hantos Z, Petak F, Stanley HE (1994) Avalanches and power law behaviour in lung inflation. Nature 368:615–618

129. Suki B, Alencar AM, Frey U, Ivanov PC, Buldyrev SV, Majumdar A, Stanley HE, Dawson CA, Krenz GS, Mishima M (2003) Fluc-

tuations, noise, and scaling in the cardio-pulmonary system. Fluct Noise Lett 3:R1–R25
130. Turcotte DL (1997) Fractals and Chaos in Geology and Geophysics. Cambridge University Press, Cambridge
131. Viscek T, Shlesinger MF, Matsushita M (eds) (1994) Fractals in Natural Sciences. World Scientific, New York
132. Viswanathan GM, Afanasyev V, Buldyrev SV, Murphy EJ, Prince PA, Stanley HE (1996) Levy flight search patterns of wandering albatrosses. Nature 381:413–415
133. Viswanathan GM, Buldyrev SV, Havlin S, da Luz MGE, Raposo E, Stanley HE (1999) Optimizing the success of random searches. Nature 401:911–914
134. Vogel H (1979) A better way to construct the sunflower head. Math Biosci 44:179–189
135. Voss RF, Clarke J (1975) 1/f noise in music and speech. Nature 258:317–318
136. Voss RF, Clarke J (1978) 1/f noise in music: Music from 1/f noise. J Acoust Soc Am 63:258–263
137. Voss RF, Wyatt J (1993) Multifractals and the Local Connected Fractal Dimension: Classification of Early Chinese Landscape Paintings. In: Crilly AJ, Earnshaw RA, Jones H (eds) Applications of Fractals and Chaos. Springer, Berlin
138. Warburton D, Schwarz M, Tefft D, Flores-Delgado F, Anderson KD, Cardoso WV (2000) The molecular basis of lung morphogenesis. Mech Devel 92:55–81
139. Watson JD, Gilman M, Witkowski J, Zoller M (1992) Recombinant DNA. Scientific American Books, New York
140. Weibel ER (2000) Symmorphosis: On Form and Function in Shaping Life. Harvard University Press, Cambridge
141. Weiss GH (1994) Aspects and applications of the random walk. North-Holland, New York
142. West GB, Brown JH, Enquist BJ (1997) A general model for the origin of allometric scaling laws in biology. Science 276: 122–126
143. West GB, Woodruff WH, Brown JH (2002) Allometric scaling of metabolic rate from molecules and mitochondria to cells and mammals. Proc Natl Acad Sci 99:2473–2478
144. West GB, Savage VM, Gillooly J, Enquist BJ, Woodruff WH, Brown JH (2003) Why does metabolic rate scale with body size? Nature 421:713–713
145. Witten TA, Sander LM (1981) Phys Rev Lett 47:1400–1404
146. Zipf GK (1949) Human Behavior and the Principle of Least–Effort. Addison-Wesley, Cambridge
147. Zolotarev VM, Uchaikin VV (1999) Chance and Stability: Stable Distributions and their Applications. VSP BV, Utrecht

Fractals and Economics

Misako Takayasu[1], Hideki Takayasu[2]
[1] Tokyo Institute of Technology, Tokyo, Japan
[2] Sony Computer Science Laboratories Inc, Tokyo, Japan

Article Outline

Glossary

Fractal An adjective or a noun representing complex configurations having scale-free characteristics or self-similar properties. Mathematically, any fractal can be characterized by a power law distribution.

Power law distribution For this distribution the probability density is given by a power law, $p(r) = c \cdot r^{-\alpha-1}$, where c and α are positive constants.

Foreign exchange market A free market of currencies, exchanging money in one currency for other, such as purchasing a United States dollar (USD) with Japanese yen (JPY). The major banks of the world are trading 24 hours and it is the largest market in the world.

Definition of the Subject

Market price fluctuation was the very first example of fractals, and since then many examples of fractals have been found in the field of Economics. Fractals are everywhere in economics. In this article the main attention is focused on real world examples of fractals in the field of economics, especially market properties, income distributions, money flow, sales data and network structures. Basic mathematics and physics models of power law distributions are reviewed so that readers can start reading without any special knowledge.

Introduction

Fractal is the scientific word coined by B.B. Mandelbrot in 1975 from the Latin word *fractus*, meaning "fractured" [25]. However, *fractal* does not directly mean fracture itself. As an image of a fractal Fig. 1 shows a photo of fractured pieces of plaster fallen on a hard floor. There are several large pieces, many middle size pieces and countless fine pieces. If you have a microscope and observe a part of floor carefully then you will find in your vision several large pieces, many small pieces and countless fine pieces, again in the microscopic world. Such scale-invariant nature is the heart of the fractal. There is no explicit definition on the word fractal, it generally means a complicated scale-invariant configuration.

Scale-invariance can be defined mathematically [42]. Let $P(\geq r)$ denote the probability that the diameter of

Fractals and Economics, Figure 1
Fractured pieces of plaster fallen on a hard floor (provided by H. Inaoka)

Fractals and Economics, Figure 2
Sierpinski gasket

a randomly chosen fractured piece is larger than r, then this distribution is called scale-invariant if this function satisfies the following proportional relation for any positive scale factor λ in a considering scale range:

$$P(\geq \lambda r) \propto P(\geq r) . \tag{1}$$

The proportional factor should be a function of λ, so we can re-write Eq. (1) as

$$P(\geq \lambda r) = C(\lambda)P(\geq r) . \tag{2}$$

Assuming that $P(\geq r)$ is a differentiable function, and differentiate Eq. (2) by λ, and then let $\lambda = 1$.

$$rP'(\geq r) = C'(1)P(\geq r) \tag{3}$$

As $C'(1)$ is a constant this differential equation is readily integrated as

$$P(\geq r) = c_0 r^{C'(1)} . \tag{4}$$

$P(\geq r)$ is a cumulative distribution and it is a non-increasing function in general, the exponent $C'(1)$ can be replaced by $-\alpha$ where α is a positive constant. Namely, from the scale-invariance with the assumption of differentiability we have the following power law:

$$P(\geq r) = c_0 r^{-\alpha} . \tag{5}$$

The reversed logic also holds, namely for any power law distribution there is a fractal configuration or a scale-invariant state.

In the case of real impact fracture, the size distribution of pieces is experimentally obtained by repeating sieves of various sizes, and it is empirically well-known that a fractured piece's diameter follows a power law with the exponent about $\alpha = 2$ independent of the details about the material or the way of impact [14]. This law is one of the most stubborn physical laws in nature as it is known to hold from 10^{-6} m to 10^5 m, from glass pieces around us to asteroids. From theoretical viewpoint this phenomenon is known to be described by a scale-free dynamics of crack propagation and the universal properties of the exponent value are well understood [19].

Usually fractal is considered geometric concept introducing the quantity fractal dimension or the concept of self-similarity. However, in economics there are very few geometric objects, so, the concept of fractals in economics are mostly used in the sense of power law distributions.

It should be noted that any geometrical fractal object accompanies a power law distribution even a deterministic fractal such as Sierpinski gasket. Figure 2 shows Sierpinski gasket which is usually characterized by the fractal dimension D given by

$$D = \frac{\log 3}{\log 2} . \tag{6}$$

Paying attention to the distribution of length r of white triangles in this figure, it is easy to show that the probability that a randomly chosen white triangle's side is larger than r, $P(\geq r)$, follows the power law,

$$P(\geq r) \propto r^{-\alpha} , \quad \alpha = D = \frac{\log 3}{\log 2} . \tag{7}$$

Here, the power law exponent of distribution equals to the fractal dimension; however, such coincidence occurs only when the considering distribution is for a length distribution. For example, in Sierpinski gasket the area s of white triangles follow the power law,

$$P(\geq s) \propto s^{-\alpha} , \quad \alpha = \frac{\log 3}{\log 4} . \tag{8}$$

The fractal dimension is applicable only for geometric fractals, however, power law distributions are applicable for any fractal phenomena including shapeless quantities. In such cases the power law exponent is the most important quantity for quantitative characterization of fractals.

According to Mandelbrot's own review on his life the concept of fractal was inspired when he was studying economics data [26]. At that time he found two basic properties in the time series data of daily prices of New York cotton market [24]:

(A) Geometrical similarity between large scale chart and an expanded chart.
(B) Power law distribution of price changes in a unit time interval, which is independent of the time scale of the unit.

He thought such scale invariance in both shape and distribution is a quite general property, not only in price charts but also in nature at large. His inspiration was correct and the concept of fractals spread over physics first and then over almost all fields of science. In the history of science it is a rare event that a concept originally born in economics has been spread widely to all area of sciences.

Basic mathematical properties of cumulative distribution can be summarized as follows (here we consider distribution of non-negative quantity for simplicity):

1. $P(\geq 0) = 1$, $P(\geq \infty) = 0$.
2. $P(\geq r)$ is a non-increasing function of r.
3. The probability density is given as $p(r) \equiv -\frac{d}{dr} P(\geq r)$. As for power law distributions there are three peculiar characteristics:
4. Difficulty in normalization. Assuming that $P(\geq r) = c_0 r^{-\alpha}$ for all in the range $0 \leq r < \infty$, then the normalization factor c_0 must be 0 considering the limit of $r \to 0$. To avoid this difficulty it is generally assumed that the power law does not hold in the vicinity of $r = 0$. In the case of observing distribution from real data there are naturally lower and upper bounds, so this difficulty should be necessary only for theoretical treatment.
5. Divergence of moments. As for moments defined by $\langle r^n \rangle \equiv \int_0^\infty r^n p(r) dr$, $\langle r^n \rangle = \infty$ for $n \geq \alpha$. In the special case of $2 \geq \alpha > 0$ the basic statistical quantity, the variance, diverges, $\sigma^2 \equiv \langle r^2 \rangle - \langle r \rangle^2 = \infty$. In the case of $1 \geq \alpha > 0$ even the average can not be defined as $\langle r \rangle = \infty$.
6. Stationary or non-stationary? In view of the data analysis, the above characteristics of diverging moments is likely to cause a wrong conclusion that the phenomenon is non-stationary by observing its averaged

value. For example, assume that we observe k samples $\{r_1, r_2, \ldots, r_k\}$ independently from the power law distribution with the exponent, $1 \geq \alpha > 0$. Then, the sample average, $\langle r \rangle_k \equiv \frac{1}{k}\{r_1 + r_2 + \cdots + r_k\}$, is shown to diverge as, $\langle r \rangle_k \propto k^{1/\alpha}$. Such tendency of monotonic increase of averaged quantity might be regarded as a result of non-stationarity, however, this is simply a general property of a power law distribution. The best way to avoid such confusion is to observe the distribution directly from the data.

Other than the power law distribution there is another important statistical quantity in the study of fractals, that is, the autocorrelation. For given time series, $\{x(t)\}$, the autocorrelation is defined as,

$$C(T) \equiv \frac{\langle x(t+T)x(t) \rangle - \langle x(t) \rangle^2}{\langle x(t)^2 \rangle - \langle x(t) \rangle^2}, \tag{9}$$

where $\langle \cdots \rangle$ denotes an average over realizations. The autocorrelation can be defined only for stationary time series with finite variance, in which any statistical quantities do not depend on the location of the origin of time axis.

For any case, the autocorrelation satisfies the following basic properties,

1. $C(0) = 1$ and $C(\infty) = 0$
2. $|C(T)| \leq 1$ for any $T \geq 0$.
3. The Wiener–Khinchin theorem holds, $C(T) = \int_0^\infty S(f) \cos 2\pi f df$, where $S(f)$ is the power spectrum defined by $S(f) \equiv \langle \hat{x}(f)\hat{x}(-f) \rangle$, with the Fourier transform, $\hat{x}(f) \equiv \int x(t) e^{2\pi i f t} dt$.

In the case that the autocorrelation function is characterized by a power law, $C(T) \propto T^{-\beta}$, $\beta > 0$, then the time series $\{x(t)\}$ is said to have a fractal property, in the sense that the autocorrelation function is scale-independent for any scale-factor, $\lambda > 0$, $C(\lambda T) \propto C(T)$. In the case $1 > \beta > 0$ the corresponding power spectrum is given as $S(f) \propto f^{-1+\beta}$.

The power spectrum can be applied to any time series including non-stationary situations. A simple way of telling non-stationary situation is to check the power law exponent of $S(f) \propto f^{-1+\beta}$ in the vicinity of $f = 0$, for $0 > \beta$ the time series is non-stationary.

Three basic examples of fractal time series are the followings:

1. White noise. In the case that $\{x(t)\}$ is a stationary independent noise, the autocorrelation is given by the Kronecker's function, $C(T) = \delta_T$, where

$$\delta_T = \begin{cases} 1, & T = 0 \\ 0, & T \neq 0. \end{cases}$$

The corresponding power spectrum is $S(f) \propto f^0$. This case is called white noise from an analogy that superposition of all frequency lights with the same amplitude make a colorless white light. White noise is a plausible model of random phenomena in general including economic activities.

2. Random walk. This is defined by summation of a white noise, $X(t) = X(0) + \sum_{s=0}^{t} x(s)$, and the power spectrum is given by $S(f) \propto f^{-2}$. In this case the autocorrelation function can not be defined because the data is non-stationary. Random walks are quite generic models widely used from Brownian motions of colloid to market prices. The graph of a random walk has a fractal property such that an expansion of any part of the graph looks similar to the whole graph.

3. The $1/f$ noise. The boundary of stationary and nonstationary states is given by the so-called $1/f$ noise, $S(f) \propto f^{-1}$. This type of power spectrum is also widely observed in various fields of sciences from electrical circuit noise [16] to information traffics in the Internet [53]. The graph of this $1/f$ noise also has the fractal property.

Examples in Economics

In this chapter fractals observed in real economic activities are reviewed. Mathematical models derived from these empirical findings will be summarized in the next chapter.

As mentioned in the previous chapter the very first example of a fractal was the price fluctuation of the New York cotton market analyzed by Mandelbrot with the daily data for a period of more than a hundred years [24]. This research attracted much attention at that time, however, there was no other good market data available for scientific analysis, and no intensive follow-up research was done until the 1990s. Instead of earnest scientific data analysis artificial mathematical models of market prices based on random walk theory became popular by the name of Financial Technology during the years 1960–1980.

Fractal properties of market prices are confirmed with huge amount of high resolution market data since the 1990s [26,43,44]. This is due to informationization of financial markets in which transaction orders are processed by computers and detail information is recorded automatically, while until the 1980s many people gathered at a market and prices are determined by shouting and screaming which could not be recorded. Now there are more than 100 financial market providers in the world and the number of transacted items exceeds one million. Namely, millions of prices in financial markets are changing with time scale in seconds, and you can access any market price at real time if you have a financial provider's terminal on your desk via the Internet.

Among these millions of items one of the most representative financial markets is the US Dollar-Japanese Yen (USD-JPY) market. In this market Dollar and Yen are exchanged among dealers of major international banks. Unlike the case of stock markets there is no physical trading place, but major international banks are linked by computer networks and orders are emitted from each dealer's terminal and transactions are done at an electronic broking system. Such a broking system and the computer networks are provided by financial provider companies like Reuters.

The foreign exchange markets are open 24 hours and deals are done whenever buy- and sell-orders meet. The minimum unit of a deal is one million USD (called a bar), and about three million bars are traded everyday in the whole foreign exchange markets in which more than 100 kinds of currencies are exchanged continuously. The total amount of money flow is about 100 times bigger than the total amount of daily world trade, so it is believed that most of deals are done not for the real world's needs, but they are based on speculative strategy or risk hedge, that is, to get profit by buying at a low price and selling at a high price, or to avoid financial loss by selling decreasing currency.

In Fig. 3 the price of one US Dollar paid by Japanese Yen in the foreign exchange markets is shown for 13 years [30]. The total number of data points is about 20 million, that is, about 10 thousand per day or the averaged transaction interval is seven seconds. A magnified part of the top figure for one year is shown in the second figure. The third figure is the enlargement of one month in the second figure. The bottom figure is again a part of the third figure, here the width is one day. It seems that at least the top three figures look quite similar. This is one of the fractal properties of market price (A) introduced in the previous chapter. This geometrical fractal property can be found in any market, so that this is a very universal market property.

However, it should be noted that this geometrical fractal property breaks down for very short time scale as typically shown in Fig. 4. In this figure the abscissa is 10 minutes range and we can observe each transaction separately. Obviously the price up down is more zigzag and more discrete than the large scale continuous market fluctuations shown in Fig. 3. In the case of USD-JPY market the time scale that this breakdown of scale invariance occurs typically at time scale of several hours.

The distribution of rate change in a unit time (one minute) is shown in Fig. 5. Here, there are two plots of cu-

Fractals and Economics, Figure 3
Dollar-Yen rate for 13 years (*Top*). *Dark areas* are enlarged in the following figure [30]

Fractals and Economics, Figure 4
Market price changes in 10 minutes
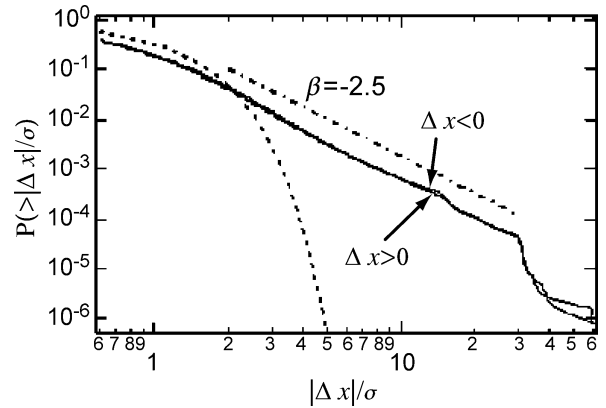
Fractals and Economics, Figure 5
Log–log plot of cumulative distribution of rate change [30]

original finding of Mandelbrot, (B) in the previous chapter, the reported exponent value is about 1.7 for cotton prices. In the case of stock markets power laws are confirmed universally for all items, however, the power exponents are not universal, taking value from near one to near five, typically around three [15]. Also the exponent values change in time year by year.

In order to demonstrate the importance of large fluctuations, Fig. 6 shows a comparison of three market prices. The top figure is the original rate changes for a week. The middle figure is produced from the same data, but it is consisted of rate changes of which absolute values are larger than 2σ, that is, about 5 % of all the data. In the bottom curve such large rate changes are omitted and the residue of 95 % of small changes makes the fluctuations. As known from these figures the middle figure is much closer to the original market price changes. Namely, the contribution from the power law tails of price change distribution is very large for macro-scale market prices.

Power law distribution of market price changes is also a quite general property which can be confirmed for any market. Up-down symmetry also holds universally in short time scale in general, however, for larger unit time the distribution of price changes gradually deforms and for very large unit time the distribution becomes closer to a Gaussian distribution. It should be noted that in special cases of market crashes or bubbles or hyper-inflations the up-down symmetry breaks down and the power law distribution is also likely to be deformed.

The autocorrelation of the time sequence of price changes generally decays quickly to zero, sometimes accompanied by a negative correlation in a very short time. This result implies that the market price changes are apparently approximated by white noise, and market prices

mulative distributions, $P(> \Delta x)$ for positive rate changes and $P(> |\Delta x|)$ for negative rate changes, which are almost identical meaning that the up-down symmetry of rate changes is nearly perfect. In this log–log plot the estimated power law distribution's exponent is 2.5. In the
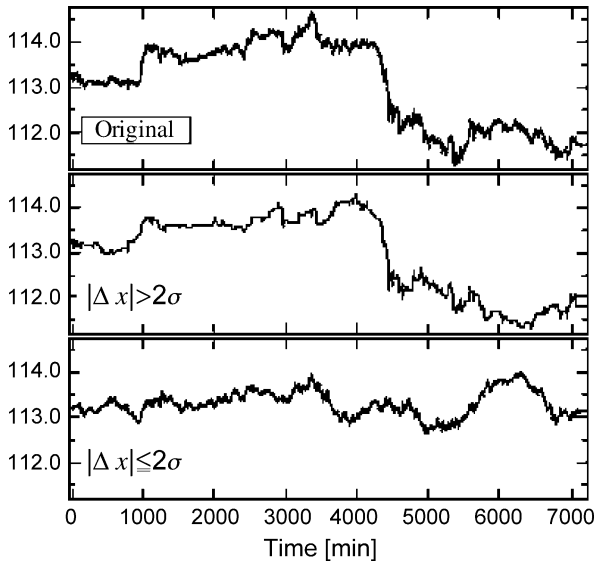
Fractals and Economics, Figure 6
USD-JPY exchange rate for a week (*top*) Rate changes smaller than 2σ are neglected (*middle*) Rate changes larger than 2σ are neglected (*bottom*)
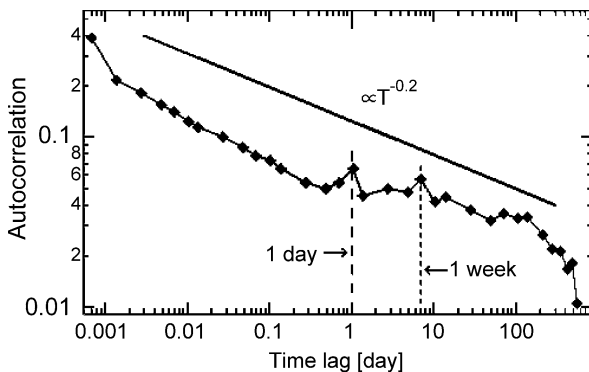
Fractals and Economics, Figure 7
Autocorrelation of volatility [30]

Fractals and Economics, Figure 8
Clustering of transaction intervals
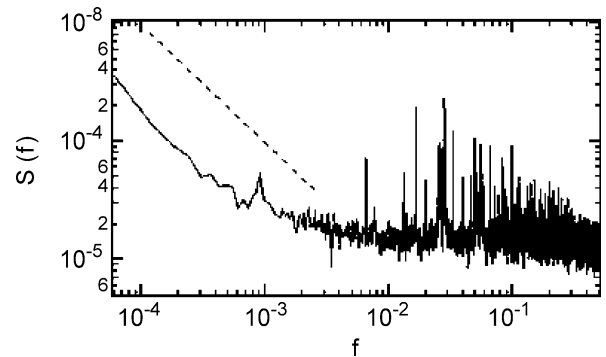
Fractals and Economics, Figure 9
Power spectrum of transaction intervals [50]

are known to follow nearly a random walk as a result. However, market price is not a simple random walk. In Fig. 7 the autocorrelation of volatility, which is defined by the square of price change, is shown in log–log scale. In the case of a simple random walk this autocorrelation should also decay quickly. The actual volatility autocorrelation nearly satisfies a power law implying that the volatility time series has a fractal clustering property. (See also Fig. 31 representing an example of price change clustering.)

Another fractal nature of markets can be found in the intervals of transactions. As shown in Fig. 8 the transac-

tion intervals fluctuate a lot in very short time scale. It is known that the intervals make clusters, namely, shorter intervals tend to gather. To characterize such clustering effect we can make a time sequence consisted of 0 and 1, where 0 denotes no deal was done at that time, and 1 denotes a deal was done. The corresponding power spectrum follows a $1/f$ power spectrum as shown in Fig. 9 [50].

Fractal properties are found not only in financial markets. Company's income distribution is known to follow also a power law [35]. A company's income is roughly given by subtraction of incoming money flow minus outgoing money flow, which can take both positive and negative values. There are about six million companies in Japan and Fig. 10 shows the cumulative distribution of annual income of these companies. Clearly we have a power law distribution of income I with the exponent very close to -1 in the middle size range, so-called the Zipf's law,

$$P(> I) \propto I^{-\beta}, \quad \beta = 1. \tag{10}$$

Although in each year every company's income fluctuates, and some percentage of companies disappear or are newly born, this power law is known to hold for more than 30
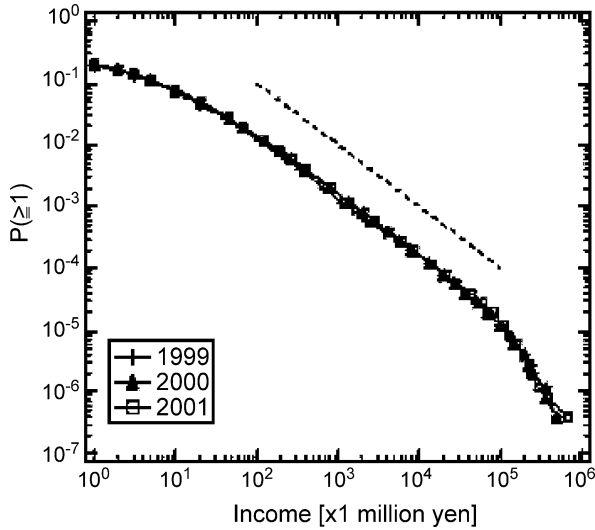
Fractals and Economics, Figure 10
Income distribution of companies in Japan
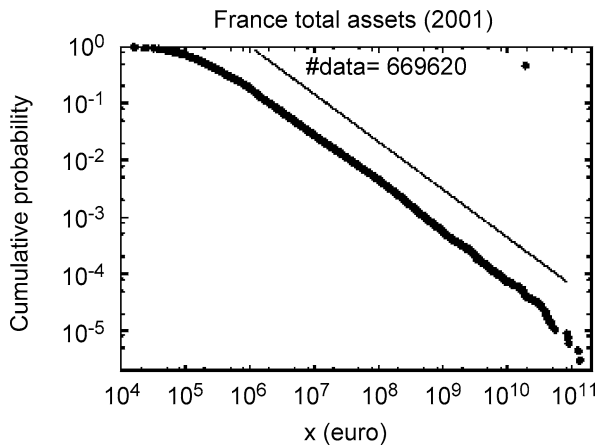
Fractals and Economics, Figure 11
Income distribution of companies in France [13]

Fractals and Economics, Figure 12
Income distribution of companies in each category [29]
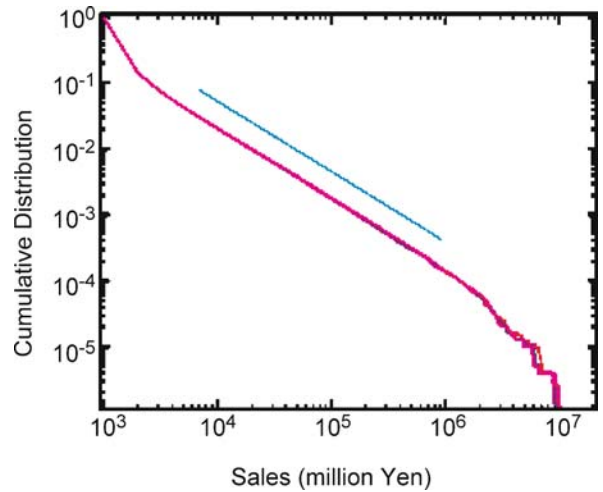
Fractals and Economics, Figure 13
The distribution of whole sales [34]

years. Similar power laws are confirmed in various countries, the case of France is plotted in Fig. 11 [13].

Observing more details by categorizing the companies, it is found that the income distribution in each job category follows nearly a power law with the exponent depending on the job category as shown in Fig. 12 [29]. The implication of this phenomenon will be discussed in Sect. "Income Distribution Models".

A company's size can also be viewed by the amount of whole sale or the number of employee. In Figs. 13 and 14 distributions of these quantities are plotted [34]. In both cases clear power laws are confirmed. The size distribution

of debts of bankrupted companies is also known to follow a power law as shown Fig. 15 [12].

A power law distribution can also be found in personal income. Figure 16 shows the personal income distribution in Japan in a log–log plot [1]. The distribution is clearly separated into two parts. The majority of people's incomes are well approximated by a log-normal distribution (the left top part of the graph), and the top few percent of people's income distribution is nicely characterized by a power law (the linear line in the left part of the graph). The majority of people are getting salaries from companies. This type of composite of two distributions is well-known from the pioneering study by Pareto about 100 years ago and it holds in various countries [8,22].
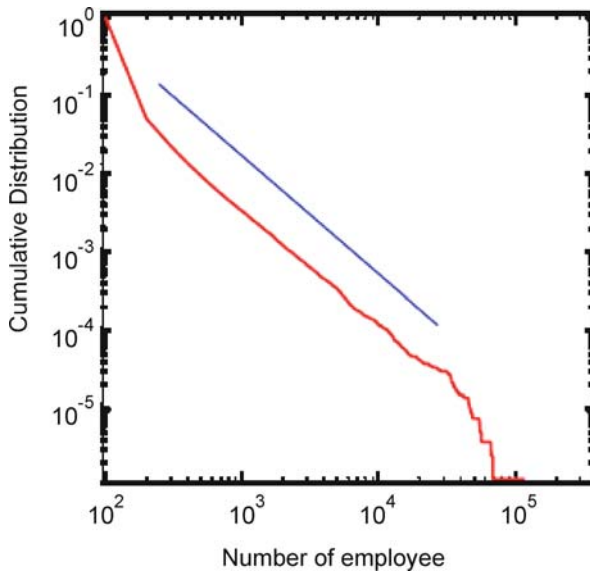
Fractals and Economics, Figure 14
The distribution of employee numbers [34]

Fractals and Economics, Figure 16
Personal income distribution in Japan [1]
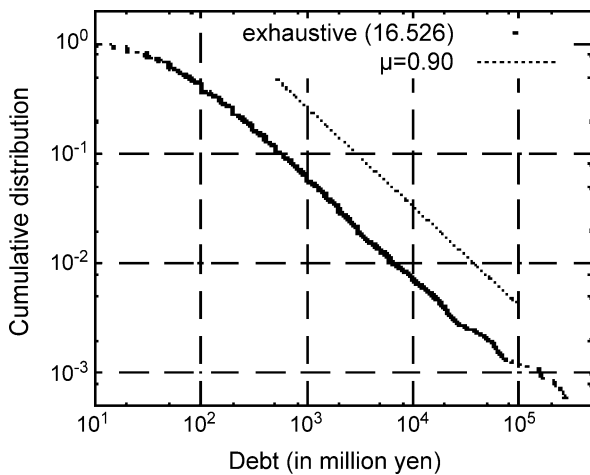
Fractals and Economics, Figure 15
The size distribution of debts of bankrupted companies [12]

A typical value of the power exponent is about two, significantly larger than the income distribution of companies. However, the exponent of the power law seems to be not universal and the value changes county by country or year by year. There is a tendency that the exponent is smaller, meaning more rich people, when the economy is improving [40].

Another fractal in economics can be found in a network of economic agents such as banks' money transfer network. As a daily activity banks transfer money to other

banks for various reasons. In Japan all of these interbank money transfers are done via a special computer network provided by the Bank of Japan. Detailed data of actual money transfer among banks are recorded and analyzed for the basic study.

The total amount of money flow among banks in a day is about 30×10^{12} yen with the number of transactions about 10 000. Figure 17 shows the distribution of the amount of money at a transaction. The range is not wide enough but we can find a power law with an exponent about 1.3 [20].

The number of banks is about 600, so the daily transaction number is only a few percent of the theoretically
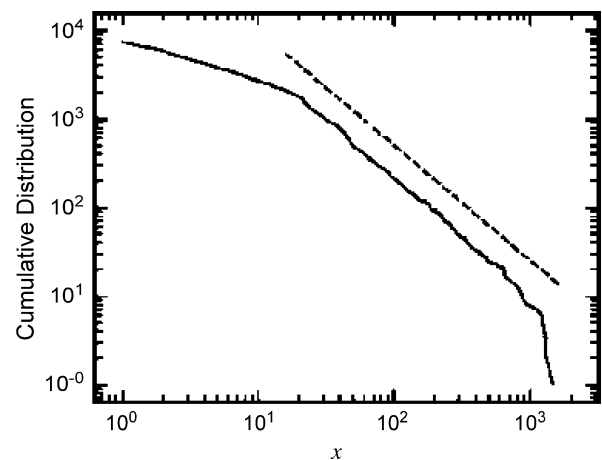
Fractals and Economics, Figure 17
The distribution of the amount of transferred money [21]
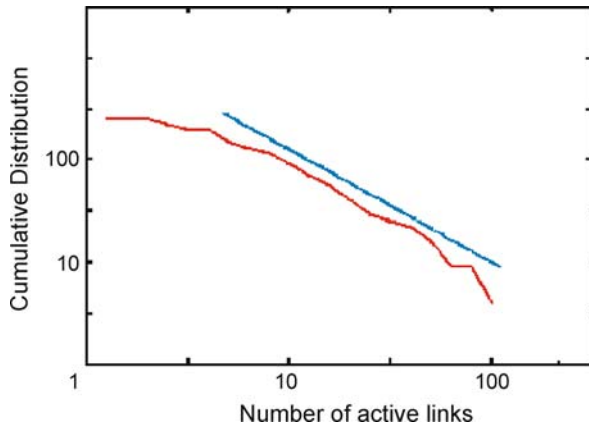
Fractals and Economics, Figure 18
The number distribution of active links per site [20]

Fractals and Economics, Figure 20
Distribution of total deposit for Japanese banks [57] Power law breaks down from 1999

possible combinations. It is confirmed that there are many pairs of banks which never transact directly. We can define active links between banks for pairs with the averaged number of transaction larger than one per day. By this criterion the number of links becomes about 2000, that is, about 0.5 percent of all possible link numbers. Compared with the complete network, the actual network topologies are much more sparse.

In Fig. 18 the number distribution of active links per site are plotted in log–log plot [21]. As is known from this graph, there is an intermediate range in which the link number distribution follows a power law. In the terminology of recent complex network study, this property is called the scale-free network [5]. The scale-free network structure among these intermediate banks is shown in Fig. 19.
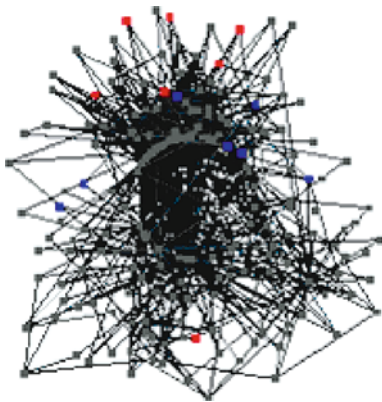
Fractals and Economics, Figure 19
Scale-free network of intermediate banks [20]

There are about 10 banks with large link numbers which deviate from the power law, also small link number banks with link number less than four are out of the power law. Such small banks are known to make a satellite structure that many banks linked to one large link number banks. It is yet to clarify why intermediate banks make fractal network, and also to clarify the role of large banks and small banks which are out of the fractal configuration.

In relation with the banks, there are fractal properties other than cash flow and the transaction network. The distribution of the whole amount of deposit of Japanese bank is approximated by a power law as shown in Fig. 20 [57]. In recent years large banks merged making a few mega banks and the distribution is a little deformed. Historically there were more than 6000 banks in Japan, however, now we have about 600 as mentioned. It is very rare that a bank disappears, instead banks are merged or absorbed. The number distribution of banks which are historically behind a present bank is plotted in Fig. 21, again a power law can be confirmed.

Other than the example of the bank network, network structures are very important generally in economics. In production process from materials, through various parts to final products the network structure is recently studied in view of complex network analysis [18]. Trade networks among companies can also be described by network terminology. Recently, network characterization quantities such as link numbers (Fig. 22), degrees of authority, and Pageranks are found to follow power laws from real trade data for nearly a million of companies in Japan [34].

Still more power laws in economics can be found in sales data. A recent study on the distribution of expenditure at convenience stores in one shopping trip shows a clear power law distribution with the exponent close to two as shown in Fig. 23 [33]. Also, book sales, movie hits,
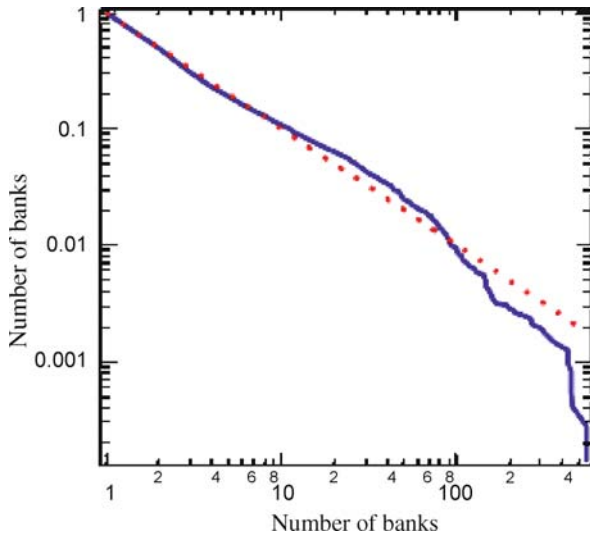
Fractals and Economics, Figure 21
Distribution of bank numbers historically behind a present bank [57]
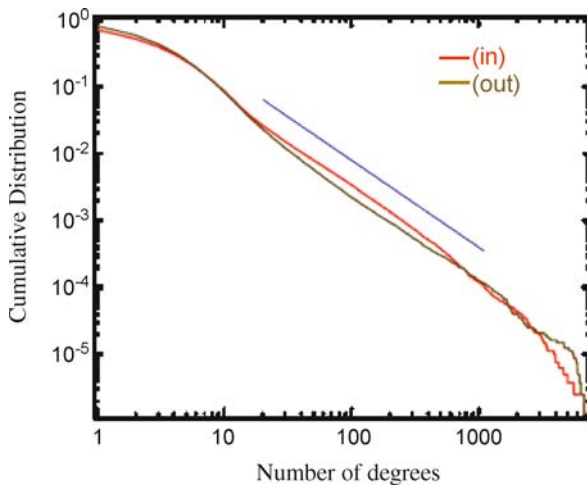
Fractals and Economics, Figure 22
Distribution of in-degrees and out-degrees in Japanese company network [34]

news paper sales are known to be approximated by power laws [39].

Viewing all these data in economics, we may say that fractals are everywhere in economics. In order to understand why fractals appear so frequently, we firstly need to make simple toy models of fractals which can be analyzed completely, and then, based on such basic models we can make more realistic models which can be directly comparable with real data. At that level of study we will be able to predict or control the complex real world economy.
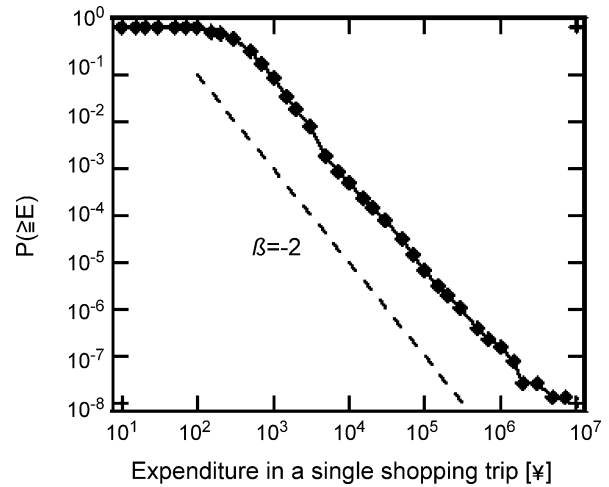
Fractals and Economics, Figure 23
Distribution of expenditure in one shopping trip [33]

Basic Models of Power Laws

In this chapter we introduce general mathematical and physical models which produce power law distributions. By solving these simple and basic cases we can deepen our understanding of the underlying mechanism of fractals or power law distributions in economics.

Transformation of Basic Distributions

A power law distribution can be easily produced by variable transformation from basic distributions.

1. Let x be a stochastic variable following a uniform distribution in the range $(0, 1]$, then, $y \equiv x^{-1/\alpha}$ satisfies a power law, $P(> y) = y^{-\alpha}$ for $y \geq 1$. This is a useful transformation in case of numerical simulation using random variable following power laws.
2. Let x be a stochastic variable following an exponential distribution, $P(> x) = e^{-x}$, for positive x, then, $y \equiv e^{x/\alpha}$ satisfies a power law, $P(> y) \propto y^{-\alpha}$. As exponential distributions occur frequently in random process such as the Poisson process, or energy distribution in thermal equilibrium, this simple exponential variable transformation can make it a power law.

Superposition of Basic Distributions

A power law distribution can also be easily produced by superposition of basic distributions.

Let x be a Gaussian distribution with the probability density given by

$$p_R(x) = \frac{\sqrt{R}}{\sqrt{2\pi}} e^{-\frac{R}{2}x^2} ,\qquad (11)$$

and R be a χ^2 distribution with degrees of freedom α,

$$w(R) = \frac{\left(\frac{1}{2}\right)^{\alpha/2}}{\Gamma\left(\frac{\alpha}{2}\right)} R^{\frac{\alpha}{2}-1} e^{-\frac{R}{2}} . \tag{12}$$

Then, the superposition of Gaussian distribution, Eq. (11), with the weight given by Eq. (12) becomes the T-distribution having power law tails:

$$p(x) = \int_0^\infty W(R) p_R(x) \mathrm{d}R$$

$$= \frac{\Gamma\left(\frac{\alpha+1}{2}\right)}{\sqrt{\pi}\,\Gamma\left(\frac{\alpha}{2}\right)} \frac{1}{(1+x^2)^{\frac{\alpha+1}{2}}} \propto |x|^{-\alpha-1} , \tag{13}$$

which is $P(>|x|) \propto |x|^{-\alpha}$ in cumulative distribution. In the special case that R, the inverse of variance of the normal distribution, distributes exponentially, the value of α is 2. Similar super-position can be considered for any basic distributions and power law distributions can be produced by such superposition.

Stable Distributions

Assume that stochastic variables, $x_1, x_2, \ldots, x_n$, are independent and follow the same distribution, $p(x)$, then consider the following normalized summation;

$$X_n \equiv \frac{x_1 + x_2 + \cdots + x_n - \mu_n}{n^{1/\alpha}} . \tag{14}$$

If there exists $\alpha > 0$ and μ_n, such that the distribution of X_n is identical to $p(x)$, then, the distribution belongs to one of the Levy stable distributions [10]. The parameter α is called the characteristic exponent which takes a value in the range $(0, 2]$. The stable distribution is characterized by four continuous parameters, the characteristic exponent, an asymmetry parameter which takes a value in $[-1, 1]$, the scale factor which takes a positive value and the location parameter which takes any real number. Here, we introduce just a simple case of symmetric distribution around the origin with the unit scale factor. The probability density is then given as

$$p(x; \alpha) = \frac{1}{2\pi} \int_{-\infty}^{\infty} e^{-i\rho x} e^{-|\rho|^\alpha} \mathrm{d}\rho . \tag{15}$$

For large $|x|$ the cumulative distribution follows the power law, $P(>x; \alpha) \propto |x|^{-\alpha}$ except the case of $\alpha = 2$. The stable distribution with $\alpha = 2$ is the Gaussian distribution.

The most important property of the stable distribution is the generalized central limit theorem: If the distribution of sum of any independent identically distributed

random variables like X_n in Eq. (14) converges in the limit of $n \to \infty$ for some value of α, then the limit distribution is a stable distribution with the characteristic exponent α. For any distribution with finite variance, the ordinary central limit theory holds, that is, the special case of $\alpha = 2$. For any infinite variance distribution the limit distribution is $\alpha \neq 2$ with a power law tail. Namely, a power law realizes simply by summing up infinitely many stochastic variables with diverging variance.

Entropy Approaches

Let x_0 be a positive constant and consider a probability density $p(x)$ defined in the interval $[x_0, \infty)$, the entropy of this distribution is given by

$$S \equiv -\int_{x_0}^{\infty} p(x) \log p(x) \mathrm{d}x . \tag{16}$$

Here, we find a distribution that maximizes the entropy with a constraint such that the expectation of logarithm of x is a constant, $\langle \log x \rangle = M$. Then, applying the variational principle to the following function,

$$L \equiv -\int_{x_0}^{\infty} p(x) \log p(x) \mathrm{d}x - \lambda_1 \left(\int_{x_0}^{\infty} p(x) \mathrm{d}x - 1 \right)$$

$$+ \lambda_2 \left(\int_{x_0}^{\infty} p(x) \log x \mathrm{d}x - M \right) \tag{17}$$

the power law is obtained,

$$P(\geq x) = \left(\frac{x}{x_0} \right)^{-\frac{1}{M - \log x_0}} . \tag{18}$$

In other words, a power law distribution maximizes the entropy in the situation where products are conserved. To be more precise, consider two time dependent random variables interacting each other satisfying the relation, $x_1(t) \cdot x_2(t) = x_1(t') \cdot x_2(t')$, then the equilibrium distribution follows a power law.

Another entropy approach to the power laws is to generalize the entropy by the following form [56],

$$S_q \equiv \frac{1 - \int_{x_0}^{\infty} p(x)^q \mathrm{d}x}{q - 1} , \tag{19}$$

where q is a real number. This function is called the q-entropy and the ordinary entropy, Eq. (15), recovers in the

limit of $q \to 1$. Maximizing the q-entropy keeping the variance constant, so-called a q-Gaussian distribution is obtained, which has the same functional form with the T-distribution, Eq. (12), with the exponent α given by

$$\alpha = \frac{q-3}{1-q}. \tag{20}$$

This generalized entropy formulation is often applied to nonlinear systems having long correlations, in which power law distributions play the central role.

Random Multiplicative Process

Stochastic time evolution described by the following formulation is called the multiplicative process,

$$x(t+1) = b(t)x(t) + f(t), \tag{21}$$

where $b(t)$ and $f(t)$ are both independent random variables [17]. In the case that $b(t)$ is a constant, the distribution of $x(t)$ depends on the distribution of $f(t)$, for example, if $f(t)$ follows a Gaussian distribution, then the distribution of $x(t)$ is also a Gaussian. However, in the case that $b(t)$ fluctuates randomly, the resulting distribution of $x(t)$ is known to follows a power law independent of $f(t)$,

$$P(> x) \propto |x|^{-\alpha}, \tag{22}$$

where the exponent α is determined by solving the following equation [48],

$$\langle |b(t)|^{\alpha} \rangle = 1. \tag{23}$$

This steady distribution exists when $\langle \log |b(t)| \rangle < 0$ and $f(t)$ is not identically 0. As a special case that $b(t) = 0$ with a finite probability, then a steady state exists. It is proved rigorously that there exists only one steady state, and starting from any initial distribution the system converges to the power law steady state.

In the case $\langle \log |b(t)| \rangle \geq 0$ there is no statistically steady state, intuitively the value of $|b(t)|$ is so large that $x(t)$ is likely to diverge. Also in the case $f(t)$ is identically 0 there is no steady state as known from Eq. (21) that $\log |x(t)|$ follows a simple random walk with random noise term, $\log |b(t)|$.

The reason why this random multiplicative process produces a power law can be understood easily by considering a special case that $b(t) = b > 1$ with probability 0.5 and $b(t) = 0$ otherwise, with a constant value of $f(t) = 1$. In such a situation the value of $x(t)$ is $1 + b + b^2 + \cdots + b^K$ with probability $(0.5)^K$. From this

we can directly evaluate the distribution of $x(t)$,

$$P\left(\geq \frac{b^{K+1}-1}{b-1}\right) = 2^{-K+1} \quad \text{i. e.}$$

$$P(\geq x) = 4(1 + (b-1)x)^{-\alpha}, \quad \alpha = \frac{\log 2}{\log b}. \tag{24}$$

As is known from this discussion, the mechanism of this power law is deeply related to the above mentioned transformation of exponential distribution in Sect. "Transformation of Basic Distributions".

The power law distribution of a random multiplicative process can also be confirmed experimentally by an electrical circuit in which resistivity fluctuates randomly [38]. In an ordinary electrical circuit the voltage fluctuations in thermal equilibrium is nearly Gaussian, however, for a circuit with random resistivity a power law distribution holds.

Aggregation with Injection

Assume the situation that many particles are moving randomly and when two particles collide they coalesce making a particle with mass conserved. Without any injection of particles the system converges to the trivial state that only one particle remains. In the presence of continuous injection of small mass particles there exists a non-trivial statistically steady state in which mass distribution follows a power law [41]. Actually, the mass distribution of aerosol in the atmosphere is known to follow a power law in general [11].

The above system of aggregation with injection can be described by the following model. Let j be the discrete space, and $x_j(t)$ be the mass on site j at time t, then choose one site and let the particle move to another site and particles on the visited site merge, then add small mass particles to all sites, this process can be mathematically given as,

$$x_j(t+1) = \begin{cases} x_j(t) + x_k(t) + f_j(t), & \text{prob} = 1/N \\ x_j(t) + f_j(t), & \text{prob} = (N-2)/N \\ f_j(t), & \text{prob} = 1/N \end{cases} \tag{25}$$

where N is the total number of sites and $f_j(t)$ is the injected mass to the site j.

The characteristic function, $Z(\rho, t) \equiv \langle e^{-\rho x_j(t)} \rangle$, which is the Laplace transform of the probability density,

satisfies the following equation by assuming uniformity,

$$Z(\rho, t+1)$$
$$= \left\{ \frac{N-2}{N} Z(\rho, t)^2 + \frac{1}{N} Z(\rho, t) + \frac{1}{N} \right\} \langle e^{-\rho f_j(t)} \rangle . \tag{26}$$

The steady state solution in the vicinity of $\rho = 0$ is obtained as

$$Z(\rho) = 1 - \sqrt{\langle f \rangle} \rho^{1/2} + \cdots . \tag{27}$$

From this behavior the following power law steady distribution is obtained.

$$P(\geq x) \propto x^{-\alpha} , \quad \alpha = \frac{1}{2} . \tag{28}$$

By introducing a collision coefficient depending on the size of particles power laws with various values of exponents realized in the steady state of such aggregation with injection system [46].

Critical Point of a Branching Process

Consider the situation that a branch grows and splits with probability q or stops growing with probability $1 - q$ as shown in Fig. 24. What is the size distribution of the branch? This problem can be solved in the following way. Let $p(r)$ be the probability of finding a branch of size r, then the next relation holds.

$$p(r+1) = q \sum_{s=1}^{r-1} p(s)p(r-s) . \tag{29}$$

Multiplying y^{r+1} and summing up by r from 0 to ∞, a closed equation of the generating function, $M(y)$, is obtained,

$$M(y) - 1 + q = q \cdot y \cdot M(y)^2 , \quad M(y) \equiv \sum_{r=0}^{\infty} y^r p(r) . \tag{30}$$

Solving this quadratic equation and expanding in terms of y, we have the probability density,

$$p(r) \propto r^{-3/2} e^{-Q(q)r} , \quad Q(q) \equiv \log 4q(1-q) . \tag{31}$$

For $q < 0.5$ the probability decays exponentially for large r, in this case all branches has a finite size. At $q = 0.5$ the branch size follows the power law, $P(\geq r) \propto r^{-1/2}$, and the average size of branch becomes infinity. For
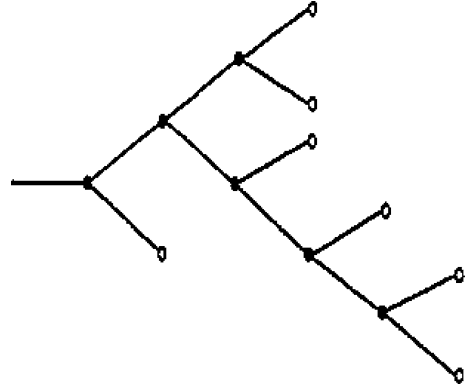
Fractals and Economics, Figure 24
Branching process (from *left* to *right*)

$q > 0.5$ there is a finite probability that a branch grows infinitely. The probability of having an infinite branch, $p(\infty) = 1 - M(1)$, is given as,

$$p(\infty) = \frac{2q - 1 + \sqrt{1 - 4q(1-q)}}{2q} , \tag{32}$$

which grows monotonically from zero to one in the range $q = [0.5, 1]$. It should be noted that the power law distribution realizes at the critical point between the finite-size phase and the infinite-size phase [42].

Compared with the preceding model of aggregation with injection, Eq. (28), the mass distribution is the same as the branch size distribution at the critical point in Eq. (31). This coincidence is not an accident, but it is known that aggregation with injection automatically chooses the critical point parameter. Aggregation and branching are reversed process and the steady occurrence of aggregation implies that branching numbers keep a constant value on average and this requires the critical point condition. This type of critical behaviors is called the self-organized criticality and examples are found in various fields [4].

Finite Portion Transport

Here, a kind of mixture of aggregation and branching is considered. Assume that conserved quantities are distributed in N-sites. At each time step choose one site randomly, and transport a finite portion, $\theta x_j(t)$, to another randomly chosen site, where θ is a parameter in the range $[0, 1]$.

$$x_j(t+1) = (1-\theta)x_j(t) ,$$
$$x_k(t+1) = x_k(t) + \theta x_j(t) . \tag{33}$$

It is known that for small positive θ the statistically steady distribution x is well approximated by a Gaussian like the case of thermal fluctuations. For θ close to 1 the fluctuation of x is very large and its distribution is close to a power law. In the limit θ goes to 1 and the distribution converges to Eq. (28), the aggregation with injection case. For intermediate values of θ the distribution accompanies a fat tail between Gaussian and a power law [49].

Fractal Tiling

A fractal tiling is introduced as the final basic model. Figure 25 shows an example of fractal tiling of a plane by squares. Like this case Euclidean space is covered by various sizes of simple shapes like squares, triangles, circles etc. The area size distribution of squares in Fig. 25 follows the power law,

$$P(\geq x) \propto x^{-\alpha}, \quad \alpha = 1/2. \tag{34}$$

Generalizing this model in d-dimensional space, the distribution of d-dimensional volume x is characterized by a power law distribution with an exponent, $\alpha = (d-1)/d$, therefore, the Zipf's law which is the case of $\alpha = 1$ realizes in the limit of $d = \infty$. The fracture size distribution measured in mass introduced in the beginning of this article corresponds to the case of $d = 3$.

A classical example of fractal tiling is the Apollonian gasket, that is, a plane is covered totally by infinite number of circles which are tangent each other. For a given river pattern like Fig. 26 the basin area distribution follows a power law with exponent about $\alpha = 0.4$ [45]. Although these are very simple geometric models, simple models may sometimes help our intuitive understanding of fractal phenomena in economics.

Fractals and Economics, Figure 26
Fractal tiling by river patterns [45]

Market Models

In this chapter market price models are reviewed in view of fractals. There are two approaches for construction of market models. One is modeling the time sequences directly by some stochastic model, and the other is modeling markets by agent models which are artificial markets in computer consisted of programmed dealers.

The first market price model was proposed by Bachelier in 1900 written as his Ph.D thesis [3], that is, five years before the model of Einstein's random walk model of colloid particles. His idea was forgotten for nearly 50 years. In 1950's Markowitz developed the portfolio theory based on a random walk model of market prices [28]. The theory of option prices by Black and Scholes was introduced in the 1970s, which is also based on random walk model of market prices, or to be more precise a logarithm of market prices in continuum description [7].

In 1982 Engle introduced a modification of the simple random walk model, the ARCH model, which is the abbreviation of auto-regressive conditional heteroscedasticity [9]. This model is formulated for market price difference as,

$$\Delta x(t) = \sigma(t) f(t), \tag{35}$$

where $f(t)$ is a random variable following a Gaussian distribution with 0 mean and variance unity, the local variance $\sigma(t)$ is given as

$$\sigma(t)^2 = c_0 + \sum_{j=1}^{k} c_k (\Delta x(t-k))^2, \tag{36}$$

Fractals and Economics, Figure 25
An example of fractal tiling

with adjustable positive parameters, $\{c_0, c_1, \ldots, c_k\}$. By the effect of this modulation on variance, the distribution of price difference becomes superposition of Gaussian distribution with various values of variance, and the distribution becomes closer to a power law. Also, volatility clustering occurs automatically so that the volatility autocorrelation becomes longer.

There are many variants of ARCH models, such as GARCH and IGARCH, but all of them are based on purely probabilistic modeling, and the probability of prices going up and that of going down are identical.

Another type of market price model has been proposed from physics view point [53]. The model is called the PUCK model, an abbreviation of potentials of unbalanced complex kinetics, which assumes the existence of market's time-dependent potential force, $U_M(x; t)$, and the time evolution of market price is given by the following set of equations;

$$x(t+1) - x(t) = -\frac{d}{dx} U_M(x; t)\bigg|_{x=x(t)-x_M(t)} + f(t), \quad (37)$$

$$U_M(x; t) \equiv \frac{b(t)}{M-1} \frac{x^2}{2}, \quad (38)$$

where M is the number of moving average needed to define the center of potential force,

$$x_M(t) \equiv \frac{1}{M} \sum_{k=0}^{M-1} x(t - k). \quad (39)$$

In this model $f(t)$ is the external noise and $b(t)$ is the curvature of quadratic potential which changes with time. When $b(t) = 0$ the model is identical to the simple random walk model. When $b(t) > 0$ the market prices are attracted to the moving averaged price, $x_M(t)$, the market is stable, and when $b(t) < 0$ prices are repelled from $x_M(t)$ so that the price fluctuation is large and the market is unstable. For $b(t) < -2$ the price motion becomes an exponential function of time, which can describe singular behavior such as bubbles and crashes very nicely.

In the simplest case of $M = 2$ the time evolution equation becomes,

$$\Delta x(t + 1) = -\frac{b(t)}{2} \Delta x(t) + f(t). \quad (40)$$

As is known from this functional form in the case $b(t)$ fluctuates randomly, the distribution of price difference follows a power law as mentioned in the previous Sect. "Random Multiplicative Process", Random multiplicative process. Especially the PUCK model derives the ARCH model by introducing a random nonlinear potential function [54]. The value of $b(t)$ can be estimated from the data and most of known empirical statistical laws including fractal properties are fulfilled as a result [55].

The peculiar difference of this model compared with financial technology models is that directional prediction is possible in some sense. Actually, from the data it is known that $b(t)$ changes slowly in time, and for non-zero $b(t)$ the autocorrelation is not zero implying that the up-down statistics in the near future is not symmetric. Moreover in the case of $b(t) < -2$ the price motion show an exponential dynamical growth hence predictable.

As introduced in Sect. "Examples in Economics" the tick interval fluctuations can be characterized by the $1/f$ power spectrum. This power law can be explained by a model called the self-modulation model [52]. Let Δt_j be the jth tick interval, and we assume that the tick interval can be approximated by the following random process,

$$\Delta t_{j+1} = \mu_j \frac{1}{K} \sum_{k=0}^{K-1} \Delta t_{j-k} + g_j, \quad (41)$$

where μ_j is a positive random number following an exponential distribution with the mean value 1, and K is an integer which means the number of moving average, g_j is a positive random variable. Due to the moving average term in Eq. (41) the tick interval automatically make clusters as shown in Fig. 27, and the corresponding power spectrum is proved to be proportional to $1/f$ as typically represented in Fig. 28.

The market data of tick intervals are tested whether Eq. (41) really works or not. In Fig. 29 the cumulative probability of estimated value of μ_j from market data is plotted where the moving average size is determined by the physical time of 150 seconds and 400 seconds. As known from this figure, the distribution fits very nicely with the exponential distribution when the moving average size is 150 seconds. This result implies that dealers in the market are mostly paying attention to the latest transaction for about a few minutes only. And the dealers' clocks in their
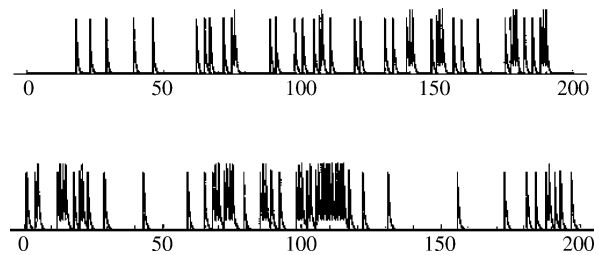
Fractals and Economics, Figure 27
Tick intervals of Poisson process (*top*) and the self-modulation process (*bottom*) [52]
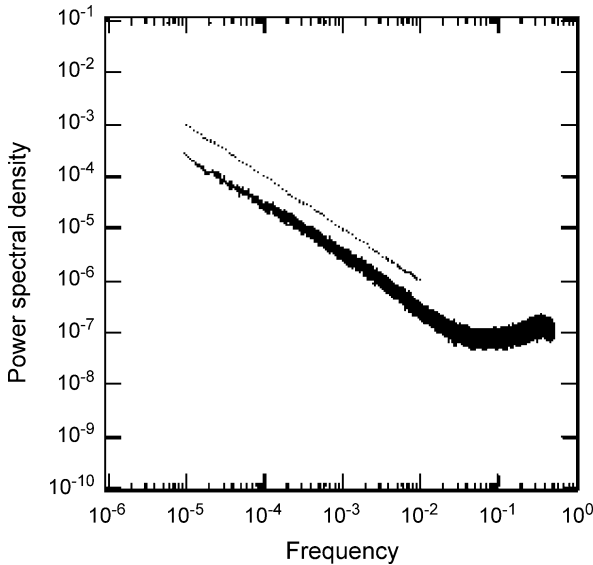
Fractals and Economics, Figure 28
The power spectrum of the self-modulation process [52]
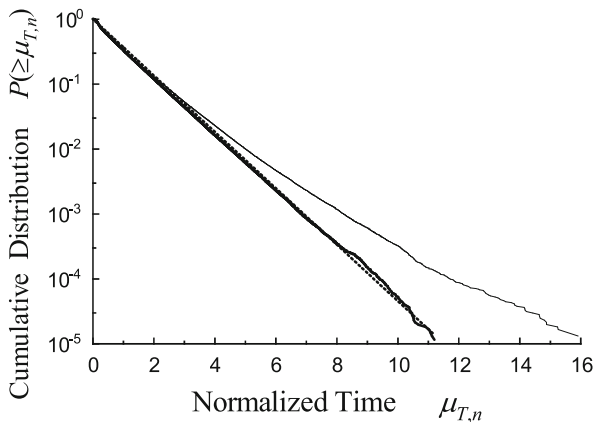
Fractals and Economics, Figure 29
The distribution of normalized time interval [50]

minds move quicker if the market becomes busier. By this self-modulation effect transactions in markets automatically make a fractal configuration.

Next, we introduce a dealer model approach to the market [47]. In any financial market dealers' final goal is to gain profit from the market. To this end dealers try to buy at the lowest price and to sell at the highest price. Assume that there are N dealers at a market, and let the jth dealer's buying and selling prices in their mind $B_j(t)$ and $S_j(t)$. For each dealer the inequality, $B_j(t) < S_j(t)$, always holds. We pay attention to the maximum price of $\{B_j(t)\}$ called the best bid, and to the minimum price of $\{S_j(t)\}$ called the

best ask. Transactions occur in the market if there exists a pair of dealers, j and k, who give the best bid and best ask respectively, and they fulfill the following condition,

$$B_j(t) \geq S_k(t) . \tag{42}$$

In the model the market price is given by the mean value of these two prices.

As a simple situation we consider a deterministic time evolution rule for these dealers. For all dealers the spread, $S_j(t) - B_j(t)$, is set to be a constant L. Each dealer has a position, either a seller or a buyer. When the jth dealer's position is a seller the selling price in mind, $S_j(t)$, decreases every time step until he can actually sell. Similar dynamics is applied to a buyer with the opposite direction of motion. In addition we assume that all dealers shift their prices in mind proportional to a market price change. When this proportional coefficient is positive, the dealer is categorized as a trend-follower. If this coefficient is negative, the dealer is called a contrarian. These rules are summarized by the following time evolution equations.

$$B_j(t+1) = B_j(t) + a_j S_j + b_j \Delta x(t) , \tag{43}$$

where S_j takes either $+1$ or -1 meaning the buyer position or seller position, respectively, $\Delta x(t)$ gives the latest market price change, $\{a_j\}$ are positive numbers given initially, $\{b_j\}$ are also parameters given initially.

Figure 30 shows an example of market price evolution in the case of three dealers. It should be noted that although the system is deterministic, namely, the future price is determined uniquely by the set of initial values, resulting market price fluctuates almost randomly even in the minimum case of three dealers. The case of $N = 2$
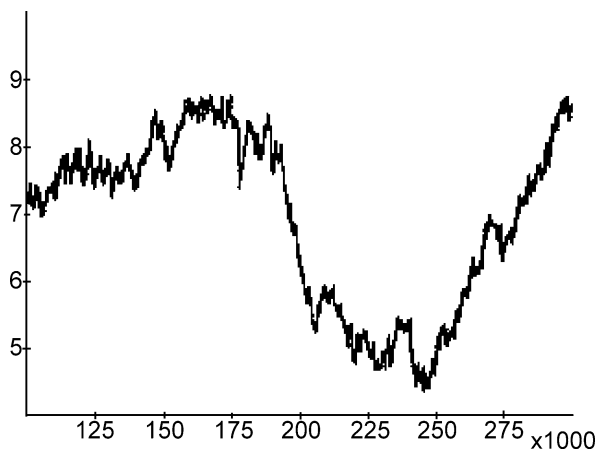
Fractals and Economics, Figure 30
Price evolution of a market with deterministic three dealers
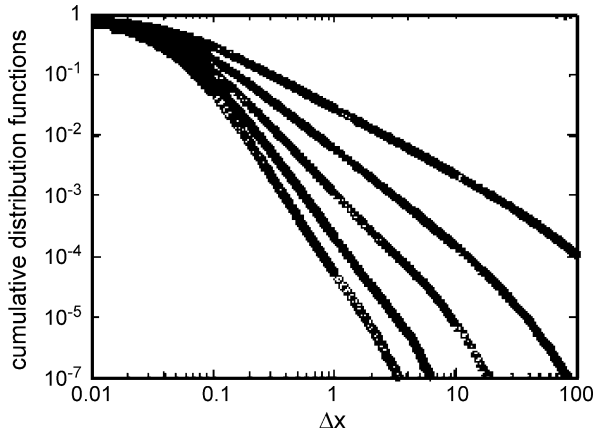
Fractals and Economics, Figure 31
Cumulative distribution of a dealer model for different values of _b_. For weaker trend-follow the slope is steeper [38]
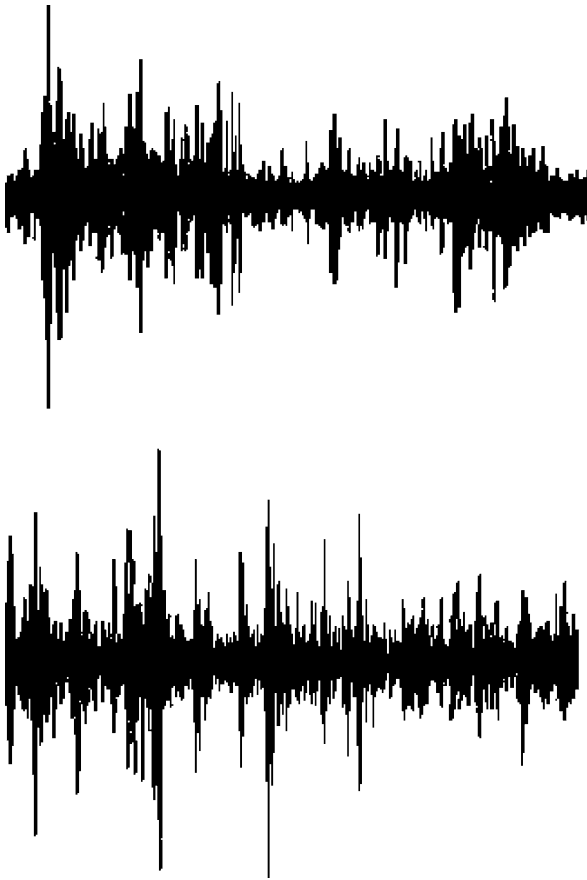
Fractals and Economics, Figure 32
Price difference time series for a real market (_top_) and a dealer model (_bottom_)

gives only periodic time evolution as expected, while for $N \geq 3$ the system can produce market price fluctuations similar to the real market price fluctuations, for example, the fractal properties of price chart and the power law distribution of price difference are realized.

In the case that the value of $\{b_j\}$ are identical for all dealers, b, then the distribution of market price difference follows a power law where the exponent is controllable by this trend-follow parameter, b as shown in Fig. 31 [37]. The volatility clustering is also observed automatically for large dealer number case as shown in Fig. 32 (bottom) which looks quite similar to a real price difference time series Fig. 32 (top).

By adding a few features to this basic dealer model it is now possible to reproduce almost all statistical characteristics of market, such as tick-interval fluctuations, abnormal diffusions etc. [58]. In this sense the study of market behaviors are now available by computer simulations based on the dealer model. Experiments on the market is either impossible or very difficult for a real market, however, in an artificial market we can repeat occurrence of bubbles and crashes any times, so that we might be able to find a way to avoid catastrophic market behaviors by numerical simulation.

Income Distribution Models

Let us start with a famous historical problem, the St. Petersburg Paradox, as a model of income. This paradox was named after Daniel Bernoulli's paper written when he was staying in the Russian city, Saint Petersburg, in 1738 [6]. This paradox treats a simple lottery as described in the following, which is deeply related to the infinite expected value problem in probability theory and also it has been attracting a lot of economists' interest in relation with the essential concept in economics, the utility [2].

Assume that you enjoy a game of chance, you pay a fixed fee, X dollars, to enter, and then you toss a fair coin repeatedly until a tail firstly appears. You win 2^n dollars where n is the number of heads. What is the fair price of the entrance fee, X?

Mathematically a fair price should be equal to the expectation value, therefore, it should be given as,

$$X = \sum_{n=0}^{\infty} 2^n \cdot \frac{1}{2^{n+1}} = \infty \,. \tag{44}$$

This mathematical answer implies that even X is one million dollars this lottery is generous enough and you should buy because expectation is infinity. But, would you dare to buy this lottery, in which you will win only one dollar with probability 0.5, and two dollars with probability 0.25, ...?

Bernoulli's answer to this paradox is to introduce the human feeling of value, or utility, which is proportional to the logarithm of price, for example. Based on this expected utility hypothesis the fair value of X is given as follows,

$$X = \sum_{n=0}^{\infty} \frac{U(2^n)}{2^{n+1}} = \sum_{n=0}^{\infty} \frac{\log(2^n)}{2^{n+1}} = 1 + \log 2 \approx 1.69 \,, \quad (45)$$

where the utility function, $U(x) = 1 + \log x$, is normalized to satisfy $U(1) = 1$. This result implies that the appropriate entry fee X should be about two dollars.

The idea of utility was highly developed in economics for description of human behavior, in the way that human preference is determined by maximal point of utility function, the physics concept of the variational principle applied to human action. Recently, in the field of behavioral finance which emerged from psychology the actual observation of human behaviors about money is the main task and the St. Petersburg paradox is attracting attention [36].

Although Bernoulli's solution may explain the human behavior, the fee $X = 2$ is obviously so small that the bookmaker of this lottery will bankrupt immediately if the entrance fee is actually fixed as two dollars and if a lot of people actually buy it. The paradox is still a paradox.

To clarify what is the problem we calculate the distribution of income of a gambler. As an income is 2^n with probability 2^{-n-1}, the cumulative distribution of income is readily obtained as,

$$P(\geq x) \propto 1/x \,. \quad (46)$$

This is the power law which we observed for income distribution of companies in Sect. "Examples in Economics".

The key of this lottery is the mechanism that the prize money doubles at each time a head appears and the coin toss stops when a tail appears. By denoting the number of coin toss by t, we can introduce a stochastic process or a new lottery which is very much related to the St. Petersburg lottery.

$$x(t + 1) = b(t)x(t) + 1 \,, \quad (47)$$

where $b(t)$ is 2 with probability 0.5 and is 0 otherwise. As introduced in Sect. "Random Multiplicative Process", this problem is solved easily and it is confirmed that the steady state cumulative distribution of $x(t)$ also follows Eq. (46). The difference between the St. Petersburg lottery and the new lottery Eq. (47) is the way of payment of entrance fee. In the case of St. Petersburg lottery the entrance fee X is paid in advance, while in the case of new lottery you have to add one dollar each time you toss a coin. This new

lottery is fair from both the gambler side and the bookmaker side because the expectation of income is given by $\langle x(t) \rangle = t$ and the amount of paid fee is also t.

Now we introduce a company's income model by generalizing this new fair lottery in the following way,

$$I(t + 1) = b(t)I(t) + f(t) \,, \quad (48)$$

where $I(t)$ denotes the annual income of a company, $b(t)$ represents the growth rate which is given randomly from a growth rate distribution $g(b)$, and $f(t)$ is a random noise. Readily from the results of Sect. "Random Multiplicative Process", we have a condition to satisfy the empirical relation, Eq. (10),

$$\langle b(t) \rangle = \int b g(b) = 1 \,. \quad (49)$$

This relation is confirmed to hold approximately in actual company data [32].

In order to explain the job category dependence of the company's income distribution already shown in Fig. 12, we plot the comparison of exponents in Fig. 33. Empirically estimated exponents are plotted in the ordinate and the solutions of the following equation calculated in each job category are plotted in the abscissa,

$$\langle b(t)^\beta \rangle = 1 \,. \quad (50)$$

The data points are roughly on a straight line demonstrating that the simple growth model of Eq. (48) seems to be meaningful.

An implication of this result is that if a job category is expanding, namely, $\langle b(t) \rangle > 1$, then the power law exponent determined by Eq. (50) is smaller than 1. On the other hand if a job category is shrinking, we have an exponent that is larger than 1.

This type of company's income model can be generalized to take into account the effect of company's size
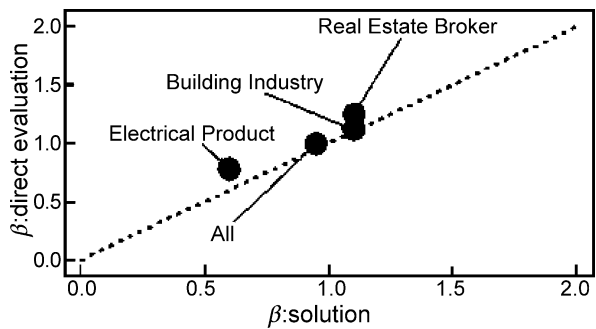
Fractals and Economics, Figure 33
Theoretical predicted exponent value vs. observed value [29]
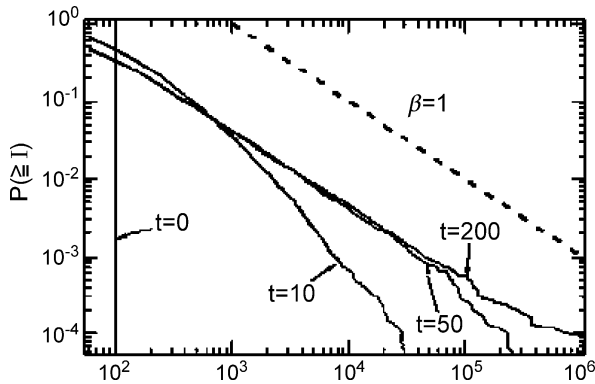
Fractals and Economics, Figure 34
Numerical simulation of income distribution evolution of Japanese companies [32]
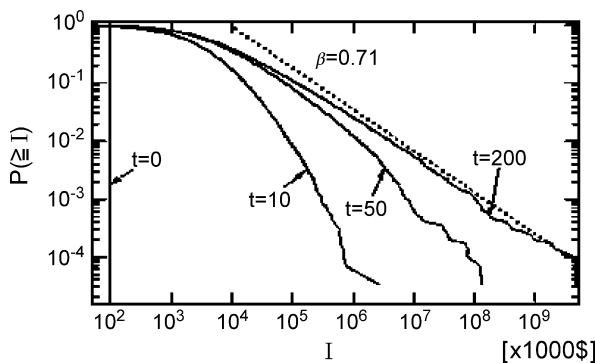
Fractals and Economics, Figure 35
Numerical simulation of income distribution evolution of USA companies [32]

dependence on the distribution of growth rate. Also, the magnitude of the random force term can be estimated from the probability of occurrence of negative income. Then, assuming that the present growth rate distribution continues we can perform a numerical simulation of company's income distribution starting from a uniform distribution as shown in Fig. 34 for Japan and in Fig. 35 for USA. It is shown that in the case of Japan, the company size distribution converges to the power law with exponent -1 in 20 years, while in the case of USA the steady power law's slope is about -0.7 and it takes about 100 years to converge [31]. According to this result extremely large companies with size about 10 times bigger than the present biggest company will appear in USA in this century. Of course the growth rate distribution will change faster than this prediction, however, this model can tell the qualitative direction and the speed of change in very macroscopic economical conditions.

Other than this simple random multiplicative model approach there are various approaches to explain empirical facts about company's statistics assuming a hierarchical structure of organization, for example [23].

Future Directions

Fractal properties generally appear in almost any huge data in economics. As for financial market models, empirical fractal laws are reproduced and the frontier of study is now at the level of practical applications. However, there are more than a million markets in the world and little is known about their interaction. More research on market interaction will be promising. Company data so far analyzed show various fractal properties as introduced in Sect. "Examples in Economics", however, they are just a few cross-sections of global economics. Especially, companies' interaction data are inevitable to analyze the underlying network structures. Not only money flow data it will be important to observe material flow data in manufacturing and consumption processes. From the viewpoint of environmental study, such material flow network will be of special importance in the near future. Detail sales data analysis is a new topic and progress is expected.

Bibliography

Primary Literature

1. Aoyama H, Nagahara Y, Okazaki MP, Souma W, Takayasu H, Takayasu M (2000) Pareto's law for income of individuals and debt of bankrupt companies. Fractals 8:293–300
2. Aumann RJ (1977) The St. Petersburg paradox: A discussion of some recent comments. J Econ Theory 14:443–445 http://en.wikipedia.org/wiki/Robert_Aumann
3. Bachelier L (1900) Theory of Speculation. In: Cootner PH (ed) The Random Character of Stock Market Prices. MIT Press, Cambridge (translated in English)
4. Bak P (1996) How Nature Works. In: The Science of Self-Organized Criticality. Springer, New York
5. Barabási AL, Réka A (1999) Emergence of scaling in random networks. Science 286:509–512; http://arxiv.org/abs/cond-mat/9910332
6. Bernoulli D (1738) Exposition of a New Theory on the Measurement of Risk; Translation in: (1954) Econometrica 22:22–36
7. Black F, Scholes M (1973) The Pricing of Options and Corporate Liabilities. J Political Econ 81:637–654
8. Brenner YS, Kaelble H, Thomas M (1991) Income Distribution in Historical Perspective. Cambridge University Press, Cambridge
9. Engle RF (1982) Autoregressive Conditional Heteroskedasticity With Estimates of the Variance of UK Inflation. Econometrica 50:987–1008
10. Feller W (1971) An Introduction to Probability Theory and Its Applications, 2nd edn, vol 2. Wiley, New York

11. Friedlander SK (1977) Smoke, dust and haze: Fundamentals of aerosol behavior. Wiley-Interscience, New York

12. Fujiwara Y (2004) Zipf Law in Firms Bankruptcy. Phys A 337:219–230

13. Fujiwara Y, Guilmi CD, Aoyama H, Gallegati M, Souma W (2004) Do Pareto-Zipf and Gibrat laws hold true? An analysis with European firms. Phys A 335:197–216

14. Gilvarry JJ, Bergstrom BH (1961) Fracture of Brittle Solids. II Distribution Function for Fragment Size in Single Fracture (Experimental). J Appl Phys 32:400–410

15. Gopikrishnan P, Meyer M, Amaral LAN, Stanley HE (1998) Inverse Cubic Law for the Distribution of Stock Price Variations. Eur Phys J B 3:139–143

16. Handel PH (1975) $1/f$ Noise-An "Infrared" Phenomenon. Phys Rev Lett 34:1492–1495

17. Havlin S, Selinger RB, Schwartz M, Stanley HE, Bunde A (1988) Random Multiplicative Processes and Transport in Structures with Correlated Spatial Disorder. Phys Rev Lett 61:1438–1441

18. Hidalgo CA, Klinger RB, Barabasi AL, Hausmann R (2007) The Product Space Conditions the Development of Nations. Science 317:482–487

19. Inaoka H, Toyosawa E, Takayasu H (1997) Aspect Ratio Dependence of Impact Fragmentation. Phys Rev Lett 78:3455–3458

20. Inaoka H, Ninomiya T, Taniguchi K, Shimizu T, Takayasu H (2004) Fractal Network derived from banking transaction – An analysis of network structures formed by financial institutions. Bank of Japan Working Paper. http://www.boj.or.jp/en/ronbun/04/data/wp04e04.pdf

21. Inaoka H, Takayasu H, Shimizu T, Ninomiya T, Taniguchi K (2004) Self-similarity of bank banking network. Phys A 339:621–634

22. Klass OS, Biham O, Levy M, Malcai O, Solomon S (2006) The Forbes 400 and the Pareto wealth distribution. Econ Lett 90:290–295

23. Lee Y, Amaral LAN, Canning D, Meyer M, Stanley HE (1998) Universal Features in the Growth Dynamics of Complex Organizations. Phys Rev Lett 81:3275–3278

24. Mandelbrot BB (1963) The variation of certain speculative prices. J Bus 36:394–419

25. Mandelbrot BB (1982) The Fractal Geometry of Nature. W.H. Freeman, New York

26. Mandelbrot BB (2004) The (mis)behavior of markets. Basic Books, New York

27. Mantegna RN, Stanley HE (2000) An Introduction to Econonomics: Correlations and Complexity in Finance. Cambridge Univ Press, Cambridge

28. Markowitz HM (1952) Portfolio Selection. J Finance 7:77–91

29. Mizuno T, Katori M, Takayasu H, Takayasu M (2001) Statistical laws in the income of Japanese companies. In: Empirical Science of Financial Fluctuations. Springer, Tokyo, pp 321–330

30. Mizuno T, Kurihara S, Takayasu M, Takayasu H (2003) Analysis of high-resolution foreign exchange data of USD-JPY for 13 years. Phys A 324:296–302

31. Mizuno T, Kurihara S, Takayasu M, Takayasu H (2003) Investment strategy based on a company growth model. In: Takayasu H (ed) Application of Econophysics. Springer, Tokyo, pp 256–261

32. Mizuno T, Takayasu M, Takayasu H (2004) The mean-field approximation model of company's income growth. Phys A 332:403–411

33. Mizuno T, Toriyama M, Terano T, Takayasu M (2008) Pareto law of the expenditure of a person in convenience stores. Phys A 387:3931–3935

34. Ohnishi T, Takayasu H, Takayasu M (in preparation)

35. Okuyama K, Takayasu M, Takayasu H (1999) Zipf's law in income distribution of companies. Phys A 269:125–131. http://www.ingentaconnect.com/content/els/03784371;jsessionid=5e5wq937wfsqu.victoria

36. Rieger MO, Wang M (2006) Cumulative prospect theory and the St. Petersburg paradox. Econ Theory 28:665–679

37. Sato AH, Takayasu H (1998) Dynamic numerical models of stock market price: from microscopic determinism to macroscopic randomness. Phys A 250:231–252

38. Sato AH, Takayasu H, Sawada Y (2000) Power law fluctuation generator based on analog electrical circuit. Fractals 8:219–225

39. Sinha S, Pan RK (2008) How a "Hit" is Born: The Emergence of Popularity from the Dynamics of Collective Choice. http://arxiv.org/PS_cache/arxiv/pdf/0704/0704.2955v1.pdf

40. Souma W (2001) Universal structure of the personal income distribution. Fractals 9:463–470; http://www.nslij-genetics.org/j/fractals.html

41. Takayasu H (1989) Steady-state distribution of generalized aggregation system with injection. Phys Rev Lett 63:2563–2566

42. Takayasu H (1990) Fractals in the physical sciences. Manchester University Press, Manchester

43. Takayasu H (ed) (2002) Empirical Science of Financial Fluctuations–The Advent of Econophysics. Springer, Tokyo

44. Takayasu H (ed) (2003) Application of Econophysics. Springer, Tokyo

45. Takayasu H, Inaoka H (1992) New type of self-organized criticality in a model of erosion. Phys Rev Lett 68:966–969

46. Takayasu H, Takayasu M, Provata A, Huber G (1991) Statistical properties of aggregation with injection. J Stat Phys 65:725–745

47. Takayasu H, Miura H, Hirabayashi T, Hamada K (1992) Statistical properties of deterministic threshold elements–The case of market price. Phys A 184:127–134

48. Takayasu H, Sato AH, Takayasu M (1997) Stable infinite variance fluctuations in randomly amplified Langevin systems. Phys Rev Lett 79:966–969

49. Takayasu M, Taguchi Y, Takayasu H (1994) Non-Gaussian distribution in random transport dynamics. Inter J Mod Phys B 8:3887–3961

50. Takayasu M (2003) Self-modulation processes in financial market. In: Takayasu H (ed) Application of Econophysics. Springer, Tokyo, pp 155–160

51. Takayasu M (2005) Dynamics Complexity in Internet Traffic. In: Kocarev K, Vatty G (eds) Complex Dynamics in Communication Networks. Springer, New York, pp 329–359

52. Takayasu M, Takayasu H (2003) Self-modulation processes and resulting generic $1/f$ fluctuations. Phys A 324:101–107

53. Takayasu M, Mizuno T, Takayasu H (2006) Potentials force observed in market dynamics. Phys A 370:91–97

54. Takayasu M, Mizuno T, Takayasu H (2007) Theoretical analysis of potential forces in markets. Phys A 383:115–119

55. Takayasu M, Mizuno T, Watanabe K, Takayasu H (preprint)

56. Tsallis C (1988) Possible generalization of Boltzmann–Gibbs statistics. J Stat Phys 52:479–487; http://en.wikipedia.org/wiki/Boltzmann_entropy

57. Ueno H, Mizuno T, Takayasu M (2007) Analysis of Japanese bank's historical network. Phys A 383:164–168
58. Yamada K, Takayasu H, Takayasu M (2007) Characterization of foreign exchange market using the threshold-dealer-model. Phys A 382:340–346

Books and Reviews

Takayasu H (2006) Practical Fruits of Econophysics. Springer, Tokyo
Chatterjee A, Chakrabarti BK (2007) Econophysics of Markets and Business Networks (New Economic Windows). Springer, New York

Fractals in Geology and Geophysics

DONALD L. TURCOTTE
Department of Geology,
University of California,
Davis, USA

Article Outline

Glossary

Fractal A collection of objects that have a power-law dependence of number on size.

Fractal dimension The power-law exponent in a fractal distribution.

Definition of the Subject

The scale invariance of geological phenomena is one of the first concepts taught to a student of geology. When a photograph of a geological feature is taken, it is essential to include an object that defines the scale, for example, a coin or a person. It was in this context that Mandelbrot [7] introduced the concept of fractals. The length of a rocky coastline is obtained using a measuring rod with a specified length. Because of scale invariance, the length of the coastline increases as the length of the measuring rod decreases according to a power law. It is not possible to obtain a specific value for the length of a coastline due to small indentations down to a scale of millimeters or less.

A fractal distribution requires that the number of objects N with a linear size greater than r has an inverse power-law dependence on r so that

$$N = \frac{C}{r^D} \tag{1}$$

where C is a constant and the power D is the fractal dimension. This power-law scaling is the only distribution that is scale invariant. However, the power-law dependence cannot be used to define a statistical distribution because the integral of the distribution diverges to infinity either for large values or small values of r. Thus fractal distributions never appear in compilations of statistical distributions. A variety of statistical distributions have power-law behavior either at large scales or small scales, but not both. An example is the Pareto distribution.

Many geological phenomena are scale invariant. Examples include the frequency-size distributions of fragments, faults, earthquakes, volcanic eruptions, and landslides. Stream networks and landforms exhibit scale invariance. In terms of these applications there must always be upper and lower cutoffs to the applicability of a fractal distribution. As a specific application consider earthquakes on the Earth. The number of earthquakes has a power-law dependence on the size of the rupture over a wide range of sizes. But the largest earthquake cannot exceed the size of the Earth, say 10^4 km. Also, the smallest earthquake cannot be smaller than the grain size of rocks, say 1 mm. But this range of scales is 10^{10}. Actual earthquakes appear to satisfy fractal scaling over the range 1 m to 10^3 km.

An example of fractal scaling is the number-area distribution of lakes [10], this example is illustrated in Fig. 1. Excellent agreement with the fractal relation given in Eq. (1) is obtained taking $D = 1.90$. The linear dimension r is taken to be the square root of the area A and the power-law (fractal) scaling extends from $r = 100$ m to $r = 300$ km.

Introduction

Fractal scaling evolved primarily as an empirical means of correlating data. A number of examples are given below. More recently a theoretical basis has evolved for the applicability of fractal distributions. The foundation of this basis is the concept of self-organized criticality. A number of simple computational models have been shown to
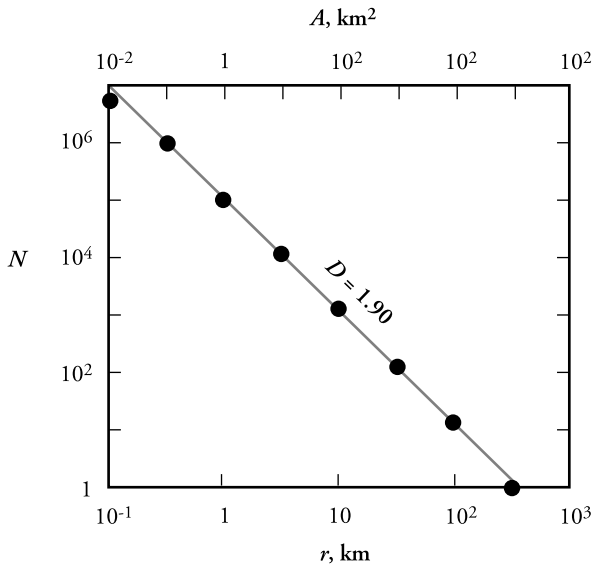
Fractals in Geology and Geophysics, Figure 1
Dependence of the cumulative number of lakes N with areas greater than A as a function of A. Also shown is the linear dimension r which is taken to be the square root of A. The straight-line correlation is with Eq. (1) taking the fractal dimension $D = 1.90$

yield fractal distributions. Examples include the sand-pile model, the forest-fire model, and the slider-block model.

Drainage Networks

Drainage networks are a universal feature of landscapes on the Earth. Small streams merge to form larger streams, large streams merge to form rivers, and so forth. Strahler [16] quantified stream networks by introducing an ordering system. When two like-order streams of order i merge they form a stream of order $i + 1$. Thus two $i = 1$ streams merge to form a $i = 2$ stream, two $i = 2$ streams merge to from a $i = 3$ stream and so forth. A bifurcation ratio R_b is defined by

$$R_b = \frac{N_i}{N_{i+1}} \qquad (2)$$

where N_i is the number of streams of order i. A length order ratio R_r is defined by

$$R_r = \frac{r_{i+1}}{r_i} \qquad (3)$$

where r_i is the mean length of streams of order i. Empirically both R_b and R_r are found to be nearly constant for a range of stream orders in a drainage basin. From Eq. (1) the fractal dimension of a drainage basin

$$D = \frac{\ln(N_i/N_{i+1})}{\ln(r_{i+1}/r_i)} = \frac{\ln R_b}{\ln R_r} \qquad (4)$$

Typically $R_b = 4.6$, $R_r = 2.2$, and the corresponding fractal dimension is $D = 1.9$. This scale invariant scaling of drainage networks was recognized some 20 years before the concept of fractals was introduced in 1967.

A major advance in the quantification of stream networks was made by Tokunaga [17]. This author was the first to recognize the importance of side branching, that is some $i = 1$ streams intersect $i = 2$, $i = 3$, and all higher-order streams. Similarly, $i = 2$ streams intersect $i = 3$ and higher- order streams and so forth. A fully self-similar, side-branching topology was developed. Applications to drainage networks have been summarized by Peckham [11] and Pelletier [13].

Fragmentation

An important application of power-law (fractal) scaling is to fragmentation. In many examples the frequency-mass distributions of fragments are fractal. Explosive fragmentation of rocks (for example in mining) give fractal distributions. At the largest scale the frequency size distribution of the tectonic plates of plate tectonics are reasonably well approximated by a power-law distribution. Fault gouge is generated by the grinding process due to earthquakes on a fault. The frequency-mass distribution of the gouge fragments is fractal. Grinding (comminution) processes are common in tectonics. Thus it is not surprising that fractal distributions are ubiquitous in geology.

As a specific example consider the frequency-mass distribution of asteroids. Direct measurements give a fractal distribution. Since asteroids are responsible for the impact craters on the moon, it is not surprising that the frequency-area distribution of lunar craters is also fractal.

Using evidence from the moon and a fractal extrapolation it is estimated that on average, a 1m diameter meteorite impacts the earth every year, that a 100 m diameter meteorite impacts every 10,000 years, and that a 10 km diameter meteorite impacts the earth every 100,000,000 years. The classic impact crater is Meteor Crater in Arizona, it is over 1 km wide and 200 m deep. Meteor Crater formed about 50,000 years ago and it is estimated that the impacting meteorite had a diameter of 30 m. The largest impact to occur in the 20th century was the June 30, 1908 Tunguska event in central Siberia. The impact was observed globally and destroyed over 1000 km^2 of forest. It is believed that this event was the result of a 30 m diameter meteorite that exploded in the atmosphere.

One of the major global extinctions occurred at the Cretaceous/Tertiary boundary 65 million years ago. Some 65% of the existing species were destroyed including dinosaurs. This extinction is attributed to a massive impact

at the Chicxulub site on the Yucatan Peninsula, Mexico. It is estimated that the impacting meteorite had a 10 km diameter. In addition to the damage done by impacts there is evidence that impacts on the oceans have created massive tsunamis. The fractal power-law scaling can be used to quantify the risk of future impacts.

Earthquakes

Earthquakes universally satisfy several scaling laws. The most famous of these is Gutenberg–Richter frequency-magnitude scaling. The magnitude M of an earthquake is an empirical measure of the size of an earthquake. If the magnitude is increased by one unit it is observed that the cumulative number of earthquakes greater than the specified magnitude is reduced by a factor of 10.

For the entire earth, on average, there is 1 magnitude 8 earthquake per year, 10 magnitude 7 earthquakes per year, and 100 magnitude 6 earthquakes per year. When magnitude is converted to rupture area a fractal relation is obtained. The numbers of earthquakes that occur in a specified region and time interval have a power-law dependence on the rupture area.

The validity of this fractal scaling has important implications for probabilistic seismic risk assessment. The number of small earthquakes that occur in a region can be extrapolated to estimate the risk of larger earthquakes [1]. As an example consider southern California. On average there are 30 magnitude 4 or larger earthquakes per year. Using the fractal scaling it is estimated that the expected intervals between magnitude 6 earthquakes will be 3 years, between magnitude 7 earthquakes will be 30 years, and between magnitude 8 earthquakes will be 300 years.

The fractal scaling of earthquakes illustrate a useful aspect of fractal distributions. The fractal distribution requires two parameters. The first parameter, the fractal dimension D (known as the b-value in seismology), gives the dependence of number on size (magnitude). For earthquakes the fractal dimension is almost constant independent of the tectonic setting. The second parameter gives the level of activity. For example, this can be the number of earthquakes greater than a specified magnitude in a region. This level of activity varies widely and is an accepted measure of seismic risk. The level is essentially zero in states like Minnesota and is a maximum in California.

Volcanic Eruptions

There is good evidence that the frequency-volume statistics of volcanic eruptions are also fractal [9]. Although it is difficult to quantify the volumes of magma and ash associated with older eruptions, the observations suggest that an eruption with a volume of 1 km^3 would be expected each 10 years, 10 km^3 each 100 years, and 100 km^3 each 1000 years. For example, the 1991 Mount Pinatubo, Philippines eruption had an estimated volume of about 5 km^3. The most violent eruption in the last 200 years was the 1815 Tambora, Indonesia eruption with an estimated volume of 150 km^3. This eruption influenced the global climate in 1816 which was known as the year without a summer. It is estimated that the Long Valley, California eruption with an age of about 760,000 years had a volume of about 600 km^3 and the Yellowstone eruptions of about 600,000 years ago had a volume of about 2000 km^3.

Although the validity of the power-law (fractal) extrapolation of volcanic eruption volumes to long periods in the past can be questioned, the extrapolation does give some indication of the risk of future eruptions to global climate. There is no doubt that the large eruptions that are known to have occurred on time scales of 10^5 to 10^6 years would have a catastrophic impact on global agricultural production.

Landslides

Landslides are a complex natural phenomenon that constitutes a serious natural hazard in many countries. Landslides also play a major role in the evolution of landforms. Landslides are generally associated with a trigger, such as an earthquake, a rapid snowmelt, or a large storm. The landslide event can include a single landslide or many thousands. The frequency-area distribution of a landslide event quantifies the number of landslides that occur at different sizes. It is generally accepted that the number of large landslides with area A has a power-law dependence on A with an exponent in the range 1.3 to 1.5 [5].

Unlike earthquakes, a complete statistical distribution can be defined for landslides. A universal fit to an inverse-gamma distribution has been found for a number of event inventories. This distribution has a power-law (fractal) behavior for large landslides and an exponential cut-off for small landslides. The most probable landslides have areas of about 40 m^2. Very few small landslides are generated.

As a specific example we consider the 11,111 landslides generated by the magnitude 6.7 Northridge (California) earthquake on January 17, 1994. The total area of the landslides was 23.8 km^2 and the area of the largest landslide was 0.26 km^2. The inventory of landslide areas had a good power-law dependence on area for areas greater than 10^3 m^2 (10^{-3} km^2). The number of landslides generated by earthquakes have a strong dependence on earthquake magnitude. Typically earthquakes with magnitudes M less than 4 do not generate any landslides [6].

Floods

Floods are a major hazard to many cities and estimates of flood hazards have serious economic implications. The standard measure of the flood hazard is the 100-year flood. This is quantified as the river discharge Q_{100} expected during a 100 year period. Since there is seldom a long enough history to establish Q_{100} directly, it is necessary to extrapolate smaller floods that occur more often.

One extrapolation approach is to assume that flood discharges are fractal (power-law) [3,19]. This scale invariant distribution can be expressed in terms of the ratio F of the peak discharge over a 10 year interval to the peak discharge over a 1 year interval, $F = Q_{10}/Q_1$. With self-similarity the parameter F is also the ratio of the 100 year peak discharge to the 10 year peak discharge, $F = Q_{100}/Q_{10}$. Values of F have a strong dependence on climate. In temperate climates such as the northeastern and northwestern US values are typically in the range $F = 2$–3. In arid and tropical climates such as the southwestern and southeastern US values are typically in the range $F = 4$–6.

The applicability of fractal concepts to flood forecasting is certainly controversial. In 1982, the US government adopted the log-Pearson type 3 (LP3) distribution for the legal definition of the flood hazard [20]. The LP3 is a thin-tailed distribution relative to the thicker tailed power-law (fractal) distribution. Thus the forecast 100 year flood using LP3 is considerably smaller than the forecast using the fractal approach. This difference is illustrated by considering the great 1993 Mississippi River flood. Considering data at the Keukuk, Iowa gauging station [4] this flood was found to be a typical 100 year flood using the power-law (fractal) analysis and a 1000 to 10,000 year flood using the federal LP3 formulation. Concepts of self-similarity argue for the applicability of fractal concepts for flood-frequency forecasting. This applicability also has important implications for erosion. Erosion will be dominated by the very largest floods.

Self-Affine Fractals

Mandelbrot and Van Ness [8] extended the concept of fractals to time series. Examples of time series in geology and geophysics include global temperature, the strength of the Earth's magnetic field, and the discharge rate in a river. After periodicities and trends have been removed, the remaining values are the stochastic (noise) component of the time series. The standard approach to quantifying the noise component is to carry out a Fourier transform on the time series [2]. The power-spectral density coefficients S_i are proportional to the squares of the Fourier coefficients. The time series is a self-affine fractal if the power-spectral density coefficients have an inverse power-law dependence on frequency f_i, that is

$$S_i = \frac{C}{f_i^{\beta}} \tag{5}$$

where C is a constant and β is the power-law exponent.

For a Gaussian white noise the values in the time series are selected randomly from a Gaussian distribution. Adjacent values are not correlated with each other. In this case the spectrum is flat and the power spectral density coefficients are not a function of frequency, $\beta = 0$. The classic example of a self-affine fractal is a Brownian walk. A Brownian walk is obtained by taking the running sum of a Gaussian white noise. In this case we have $\beta = 2$. Another important self-affine time series is a red (or pink) noise with power spectral density coefficients proportional to $1/f$, that is $\beta = 1$. We will see that the variability in the Earth's magnetic field is well approximated by a $1/f$ noise.

Self-affine fractal time series in the range $\beta = 0$ to 1 are known as fractional Gaussian noises. These noises are stationary and the standard deviation is a constant independent of the length of the time series. Self-affine time series with β larger than 1 are known as fractional Brownian walk. These motions are not stationary and the standard deviation increases as a power of the length of the time series, there is a drift. For a Brownian walk the standard deviation increases with the square root of the length of the time series.

Topography

The height of topography along linear tracks can be considered to be a continuous time series. In this case we consider the wave number k_i (1/wave length) instead of frequency. Topography is a self-affine fractal if

$$S_i = \frac{C}{k_i^{\beta}} \tag{6}$$

Spectral expansions of global topography have been carried out, an example [15] is given in Fig. 2. Excellent agreement with the fractal relation given in Eq. (6) is obtained taking $\beta = 2$, topography is well approximated by a Brownian walk. It has also shown that this fractal behavior of topography is found for the moon, Venus, and Mars [18].

Earth's Magnetic Field

Paleomagnetic studies have given the strength and polarity of the Earth's magnetic field as a function of time over millions of years. These studies have also shown that the field has experienced a sequence of reversals.
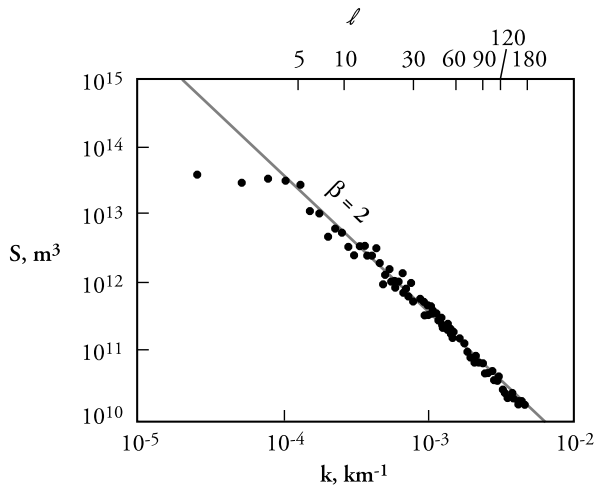
Fractals in Geology and Geophysics, Figure 2
Power spectral density S as a function of wave number k for a spherical harmonic expansion of the Earth's topography (degree l). The straight-line correlation is with Eq. (6) taking $\beta = 2$, a Brownian walk

Spectral studies of the absolute amplitude of the field have been shown that it is a self-affine fractal [12,14]. The power-spectral density is proportional to one over the frequency, it is a $1/f$ noise. When the fluctuations of the $1/f$ noise take the magnitude to zero the polarity of the field reverses. The predicted distribution of polarity intervals is fractal and is in good agreement with the observed polarity intervals.

Future Directions

There is no question that fractals are a useful empirical tool. They provide a rational means for the extrapolation and interpolation of observations. The wide applicability of power-law (fractal) distributions is generally accepted, but does this applicability have a more fundamental basis? Fractality appears to be fundamentally related to chaotic behavior and to numerical simulations exhibiting self-organized criticality. The entire area of fractals, chaos, self-organized criticality, and complexity remains extremely active, and it is impossible to predict with certainty what the future holds.

Bibliography

Primary Literature

1. Kossobokov VG, Keilis-Borok VI, Turcotte DL, Malamud BD (2000) Implications of a statistical physics approach for earthquake hazard assessment and forecasting. Pure Appl Geophys 157:2323
2. Malamud BD, Turcotte DL (1999) Self-affine time series: I. Generation and analyses. Adv Geophys 40:1
3. Malamud BD, Turcotte DL (2006) The applicability of power-law frequency statistics to floods. J Hydrol 332:168
4. Malamud BD, Turcotte DL, Barton CC (1996) The 1993 Mississippi river flood: A one hundred or a one thousand year event? Env Eng Geosci 2:479
5. Malamud BD, Turcotte DL, Guzzetti F, Reichenbach P (2004) Landslide inventories and their statistical properties. Earth Surf Process Landf 29:687
6. Malamud BD, Turcotte DL, Guzzetti F, Reichenbach P (2004) Landslides, earthquakes, and erosion. Earth Planet Sci Lett 229:45
7. Mandelbrot BB (1967) How long is the coast of Britain? Statistical self-similarity and fractional dimension. Science 156:636
8. Mandelbrot BB, Van Ness JW (1968) Fractional Brownian motions, fractional noises and applications. SIAM Rev 10:422
9. McClelland L et al (1989) Global Volcanism 1975-1985. Prentice-Hall, Englewood Cliffs
10. Meybeck M (1995) Global distribution of lakes. In: Lerman A, Imboden DM, Gat JR (eds) Physics and Chemistry of Lakes, 2nd edn. Springer, Berlin, pp 1-35
11. Peckham SD (1989) New results for self-similar trees with applications to river networks. Water Resour Res 31:1023
12. Pelletier JD (1999) Paleointensity variations of Earth's magnetic field and their relationship with polarity reversals. Phys Earth Planet Int 110:115
13. Pelletier JD (1999) Self-organization and scaling relationships of evolving river networks. J Geophys Res 104:7259
14. Pelletier JD, Turcotte DL (1999) Self-affine time series: II. Applications and models. Adv Geophys 40:91
15. Rapp RH (1989) The decay of the spectrum of the gravitational potential and the topography of the Earth. Geophys J Int 99:449
16. Strahler AN (1957) Quantitative analysis of watershed geomorphology. Trans Am Geophys Un 38:913
17. Tokunaga E (1978) Consideration on the composition of drainage networks and their evolution. Geogr Rep Tokyo Metro Univ 13:1
18. Turcotte DL (1987) A fractal interpretation of topography and geoid spectra on the earth, moon, Venus, and Mars. J Geophys Res 92:E597
19. Turcotte DL (1994) Fractal theory and the estimation of extreme floods. J Res Natl Inst Stand Technol 99:377
20. US Water Resources Council (1982) Guidelines for Determining Flood Flow Frequency. Bulletin 17B. US Geological Survey, Reston

Books and Reviews

Feder J (1988) Fractals. Plenum Press, New York
Korvin G (1992) Fractal Models in the Earth Sciences. Elsevier, Amsterdam
Mandelbrot BB (1982) The Fractal Geometry of Nature. Freeman, San Francisco
Turcotte DL (1997) Fractals and Chaos in Geology and Geophysics, 2nd edn. Cambridge University Press, Cambridge

Fractals Meet Chaos

TONY CRILLY
Middlesex University, London, UK

Article Outline

Glossary

Dimension The traditional meaning of dimension in modern mathematics is "topological dimension" and is an extension of the classical Greek meaning. In modern concepts of this dimension can be defined in terms of a separable metric space. For the practicalities of Fractals and Chaos the notion of dimension can be limited to subsets of Euclidean n-space where n is an integer. The newly arrived "fractal dimension" is *metrically* based and can take on fractional values. Just as for topological dimension.

As for topological dimension itself there is a profusion of different (but related) concepts of metrical dimension. These are widely used in the study of fractals, the ones of principal interest being:

- Hausdorff dimension (more fully, Hausdorff–Besicovitch dimension),
- Box dimension (often referred to as Minkowski–Bouligand dimension),
- Correlation dimension (due to A. Rényi, P. Grassberger and I. Procaccia).

Other types of metric dimension are also possible. There is "divider dimension" (based on ideas of an English mathematician/meteorologist L. F. Richardson in the 1920s); the "Kaplan–Yorke dimension" (1979) derived from Lyapunov exponents, known also as the "Lyapunov dimension"; "packing dimension" introduced by Tricot (1982). In addition there is an overall *general* dimension due to A. Rényi (1970) which admits box dimension, correlation dimension and information dimension as special cases. With many

of the concepts of dimension there are upper and lower refinements, for example, the separate *upper* and *lower* box dimensions. Key references to the vast (and highly technical) subject of mathematical dimension include [31,32,33,60,73,92,93].

Hausdorff dimension (Hausdorff–Besicovitch dimension). In the study of fractals, the most sophisticated concept of dimension is Hausdorff dimension, developed in the 1920s.

The following definition of Hausdorff dimension is given for a subset A of the real number line. This is readily generalized to subsets of the plane, Euclidean 3-space and Euclidean n-space, and more abstractly to separable metric spaces by taking neighborhoods as disks instead of intervals. Let $\{U_i\}$ be an *r-covering* of A, (a covering of A where the width of *all* intervals U_i, satisfies $w(U_i) \leq r$). The measure m_r is defined by

$$m_r(A) = \inf\left(\sum_{i=1}^{\infty} w(U_i)\right),$$

where the infimum (or greatest of the minimum values) is taken over all *r-coverings* of A.
The Hausdorff dimension D_H of A is:

$$D_H = \lim_{r \to 0} m_r(A),$$

provided the limit exists. The subset $E = \{1/n: n = 1, 2, , 3, \ldots\}$ of the unit interval has $D_H = 0$ (the Hausdorff dimension of a countable set is always zero). The Hausdorff dimension is the basis of "fractal dimension" but because it takes into account intervals of unequal widths it may be difficult to calculate in practice.

Box dimension (or Minkowski–Bouligand dimension, known also as capacity dimension, cover dimension, grid dimension). The box counting dimension is a more direct and practical method for computing dimension in the case of fractals. To define it, we again confine our attention to the real number line in the knowledge that box dimension is readily extended to subsets of more general spaces.
As before, let $\{U_i\}$ be an *r-covering* of A, and let $N_r(A)$ be the least number of sets in such a covering. The box dimension D_B of A is defined by:

$$D_B = \lim_{r \to 0} \frac{\log N_r(A)}{\log 1/r}.$$

The box dimension of the subset $E = \{1/n: n = 1, 2, 3, \ldots\}$ of the unit interval can be calculated to give $D_B = 0.5$.

In general Hausdorff and Box dimensions are related to each other by the inequality $D_B \geq D_H$, as happens in the above example. The relationship between D_H and D_B is investigated in [49]. For compact, self-similar fractal sets $D_B = D_H$ but there are fractal sets for which $D_B > D_H$ [76]. Though Hausdorff dimension and Box dimension have similar properties, Box dimension is only finitely additive, while Hausdorff dimension is countably additive.

Correlation dimension For a set of n points A on the real number line, let $P(r)$ be the probability that two different points of A chosen at random are closer than r apart. For a large number of points n, the graph of $v = \log P(r)$ against $u = \log r$ is approximated to its slope for small values of r, and theoretically, a straight line. The correlation dimension D_C is defined as its slope for small values of r, that is,

$$D_C = \lim_{r \to 0} \frac{dv}{du} \, .$$

The correlation dimension involves the separation of points into "boxes", whereas the box dimension merely counts the boxes that cover A.

If P_i is the probability of a point of A being in box i (approximately n_i/n where n_i is the number of points in box i and n is the totality of points in A) then an alternative definition of correlation dimension is

$$D_C = \lim_{r \to 0} \frac{\log \sum_i P_i^2}{\log r} \, .$$

Attractor A point set in phase space which "attracts" trajectories in its vicinity. More formally, a bounded set A in phase space is called an attractor for the solution $x(t)$ of a differential equation if

- $x(0) \in A \Rightarrow x(t) \in A \ for \ all \ \ t$. Thus, an attractor is invariant under the dynamics (trajectories which start in A remain in A).
- There is a neighborhood $U \supset A$ such that any trajectory starting in U is attracted to A (the trajectory gets closer and closer to A).
- If $B \subset A$ and if B satisfies the above two properties then $B = A$.

An attractor is therefore the minimal set of points A which attracts all orbits starting at some point in a neighborhood of A.

Orbit is a sequence of points $\{x_i\} = x_0, x_1, x_2, \ldots$ defined by an iteration $x_n = f^n(x_0)$. If n is a positive it is called a forwards orbit, and if n is negative a backwards orbit. If $x_0 = x_n$ for some finite value n, the orbit is periodic. In this case, the smallest value of n for

which this is true is called the period of the orbit.

For an invertible function f, a point x is homoclinic to a if

$$\lim f^n(x) = \lim f^{-n}(x) = a \quad \text{as} \ \ n \to \infty \, .$$

and in this case the orbit $\{f^n(x_0)\}$ is called a homoclinic orbit – the orbit which converges to the same saddle point a forwards or backwards. This term was introduced by Poincaré.

The terminology "orbit" may be regarded as applied to the solution of a difference equation, in a similar way the solution of a differential equation $x(t)$ is termed a trajectory. Orbit is the term used for discrete dynamical system and trajectory for the continuous time case.

Basin of attraction If a is an attractive fixed point of a function f, its basin of attraction $B(a)$ is the subset of points defined by

$$B(a) = \{x \colon f^k(x) \to a \, , \ \text{as} \ k \to \infty\} \, .$$

It is the subset containing all the initial points of orbits attracted to a. The basins of attraction may have a complicated structure. An important example applies to the case where a is a point in the complex plane C.

Julia set A set J_f is the boundary between the basins of attraction of a function f. For example, in the case where $z = \pm 1$ are attracting points (solutions of $z^2 - 1 = 0$), the Julia set of the "Newton–Fourier" function $f(z) = z - ((z^2 - 1)/2z)$ is the set of complex numbers which lie along the imaginary axis $x = 0$ (as proved by Schröder and Cayley in the 1870s). The case of the Julia set involved with the solutions of $z^3 - 1 = 0$ was beyond these pioneers and is fractal in nature. An alternative definition for a Julia set is the closure of the subset of the complex plane whose orbits of f tend to infinity.

Definition of the Subject

Though "Chaos" and "Fractals" are yoked together to form "Fractals and Chaos" they have had separate lines of development. And though the names are modern, the mathematical ideas which lie behind them have taken more than a century to gain the prominence they enjoy today. Chaos carries an applied connotation and is linked to differential equations which model physical phenomena. Fractals is directly linked to subsets of Euclidean space which have a fractional dimension, which may be obtained by the iteration of functions.

This brief survey seeks to highlight some of the significant points in the history of both of these subjects. There are brief academic histories of the field. A history of

Chaos has been shown attention [4,46], while an account of the early history of the iteration of complex functions (up to Julia and Fatou) is given in [3]. A broad survey of the whole field of fractals is given in [18,50]. Accounts of Chaos and Fractals tend to concentrate on one side at the expense of the other. Indeed, it is only quite recently that the two subjects have been seen to have a substantial bearing on each other [72].

This account treats a "prehistory" and a "modern period". In the prehistory period, before about 1960, topics which contributed to the modern theory are not so prominent. There is a lack of impetus to create a new field. In the modern period there is a greater sense of continuity, driven on by the popular interest in the subject. The modern theory coincided with a rapid increase in the power of computers unavailable to the early workers in the prehistory period. Scientists, mathematicians, and a wider public could now "see" the beautiful geometrical shapes displayed before them [25].

The whole theory of "fractals and chaos" necessarily involves nonlinearity. It is a mathematical theory based on the properties of processes which are assumed to be modeled by nonlinear differential equations and nonlinear functions. Chaos shows itself when solutions to these differential equations become unstable. The study of stability is rooted in the mathematics of the nineteenth century. Fractals are derived from the geometric study of curves and sets of points generally, and from abstract iterative schemes. The modern theory of fractals is the outcome of explorations by mathematicians and scientists in the 1960s and 1970s, though, as we shall see, it too has an extensive prehistory.

Recently, there has been an explosion of published work in Fractals and Chaos. Just in Chaos theory alone a bibliography of six hundred articles and books compiled by 1982, grew to over seven thousand by the end of 1990 [97]. This is most likely an excessive underestimate. Fractals and Chaos has since grown into a wide-ranging and variegated theory which is rapidly developing. It has widespread applications in such areas as

Astronomy the motions of planets and galaxies

Biology population dynamics chemistry chemical reactions

Economics time series, analysis of financial markets

Engineering capsize of ships, analysis of road traffic flow

Geography measurement of coastlines, growth of cities, weather forecasting

Graphic art analysis of early Chinese Landscape Paintings, fractal geometry of music

Medicine dynamics of the brain, psychology, heart rhythms.

There are many works which address the application of Chaos and Fractals to a broad sweep of subjects. In particular, see [13,19,27,59,63,64,87,96].

Introduction

"Fractals and Chaos" is the popular name for the subject which burst onto the scientific scene in the 1970s and 1980s and drew together practical scientists and mathematicians. The subject reached the popular ear and perhaps most importantly, its eye. Computers were becoming very powerful and capable of producing remarkable visual images. Quite elementary mathematics was capable of creating beautiful pictures that the world had never seen before.

The popular ear was captivated – momentarily at least – by the neologisms created for the theory. "Chaos" was one, but "fractals", the "horseshoe", and the superb "strange attractors" became an essential part of the scientific vocabulary. In addition, these were being passed around by those who dwelled far from the scientific front. It seemed all could take part, from scientific and mathematical researchers to lay scientists and amateur mathematicians. All could at least experience the excitement the new subject offered. Fractals and Chaos also carried implications for the philosophy of science.

The widespread interest in the subject owes much to articles and books written for the popular market. J. Gleick's *Chaos* was at the top of the heap in this respect and became a best seller. He advised his readership that chaos theory was one of the great discoveries of the twentieth century and quoted scientists who placed chaos theory alongside the revolutions of Relativity and Quantum Mechanics. Gleick claimed that "chaos [theory] eliminates the Laplacean fantasy of deterministic predictability". In a somewhat speculative flourish, he reasoned: "Of the three theories, the revolution in chaos applies to the universe we see and touch, to objects at human scale. Everyday experience and real pictures of the world become legitimate targets for inquiry. There has long been a feeling, not always expressed openly, that theoretical physics has strayed far from human intuition about the world [38]."

More properly chaos is "deterministic chaos". The equations and functions used to model a dynamical system are stated exactly. The contradictory nature of chaos is that the mathematical solutions appeared to be random. On the one hand the situation is deterministic, but on the other there did not seem to be any order in the solutions. Chaos theory resolves this difficulty by conceptualizing the notion of orbits, trajectories phase space, attractors, and fractals. What appeared paradoxical seen through the

old lenses offered explanation through a more embracing mathematical theory.

In the scientific literature the modern Chaos is presented through "dynamical systems", and this terminology gives us a clue to its antecedents. Dynamical systems conveys the idea of a physical system. It is clear that mechanical problems, for example, those involving motion are genuine dynamical systems because they evolve in space through time. Thus the pendulum swings backwards and forwards in time, planets trace out orbits, and the beat of a heart occurs in time.

The differential equations which describe physical dynamical system give rise to chaos, but how do fractals enter the scene? In short, the trajectories in the phase space which describes the physical system through a process of spreading and folding pass through neighborhoods of the attractor set infinitely often, and on magnification reveal them running closer and closer together. This is the fine structure of the attractor. Measuring the metric dimension of this set results in a fraction and it is there that the connection with fractals is made. But this is running ahead of the story.

Dynamical Systems

The source of "Chaos" lies in the analysis of physical systems and goes back to the eighteenth century and the work of the Swiss mathematician Leonhard Euler. The most prolific mathematician of all time, Euler (whose birth tercentenary occurred in 2007) was amongst natural philosophers who made a study of differential equations in order to solve practical mechanical and astronomical problems. Chief among the problems is the problem of fluid flow which occurs in hydrodynamics.

Sensitive Initial Conditions

The key characteristic of "chaotic solutions" is their sensitivity to initial conditions: two sets of initial conditions close together can generate very different solution trajectories, which after a long time has elapsed will bear very little relation to each other. Twins growing up in the same household will have a similar life for the childhood years but their lives may diverge completely in the fullness of time. Another image used in conjunction with chaos is the so-called "butterfly effect" – the metaphor that the difference between a butterfly flapping its wings in the southern hemisphere (or not) is the difference between fine weather and hurricanes in Europe. The butterfly effect notion most likely got its name from the lecture E. Lorenz gave in Washington in 1972 entitled "Predictability: Does the Flap of a Butterfly's wings in Brazil Set off a Tornado in

Texas?" [54]. An implication of chaos theory is that prediction in the long term is impossible for we can never know for certain whether the "causal" butterfly really did flap its wings.

The sensitivity of a system to initial conditions, the hallmark of what makes a chaotic solution to a differential equation is derived from the stability of a system. Writing in 1873, the mathematical physicist James Clerk Maxwell alluded to this sensitivity in a letter to man of science Francis Galton:

> Much light may be thrown on some of these questions [of mechanical systems] by the consideration of stability and instability. When the state of things is such that an infinitely small variation of the present state will alter only by an infinitely small quantity the state at some future time, the condition of the system, whether at rest or in motion, is said to be stable; but when an infinitely small variation in the present state may bring about a finite difference in the state of the system in a finite time, the condition is said to be unstable [44].

H. Poincaré too, was well aware that small divergences in initial conditions could result in great differences in future outcomes, and said so in his discourse on *Science and Method*: "It may happen that slight differences in the initial conditions produce very great differences in the final phenomena; a slight error in the former would make an enormous error in the latter. Prediction becomes impossible." As an example of this he looked at the task of weather forecasting:

> Why have the meteorologists such difficulty in predicting the weather with any certainty? Why do the rains, the tempests seem to us to come by chance, so that many persons find it quite natural to pray for rain or shine, when they would think it ridiculous to pray for an eclipse? We see that great perturbations generally happen in regions where the atmosphere is in unstable equilibrium. … one tenth of a degree more or less at any point, and the cyclone bursts here and not there, and spreads its ravages over countries it would have spared. … Here again we find the same contrast between a very slight cause, unappreciable to the observer, and important effects, which are sometimes tremendous disasters [75].

So here is the answer to one conundrum – Poincaré is the true author of the butterfly effect! Weather forecasting is ultimately a problem of fluid flow in this case air flow.

Fluid Motion and Turbulence

The study of the motion of fluids predates Poincaré and Lorenz. Euler published "General Principles of the Motion of Fluids" in 1755 in which he wrote down a set of partial differential equations to describe the motion of non-viscous fluids. These were improved by the French engineer C. Navier when in 1821 he published a paper which took viscosity into account. From a different starting point in 1845 the British mathematical physicist G. G. Stokes derived the same equations in a paper entitled "On the Theories of the Internal Friction of Fluids in Motion."

These are the Navier–Stokes equations, a set of nonlinear partial differential equations which generally apply to fluid motion and which are studied by the mathematical physicist. In modern vector notation (in the case of unforced incompressible flow) they are

$$\frac{\partial \mathbf{v}}{\partial t} + \mathbf{v}.\nabla \mathbf{v} = -\frac{1}{\rho}\nabla p + \frac{\mu}{\rho}\nabla^2 \mathbf{v},$$

where $\mathbf{v}$ is velocity, p is pressure, and ρ and μ are density and viscosity constants. It is the nonlinearity of the Navier–Stokes equations which makes them intractable, the nonlinearity manifested by the "products of terms" like $\mathbf{v}.\nabla\mathbf{v}$ which occur in them. They have been studied intensively since the nineteenth century particularly in special forms obtained by making simplifying assumptions.

L. F. Richardson, who enters the subject of Fractals and Chaos at different points made attempts to solve nonlinear differential equations by numerical methods. In the 1920s, Richardson adapted words from Gulliver's Travels in one of his well-known refrains on turbulence: "big whirls have little whirls that feed on their velocity, and little whirls have lesser whirls and so on to viscosity – in the molecular sense" [79]. This numerical work was hampered by the lack of computing power. The only computers available in the 1920s were human ones, and the traditions of using paid "human computers" was still in operation. Richardson visualized an orchestra of human computers harmoniously carrying out the vast array of calculations under the baton of a mathematician. There were glimmers of all this changing, and during the 1920s the appearance of electrical devices gave an impetus to the mathematical study of both numerical analysis and the study of nonlinear differential equations. John von Neumann for one saw the need for the electronic devices as an aid to mathematics. Notwithstanding this development, the problem of fluid flow posed intrinsic mathematical difficulties.

In 1932, Horace Lamb addressed the British Association for the Advancement of Science, with a prophetic statement dashed with his impish touch of humor: "I am an old man now, and when I die and go to Heaven there are two matters on which I hope for enlightenment. One is quantum electro-dynamics, and the other is the turbulent motion of fluids. And about the former I am really rather optimistic." Forty years on Werner Heisenberg continued in the same vein. He certainly knew about quantum theory, having invented it, but chose relativity as the competing theory with turbulence for his own lamb's tale. On his death bed it is said he compared quantum theory with turbulence and is reputed to have singled out turbulence as of the greater difficulty.

In 1941, A. Kolmogorov, the many-sided Russian mathematician published two papers on problems of turbulence caused by the jet engine and astronomy. Kolmogorov also made contributions to probability and topology though these two papers are the ones rated highly by fluid dynamicists and physicists. In 1946 he was appointed to head the Turbulence Laboratory of the Academic Institute of Theoretical Geophysics. With Kolmolgov, an influential school of mathematicians that included L. S. Pontryagin, A. A. Andronov, D. V. Anosov and V. I. Arnol'd became active in Russia in the field of dynamical systems [24].

Qualitative Differential Equations and Topology

Poincaré's qualitative study of differential equations pioneered the idea of viewing the solution of differential equations as curves rather than functions and replacing the local with the global. Poincaré's viewpoint was revolutionary. As he explained it in *Science and Method*: "formerly an equation was considered solved only when its solution had been expressed by aid of a finite number of known functions; but that is possible scarcely once in a hundred times. What we always can do, or rather what we should always seek to do, is to solve the problem *qualitatively* [his italics] so to speak; that is to say; seek to know the general form of the curve [trajectory] which represents the unknown function."

For the case of the plane, $m = 2$, for instance, what do the solutions to differential equations look like across the whole plane, viewing them as trajectories $x(t)$ starting at an initial point? This contrasts with the traditional view of solving differential equations whereby specific functions are sought which satisfy initial and boundary conditions.

Poincaré's attack on the three body problem (the motion of the moon, earth, sun, is an example) was stimulated by a prize offered in 1885 to commemorate the sixtieth birthday of King Oscar II of Sweden. The problem set was for an *n*-body problem but Poincaré's essay on the re-

stricted three-body problem was judged so significant that he was awarded the prize in January 1889. Before publication, Poincaré went over his working and found a significant error. It proved a profound error. According to Barrow-Green his description of doubly asymptotic trajectories is the first mathematical description of chaotic motion in a dynamical system [6,7]. Poincaré introduced the "Poincaré section" the surface section that transversely intersects the trajectory in phase space and defines a sequence of points on it. In the case of the damped pendulum, for example, a sequence of points is obtained as the spiral trajectory in phase space hits the section. This idea brings a physical dynamical systems problem in conjunction with topology, a theme developed further by Birkhoff in 1927 [1,10].

Poincaré also introduced the idea of phase space but his study mainly revolves around periodic trajectories and not the non periodic ones typical of chaos. Poincaré said of periodic orbits, that they were "the only gap through which may attempt to penetrate, into a place where up to now was reputed to be unreachable". The concentration on periodic trajectories has a parallel in nearby pure mathematics – where irregular curves designated as "monsters" were ignored in favor of "normal curves".

G. D. Birkhoff learned from Poincaré. It was said that, apart from J. Hadamard, no other mathematician knew Poincaré's work as well as Birkhoff did. Like his mentor, Birkhoff adopted phase space as his template, and emphasised periodic solutions and treated conservative systems and not dissipative systems (such as the damped pendulum which loses energy). Birkhoff spent the major part of his career, between 1912 and 1945 contributing to the theory of dynamical systems. His aim was to provide a qualitative theory which characterized equilibrium points in connection with their stability. A dynamical system was defined by a set of n differential equations $dx_i/dt = F_i(x_1, \ldots, x_n)$ defined locally. An interpretation of these could be a situation in chemistry where $x_1(t), \ldots, x_n(t)$ might be the concentrations of n chemical reactants at time t where the initial values $x_1(0), \ldots, x_n(0)$ are given, though applications were not Birkhoff's main concern.

In 1941, towards the end of his career, Birkhoff reappraised the field in "Some unsolved problems of theoretical dynamics" Later M. Morse discussed these and pointed out that Birkhoff listed the same problems he had considered in 1920. In 1920 Birkhoff had written about dynamical systems in terms of the "general analysis" propounded by E. H. Moore in Chicago where he had been a student. In the essay of 1941 modern topological language was used:

As was first realized about fifty years ago by the great French mathematician, Henri Poincaré, the study of dynamical systems (such as the solar system) leads directly to extraordinary diverse and important problems in point-set theory, topology and the theory of functions of real variables.

The idea was to describe the phase space by an abstract topological space. In his talk of 1941, Birkhoff continued:

The kind of abstract space which it seems best to employ is a compact metric space. The individual points represent "states of motion", and each curve of motion represents a complete motion of the abstract dynamical system [11].

Using Birkhoff's reappraisal, Morse set out future goals: " 'Conservative flows' are to be studied both in the topological and the statistical sense, and abstract variational theory is to enter. There is no doubt about the challenge of the field, and the need for a powerful and varied attack [68]."

Duffing, Van der Pol and Radar

It is significant that chaos theory was first derived from practical problems. Poincaré was an early pioneer with a problem in astronomy but other applications shortly arrived. C. Duffing (1918) introduced a second order nonlinear differential equation which described a mechanical oscillating device. In a simple form of it (with zero forcing term):

$$d^2x/dt^2 + \alpha dx/dt + (\beta x^3 + \gamma x) = 0 \,,$$

Duffing's equation exhibits chaotic solutions. B. van der Pol working at the Radio Scientific Research group at the Philips Laboratory at Eindhoven, described "irregular noise" in an electronic diode. Van der Pol's equations of the form (with right hand side forcing term):

$$d^2x/dt^2 + k(x^2 - 1)dx/dt + x = A\cos\Omega t$$

described "relaxational" oscillations or arrhythmic beats of an electrical circuit. Such "relaxational" oscillations are of the type which also occur in the beating of the heart. Richardson noted the transition from periodic solutions to van der Pol's equation (1926) to unstable solutions. Both Duffing's equation and Van der Pol's equation play an important part in chaos theory.

In 1938 the English mathematician Mary Cartwright answered a call for help from the British Department of Scientific and Industrial Research. A solution to the differential equations connected with the new radar technology

was wanted, and van der Pol's equation was relevant. It was in the connection this equation that she collaborated with J. E. Littlewood in the 1940s and made discoveries which presaged modern Chaos.

The mathematical difficulty was caused by the equation being nonlinear but otherwise the equation appear nondescript. Yet in mathematics, the nondescript can yield surprises. A corner of mathematics was discovered that Cartwright described as "a curious branch of mathematics developed by different people from different standpoints – straight mechanics, radio oscillations, pure mathematics and servo-mechanisms of automatic control theory [65]" The mathematician and physicist Freeman Dyson wrote of this work and its practical significance:

> Cartwright had been working with Littlewood on the solutions of the equation, which describe the output of a nonlinear radio amplifier when the input is a pure sine-wave. The whole development of radio in World War II depended on high power amplifiers, and it was a matter of life and death to have amplifiers that did what they were supposed to do. The soldiers were plagued with amplifiers that misbehaved, and blamed the manufacturers for their erratic behavior. Cartwright and Littlewood discovered that the manufacturers were not to blame. The equation itself was to blame. They discovered that as you raise the gain [the ratio of output to input] of the amplifier, the solutions of the equation become more and more irregular. At low power the solution has the same period as the input, but as the power increases you see solutions with double the period, and finally you have solutions that are not periodic at all [30].

The story is now familiar: there is the phenomenon of period doubling of solutions followed by chaotic solutions as the gain of the amplifier is raised still higher. A further contribution to the theory of this equation was made by N. Levinson of the Massachusetts Institute of Technology in the United States.

Curves and Dimension

The subject of "Fractals" is a more recent development. First inklings of them appeared in "Analysis situs" in the latter part of the nineteenth century when the young subject of topology was gaining ground. Questions were being asked about the properties of sets of points in Euclidean spaces, the nature of curves, and the meaning of dimension itself.

Crinkly Curves

In 1872, K. Weierstrass introduced the famous function defined by a convergent series:

$$f(x) = \sum_{k=0}^{\infty} b^k \cos(a^k \pi x) \quad (a > 1, \ 0 < b < 1, \ ab > 1),$$

which was continuous everywhere but differentiable nowhere. It was the original "crinkly curve" but in 1904, H. von Koch produced a much simpler one based only on elementary geometry. While the idea of a function being continuous but not differentiable could be traced back to A-M Ampère at the beginning of the nineteenth century, von Koch's construction has similarities with examples produced by B. Bolzano. Von Koch's curve has become a classic half-way between a "monster curves" and regularity – an example of a curve of infinite length which encloses a finite area, as well as being an iconic fractal.

In retrospect G. Cantor's middle third set (also discovered by H.J.S. Smith (1875) a professor at Oxford), is also a fractal. It is a totally disconnected and uncountable set, with the curiosity that after subtractions of the middle-third segments from the unit interval the ultimate Cantor set has same cardinality as the original unit interval.

At the beginning of the twentieth century searching questions about the theory of curves were being asked. The very basic question "what is a curve" had been brought to life by G. Peano by his space filling curve which is defined in accordance with Jordan's definition of a curve but fills out a "two-dimensional square." Clearly the theory of dimension needed serious attention for how could an ostensibly two-dimensional "filled-in square" be a curve?

The Iteration of Functions

A principal source of fractals is obtained by the iteration of functions, what is now called a discrete dynamical system, or a system of symbolic dynamics [15,23].

One of the earliest forays in this field was made by the English mathematician A. Cayley, but he was not the first as D. S. Alexander has pointed out. F.W. K. E. (Ernst) Schröder anticipated him, and may even have been the inspiration for Cayley's attraction to the problem.

Is there a link between the two mathematicians? Schröder studied at Heidelberg and where L. O. Hesse was his doctoral adviser. Hesse who contributed to algebra, geometry, and invariant theory and was known to Cayley. Nowadays Schröder is known for his contributions to logic but in 1871 he published "Ueber iterite Functionen" in the newly founded *Mathematische Annalen*. The principal objective of this journal was the publication of articles

on invariant theory then a preoccupation of the English and German schools of mathematics, and Cayley was at the forefront of this research.

So did Cayley know of Schröder's work? Cayley had an appetite for all things (pure) mathematical and had acquired encyclopedic knowledge of the field, just about possible in the 1870s. In 1871 Cayley published an article ("Note on the theory of invariants") in the same volume and issue of the *Mathematische Annalen* in which Schröder's article appeared, and actually published three articles in the volume. In this period of his life, when he was in his fifties he allowed his interests full rein and became a veritable magpie of mathematics. He covered a wide range of topics, too many to list here, but his contributions were invariably short. Moreover he dropped invariant theory at this point and only resumed in 1878 when J. J. Sylvester reawakened his interest. There was time to rediscover the four color problem and perhaps Schröder's work on the iteration of functions. It was a field where Schroeder had previously "encountered very few collaborators." The custom of English mathematicians citing previous work of others began in the 1880s and Cayley was one of the first to do this. The fact that Cayley did not cite Schröder is not significant.

In February 1879, Cayley wrote to Sir William Thomson (the later Lord Kelvin) about the Newton–Fourier method of finding the root of an equation, named after I. Newton (c.1669) and J. Fourier (1818) that dealt with the real variable version of the method. This achieved a degree of significance in Cayley's mind, for a few days later he wrote to another colleague about it:

> I have a beautiful question which is bothering me – the extension of the Newton–Fourier method of approximation to imaginary values: it is very easy and pretty for a quadric equation, but I do not yet see how it comes out for a cubic. The general notion is that the plane must be divided into regions; such that starting with a point P in one of these say the A-region … [his ellipsis], the whole series of derived points $P_1, P_2, P_3, \ldots$ up is P_∞ (which will be the point A) lies in this [planar] region; … and so for the B. and C. regions. But I do not yet see how to find the bounding curves [of these regions] [21].

So Cayley's regions are the modern basins of attraction for the point A, the bounding curves now known as Julia sets. He tried out the idea before the Cambridge Philosophical Society, and by the beginning of March had done enough to send the problem for publication:

> In connexion herewith, throwing aside the restrictions as to reality, we have what I call the Newton–

Fourier Imaginary Problem, as follows.

> Take $f(u)$ a given rational and integral function [a polynomial] of u, with real or imaginary coefficients; z, a given real or imaginary value, and from this derive z_1 by the formula

$$z_1 = z - \frac{f(z)}{f'(z)}$$

> and thence $z_1, z_2, z_3, \ldots$ each from the preceding one by the like formula. … The solution is easy and elegant in the case of a quadric equation: but the next succeeding case of the cubic equation appears to present considerable difficulty [17].

Cayley's connection with German mathematicians was close in the 1870s, and later on in 1879 (and in 1880) he went on tours of Germany, and visited mathematicians. No doubt he took the problem with him, and it was one he returned to periodically.

Both Cayley and Schröder solved this problem for the roots of $z^2 = 1$ but their methods differ. Cayley's is geometrical whereas Schröder's was analytical. There are only two fixed points $z = \pm 1$ and the boundary (Julia set) between the two basins of attraction of the root finding function is the "y-axis." The algorithm for finding the complex roots of the cubic equation $z^3 = 1$ has three stable fixed points at $z = 1, z = \exp(2\pi i/3), z = \exp(-2\pi i/3)$ and with the iteration $z \to z - (z^3 - 1)/3z^2$, three domains, basins of attraction exist with highly interlaced boundaries. The problem of determining the Julia set for the quadratic was straightforward but the cubic seemed impossible [42]. For good reason, with the aid of modern computing machinery the Julia set in the case of the cubic is an intricately laced trefoil. But some progress was made before computers entered the field. After WWI G. Julia and P. Fatou published lengthy papers on iteration and later C. L. Siegel studied the field [35,48,85].

Topological Dimension

The concept of dimension has been an enduring study for millennia, and fractals has prolonged the centrality of this concept in mathematics. The ordinary meaning of dimensions one, two, three dimensions applied to line, plane, solid of the Greeks, was initially extended to n-dimensions. Since the 1870s mathematicians ascribed newer meanings to the term.

At the beginning of the twentieth century, Topology was in its infancy, and emerging from "analysis situs." An impetus was the Hausdorff's definition of a topological space in terms of neighborhoods in 1914. The evolution of topological dimension was developed in the hands

of such mathematicians as L. E. J. Brouwer, H. Poincaré, K. Menger, P. Urysohn, F. Hausdorff but in the new surroundings, the concepts of curve and dimension proved elusive.

Poincaré made several attempts to define dimension. One was in terms of group theory, one he described as a "dynamical theory." This was unsatisfactory for why should dimension depend on the idea of a group. He gave another definition that he described as a "statical theory." Accordingly in papers of 1903, and 1912, a topological notion of n-dimensions (where n a natural number) was based on the notion of a cut, and Poincaré wrote:

> If to *divide* a continuum it suffices to consider as cuts a certain number of elements all distinguishable from one another, we say this continuum is of *one dimension*; if, on the contrary, to divide a continuum it is necessary to consider as cuts a system of elements themselves forming one or several continua, we shall say that this continuum is of *several dimensions*. If to divide a continuum C, cuts which form one or several continua of one dimension suffice, we shall say that C is a continuum of *two dimensions*; if cuts which form one or several continua of at most two dimensions suffice, we shall say that C is a continuum of *three dimensions*; and so on [74].

To illustrate the pitfalls of this game of cat and mouse where "definition" attempts to capture the right notion, we see this definition yielded some curious results – the dimension of a double cone ostensibly of two dimensions turns out to be of one dimension, since one can delete the zero dimensional point where the two ends of the cone meet.

Curves were equally difficult to pin down. Menger used a physical metaphor to get at the pure notion of a curve:

> We can think of a curve as being represented by fine wires, surfaces as produced from thin metal sheets, bodies as if they were made of wood. Then we see that in order to separate a point in the surface from points in a neighborhood or from other surfaces, we have to cut the surfaces along continuous lines with a scissors. In order to extract a point in a body from its neighborhood we have to saw our way through whole surfaces. On the other hand in order to excise a point in a curve from its neighborhood irrespective of how twisted or tangled the curve may be, it suffices to pinch at discrete points with tweezers. This fact, that is independent of the particular form of curves or surfaces we consider, equips us with a strong conceptual description [22].

Menger set out his ideas about basic notions in mathematical papers and in the books *Dimensionstheorie* (1928) and *Kurventheorie* (1932). In *Dimensionstheorie* he gave an inductive definition of dimension, on the implicit understanding that dimension only made sense for n an integer (≥ -1):

> A space is called at most n-dimensional, if every point is contained in arbitrarily small neighborhoods with an most $(n-1)$-dimensional boundaries. A space that is not at most $(n-1)$-dimensional we call at least n-dimensional. … A Space is called n-dimensional, if it is both at most n-dimensional and also at least n-dimensional, in other words, if every point is contained in arbitrarily small neighborhoods with at most $(n-1)$-dimensional boundaries, but at least one point is not contained in arbitrarily small boundaries with less than $(n-1)$-dimensional boundaries. … The empty set and only this is (-1)-dimensional (and at most (-1)-dimensional. A space that for no natural number n is n-dimensional we call infinite dimensional.

Different notions of topological dimension defined in the 1920s and 1930s, were ind X (the small inductive dimension), Ind X (the large inductive dimension), and dim X (the Lebesgue covering dimension). Researchers investigated the various inequalities between these and the properties of abstract topological spaces which ensured all these notions coincided. By 1950, the theory of topological dimension was still in its infancy [20,47,66].

Metric Dimension

For the purposes of fractals, it is the *metric* definitions of dimension which are fundamental.

For instance, how can we define the dimension of the Cantor's "middle third" set which takes into account its metric structure? What about the iconic fractal known as the von Koch curve snow flake curve (1904)? These sets pose interesting questions. Pictorially von Koch's curve is made up of small 1 dimensional lines and we might be persuaded that it too should be 1 dimensional. But the real von Koch curve is defined as a limiting curve and for this there are differences between it and a line. Between any two points on a 1 dimensional line there is a finite distance but between any two points on the Koch curve the distance is infinite. This suggests its dimensionality is greater than 1 while at the same time it does not fill out a 2 dimensional region. The Sierpinski curve (or gasket) is another example, but perhaps more spectacular is the "sponge curve". In

Kurventheorie (1932) Menger gave what is now known as the Menger Sponge, the set obtained from the unit cube by successively deleting sub-blocks in the same way as Cantor's middle third set is obtained from the unit interval.

A colorful approach to defining metric dimension is provided by L. F. Richardson. Richardson (1951) took on the practical measurement of coastlines from maps. To measure the coastline, Richardson used dividers set a distance l apart. The length of the coastline will be $L = \sum l$ after walking the dividers around the coast. Richardson was a practical man, and he found by plotting L against l, the purely empirical result that $L \propto l^{-\alpha}$ for a constant α that depends on the chosen coastline. So, for a given country we get an approximate length $L = cl^{-\alpha}$ but the smaller the dividers are set apart the longer becomes the coastline!

The phenomenon of the coastline length being "divider dependent" explains the discrepancy of 227 kilometers in the measurement by of the border between Spain and Portugal (987 km stated by Spain, and 1214 km stated by Portugal). The length of the Australian coastline turns out to be 9400 km if one divider space represents 2000 km, and is 14,420 km if the divider space represent 100 km. Richardson cautions "to speak simply of the "length" of a coast is therefore to make an unwarranted assumption. When a man says that he "walked 10 miles along the coast," he usually means that he walked 10 miles [on a smooth curve] *near* the coast." Richardson goes into incredible numerical data and his work is of an empirical nature, but it paid off, and he was able to combine two branches of science, the empirical and the mathematical. Moreover here we have a link with chaos: Richardson described the "sensitivity to initial conditions" where a small difference in the setting of the dividers can result in a large difference in the "length" of the coastline.

The constant α is a characteristic of the coastline or frontier whose value is dependent on its place in the range between smoothness and jaggedness. So if the frontier is a straight line, α would be zero, and increases the more irregular the coast line. In the case of Australia, α was found to be about 0.13 and for the very irregular west coast of Britain, α was found to be about 0.25. The value $1 + \alpha$ anticipates the mathematical concept known as "divider dimension". So, for example, the divider dimension of the west coast of Britain would be 1.25. The divider dimension idea is fruitful – it can be applied to Brownian motion, a mathematical theory set out by Norbert Wiener in the 1920s but it is not the main one when applied to fractal dimension.

"Fractal dimension" means Hausdorff dimension, or as we already noted, the Hausdorff–Besicovitch dimension. Generally Hausdorff dimension is difficult to cal-

culate, but for the self-similar sets and curves such as Cantor's middle-third set, the Sierpinski curve, the von Koch curve, Menger's sponge, it is equivalent to calculating the box dimension or Minkowski–Bouligand dimension (see also [61]). The middle third set has fractal dimension $D_H = \log 2/\log 3 = 0.63\dots$, and the Sierpinski curve has $D_H = \log 3/\log 2 = 1.58\dots$, the von Koch curve $D_H = \log 4/\log 3 = 1.26\dots$, and Menger's sponge embedded in three-dimensional Euclidean space has Hausdorff dimension $D_H = \log 20/\log 3 = 2.72\dots$.

Chaos Comes of Age

The modern theory of Chaos occurred around the end of the 1950s. S. Smale, Y. Ueda, E. N. Lorenz made discoveries which ushered in the new age. The subject became extremely popular in the early 1970s and the "avant-garde" looked back to this period as the beginning of chaos theory. They contributed some of the most often cited papers in further developments.

Physical Systems

In 1961 Y. Ueda, a third year undergraduate student in Japan discovered a curious phenomenon in connection with the single "Van der Pol" type nonlinear equation

$$d^2x/dt^2 + k(\gamma x^2 - 1)dx/dt + x^3 = \beta \cos \lambda t ,$$

another example of an equation used to model an oscillator. With the parameter values set at $k = 0.2, \gamma = 8, \beta = 0.35$ Ueda found the dynamics was "chaotic" – though of course he did not use that term. With an analogue computer, of a type then used to solve differential equations, the attractor in phase space appeared as a "shattered egg." Solutions for many other values of the parameters revealed orderly behavior, but what was special about the values 0.2, 8, 0.35? Ueda had no idea he had stumbled on a major discovery. Forty years later he reminisced on the higher principles of the scientific quest: "but while I was toiling alone in my laboratory," he said, "I was never trying to pursue such a grandiose dream as making a revolutionary new discovery, not did I ever anticipate writing a memoir about it. I was simply trying to find an answer to a persistent question, faithfully trying to follow the lead of my own perception of a problem [2]."

In 1963, E. N. Lorenz published a landmark paper (with at least 5000 citations to date) but it was one published in a journal off the beaten track for mathematicians. Lorenz investigated a simple model for atmospheric convection along the lines of the Rayleigh–Bernard convection model. To see what Lorenz actually did we quote the abstract to his original paper:

Finite systems of deterministic ordinary nonlinear differential equations may be designed to represent forced dissipative hydrodynamical flow. Solutions of these equations can be identified with trajectories in phase space. For those solutions with bounded solutions, it is found that non periodic solutions are ordinarily unstable with respect to small modifications, so that slightly differing initial states can evolve into considerably different states. Systems with bounded solutions are shown to possess bounded numerical solutions. A simple system representing cellular convection is solved numerically. All of the solutions are found to be unstable, and almost all of them are non periodic [53].

The oft quoted equations (with simplifying assumptions, and a truncated version of the Navier–Stokes equations) used to model Rayleigh–Bernard convection are the nonlinear equations [89]:

$$dx/dt = \sigma(y - x)$$
$$dy/dt = rx - y - xz$$
$$dz/dt = xy - bz \,,$$

where σ is called the Prandtl number, r is the Rayleigh number and b is a geometrically determined parameter. From the qualitative viewpoint, the solution of these equations, $(x_1(t), x_2(t), x_3(t))$ with initial values $(x_1(0), x_2(0), x_3(0))$ traces out a trajectory in 3-dimensional phase space. Lorenz solved the equations numerically by a forward difference procedure based on the time honored Runge–Kutta method which set up an iterative scheme of difference equations. Lorenz discovered that the case where the parameters have specific values of $\sigma = 10, r = 28, b = 8/3$ gave chaotic trajectories which wind around the famous Lorenz attractor in the phase space – and by accident, he discovered the butterfly effect. The significance of Lorenz's work was the discovery of Chaos in low dimensional systems – the equations described dynamical behavior of a kind seen by only a few – including Ueda in Japan. One can imagine his surprise and delight, for though working in Meteorology, Lorenz had been a graduate student of G. D. Birkhoff at Harvard. Once the chink in the mathematical fabric was made, the same kind of mathematical behavior of chaos was discovered in other systems of differential equations.

The Lorenz attractor became the subject of intensive research. Once it was discovered by mathematicians – not many mathematicians read the meteorology journals – it excited much attention and helped to launch the chaos craze. But one outstanding problem remained: did the Lorenz attractor actually exist or was its presence due to the accumulation of numerical errors in the approximate methods used to solve the differential equations. Computing power was in its infancy and Lorenz made his calculations on a fairly primitive string and sealing-wax Royal McBee LGP-30 machine.

Some believed that the numerical evidence was sufficient for the existence of a Lorenz attractor but this did not satisfy everyone. This question of actual existence resisted all attempts at its solution because there were no mathematical tools to solve the equations explicitly. The problem was eventually cracked by W. Tucker in 1999 then a postgraduate student at the University of Uppsala [94]. Tucker's proof that the Lorenz attractor actually exists involved a rigorous computer algorithm in conjunction with a rigorous set of bounds on the possible numerical errors which could occur. It is very technical [95].

Other sets of differential equations which exemplified the chaos phenomena, one even more basic than Lorenz's is due to O. Rössler, equations which modeled chemical reactions:

$$dx/dt = -y - z$$
$$dy/dt = x + \alpha y$$
$$dz/dt = \alpha - \mu z + xz \,.$$

Rössler discovered chaos for the parameter values $\alpha = 0.2$ and $\mu = 5.7$ [77]. This is the simplest system yet found, for it has only one nonlinear term xz compared with two in the case of Lorenz's equations. Other examples of chaotic solutions are found in the dripping faucet experiment caried out by Shaw (1984) [84].

Strange Attractors

An attractor is a set of points in phase space with the property that a trajectory with an initial point near it will be attracted to it – and if the trajectory touches the attractor it is trapped to stay within it. But what makes an attractor "strange"?

The simple pendulum swings to and fro. This is a dissipative system as the loss of energy causes the bob to come to rest. Whatever the initial point stating point of the bob, the trajectory in phase space will spiral down to the origin. The attractor in this case is simply the point at the origin; the rest point where displacement is nil and velocity is nil. This point is said to attract all solutions of the differential equations which models the pendulum. This single point attractor is hardly strange.

If the pendulum is configured to swing to and fro with a fixed amplitude, the attractor in phase space will be a circle. In this case the system conserves energy and the whole

system is called a conservative system. If the pendulum is slightly perturbed it will return to the previous amplitude, and in this sense the circle or a limit cycle will have attracted the displaced trajectory. Neither is the circle strange.

The case of the *driven* pendulum is different. In the driven pendulum the anchor point of the pendulum oscillates with a constant amplitude a and constant drive frequency f. The equation of motion of the angular displacement from the vertical θ with damping parameter q can be described as a second order nonlinear (because of the presence of the $\sin\theta$ term in the differential equation):

$$\mathrm{d}^2\theta/\mathrm{d}t^2 + (1/q)\,\mathrm{d}\theta/\mathrm{d}t + \sin\theta = a\cos ft\,.$$

Alternatively this motion can be described by three simultaneous equations:

$$\mathrm{d}w/\mathrm{d}t = -(1/q)w - \sin\theta + a\cos\phi$$
$$\mathrm{d}\theta/\mathrm{d}t = w$$
$$\mathrm{d}\phi/\mathrm{d}t = f\,,$$

where ϕ is the phase of the drive term. The three variables (w, θ, ϕ) describe the motion of the driven pendulum in three–dimensional phase space and its shape will depend on the values of the parameters (a, f, q). For some values, just as for the Lorenz attractor, the motion will be chaotic [5].

For an attractor to be strange, it should be fractal in structure. The notion of strange attractor is due to D. Ruelle and F. Takens in papers of 1971 [82]. In their work there is no mention of fractals, simply because fractals had not risen to prominence at that time. Strange attractors for Ruelle and Takens were infinite sets of points in phase space corresponding to points of a physical dynamic system which appeared to have a complicated evolution – they were just very weird sets of points. But the principle was gained. Dynamical systems, like the gusts in wind turbulence, are in principle modeled by deterministic differential equations but now their solution trajectories seemed random. In classical physics mechanical processes were supposed to be as uncomplicated as the pendulum where the attractor was a single point or a circular limit cycle. The seemingly random processes that now appeared offered the mathematician and physicist a considerable challenge.

The name "strange attractor" caught on and quickly captured the scientific and popular imagination. Ruelle asked Takens if he had dreamed up the name, and he replied: "Did you ever ask God whether he created this damned universe? … I don't remember anything … I often create without remembering it …" and Ruelle wrote

the "creation of strange attractors thus seems to be surrounded by clouds and thunder. Anyway, the name is beautiful, and I well suited to these astonishing objects, of which we understand so little [80]." Ruelle wrote "These systems of curves [arising in the study of turbulence], these clouds of points, sometimes evoke galaxies or fireworks, other times quite weird and disturbing blossomings. There is a whole world of forms still to be explored, and harmonies still to be discovered [45,91]."

Initially a strange attractor was a set which was extraordinary in some sense and there were naturally attempts to pin this down. In particular distinctions between "strange attractors" and "chaotic behavior" were drawn out [14]. Accordingly, a strange attractor is an attractor which is *not* (i) a finite set of points (ii) a closed curve (iii) a smooth or piecewise smooth surface or a volume bounded by such a surface. "Chaotic behavior" relates to the behavior of trajectories on the points of the attractor where nearby orbits around the attractor diverge with time. A strange nonchaotic attractor is one (with the above exclusions) where the trajectories are not chaotic, that is, there is an absence of sensitivity to initial conditions.

Pure Dynamical Systems

Chaos had its origins in real physical problems and the consequent differential equations, but the way had been set by Poincaré and Birkhoff for the entry of pure mathematicians. S. Smale gained his doctorate from the University of Michigan in 1956 supervised by the topologist R. Bott, stepped into this role. In seeking out fresh fields for research he surveyed Birkhoff's *Collected Papers* and read the papers of Levinson on the van der Pol equation..

On a visit to Rio de Janeiro in late 1959, Smale focused on general dynamical systems. What better place to do research than on the beach:

> My work was mostly scribbling down ideas and trying to see how arguments could be sequenced. Also I would sketch crude diagrams of geometric objects flowing through space and try to link the pictures with formal deductions. Deeply involved in this kind of thinking and writing on a pad of paper, the distraction of the beach did not bother me. Moreover, one could take time off from the research to swim [88].

He got into hotter water to the north when it was brought to the attention of government officials that national research grants were being spent at the seaside. They seemed unaware that a mathematical researcher is always working. Smale certainly had a capacity for hard concentrated work.

Receiving a letter from N. Levinson, about the Van der Pol equation, he reported:

> I worked day and night to try to resolve the challenge to my beliefs that letter posed. It was necessary to translate Levinson's analytic argument into my own geometric way of thinking. At least in my own case, understanding mathematics doesn't come from reading or even listening. It comes from rethinking what I see or hear. I must redo the mathematics in the context of my particular background. … In any case I eventually convinced myself that Levinson was correct, and that my conjecture was wrong. Chaos was already implicit in the analyzes of Cartwright and Littlewood! The paradox was resolved, I had guessed wrongly. But while learning that, I discovered the horseshoe [8].

The famous "horseshoe" mapping allowed him to construct a proof of the Poincaré conjecture in dimensions greater than or equal to five, and for this, he was awarded a Field's medal in 1966.

Iteration provides an example of a dynamical system, of a kind involving discrete time intervals. These were the systems suggested by Schröder and Cayley, and take on a pure mathematical guise with no physical experimenting to act as a guide. In the late 1950s, a research group in Toulouse led by I. Gumowski and his student C. Mira pursued an exploration of nonlinear systems from this point of view. Their approach was inductive, and started with the simplest example of nonlinear maps of the plane which were then iterated. Functions chosen for exploration were the family,

$$x \to y - F(x)$$
$$y \to x + F(y - F(x)),$$

for various rational functions $F(x)$. The pictures in the plane were noted for their spectacular "chaos esthétique."

The connection of chaos with the iteration of functions was pioneered by Gumowski and Mira, O. M. Sharkovsky (in 1964), Smale (in 1967) [86], and N. Metropolis, M. Stein, P. Stein (in 1973) [67]. M. Feigenbaum (in the late 1970s) studied the quadratic logistic map $x \to \lambda x(1 - x)$ and its iterates and discovered period doubling and the connection with physical phenomena. For the value $\lambda = 3.5699456\dots$ the orbit is non periodic and the attractor is the Cantor set so it is a strange attractor.

Feigenbaum studied one dimensional iterative maps of a general type $x \to \lambda f(x)$ for a general f and discovered properties independent of the form of the recursion function. The Feigenbaum number, the bifurcation rate δ with

value $\delta = 4.6692\dots$ is the limit of the parameter intervals in which period doubling takes place before the onset of chaos. The intervals are shortened at each period doubling by the inverse of the Feigenbaum number, a universal value for many functions [36,37]. In the case of fluid flow, it was discovered that the transition from smooth flow to turbulence is linked with period doubling.

The iteration of functions, a very simple process, was brought to the attention of the wider scientific public by R. May. In his oft quoted piece on the one dimension quadratic mapping, he stressed the importance it held for mathematical education:

> I would therefore urge that people be introduced to, say, equation $x \to \lambda x(1 - x)$ early in their mathematical education. This equation can be studied phenomenologically by iterating it on a calculator, or even by hand. Its study does not involve as much conceptual sophistication as does elementary calculus. Such study would greatly enrich the student's intuition about nonlinear systems.

> Not only in research, but also in the everyday world of politics and economics, we would all be better off if more people realized that simple nonlinear systems do not necessarily possess simple dynamical properties [62].

The example of $x \to \lambda x(1 - x)$ provides another link with the mathematical past, which amply demonstrates that Fractals and Chaos is not merely a child of the sixties but is joined to the mainstream of mathematics which slowly evolves over centuries. The Schwarzian derivative $y = f(x)$ defined by

$$S(y) = \frac{dx}{dy}\frac{d^3 y}{dx^3} - \frac{3}{2}\left(\frac{dx}{dy}\frac{d^2 y}{dx^2}\right)^2,$$

is connected with complex analysis and also introduced into invariant theory (where it was called a differentiant) of a hundred years previously. It was introduced by D. Singer in 1978 in the context of one dimensional dynamical systems [26].

While the quadratic map is one dimensional, we can go into two dimensions with pairs of nonlinear difference equations. A beautiful example is the one $(x, y) \to (y - x^2, a + bx)$ or in the equivalent form $(x, y) \to (1 + y - ax^2, bx)$. In this case, for some values of the parameters ellipses are generated, but in others we gain strange attractors. This is the Hénon scheme $b = 0.3$, and a near 1.4 the attractor has an infinity of attracting points, and being a fractal set it is a strange attractor [81]. A Poincaré

section of this attractor is a Cantor set, further evidence for the set being a strange attractor.

The Hénon map is one of the earliest examples illustrating chaos. It has Jacobian (which measures area) b for all points in the plane, so if $|b| < 1$ the Hénon mapping of the plane contracts area by a constant factor b for any point in the plane [41]. The effect of iterating this mapping is to stretch and fold its image in the manner of the Smale horseshoe mapping. The Hénon attractor is the prototype strange attractor. It arises in the case of a diffeomorphism of the plane which stretches and folds an open set U with the property that $f(\bar{U}) \subset U$.

The Advent of Fractals

The term was coined by B. B. Mandelbrot in the 1970s and now Mandelbrot is regarded as the "father of fractals" [55,56]. His first forays were to study fractals which were invariant under linear transformations. His uncle (Szolem Mandelbrojt, a professor of Mathematics and Mechanics at the Collège de France and later at the French Académie des Sciences) encouraged him to read the original papers by G. Julia and P. Fatou as a young man.

Euclid's *Elements*

Fractals are revolutionary because they challenge one of the sacred texts of mathematics. Euclid's *Elements* had reigned over mathematics for well over two thousand years and enjoys the distinction of still be referred to by working mathematicians. Up to the nineteenth century it was the essential fare of mathematics taught in schools. In Britain the *Elements* exerted the same authority as the Bible which was the other book committed to memory by generations of school pupils. Towards the end of the nineteenth century its central position in the curriculum was challenged, and it was much for the same reasons that Mandelbrot challenged it in the 1980s.

Praised for its logical structure, learning Euclid's deductive proofs by heart was simply not the way to teach an understanding of geometry. In France it had been displaced at the beginning of the century but in Britain and many other countries the attack on its centrality came at the end of the century. Ironically one of the main upholders of the sovereignty of Euclid and a staunch defender of the sacred book was Cayley. He was just the man who a few years before had shown himself to be more in the Mandelbrot mode, who threw aside "the restrictions as to reality" in his investigation of primitive symbolic dynamic systems. And there is a second irony, for Cayley was a man who loved nature and likened the beauty of the natural world to the terrain of mathematics itself – and in prac-

tical terms used his mountaineering exploits as a way of discovering new ideas in the subject.

A century later Mandelbrot's criticism came but from a different direction but containing enough phrases which would have gained a nod of agreement from those who mounted their opposition all those years before. Mandelbrot's opening paragraph launched the attack:

Why is geometry often described as "cold" and "dry"? One reason lies in its inability to describe the shape of a cloud, a mountain, a coastline, or a tree. Clouds are nor spheres, mountains are not cones, coastlines are not circles, and bark is not smooth, nor does lightning travel in a straight line [57].

Mandelbrot was setting out an ambitious claim, and behind it was the way he used for conducting mathematical research. He advocated a different methodology from the Euclidean style updated by the Bourbakian mantra of "definition, theorem, proof". In fact he describes himself as an exile, driven from France by the Bourbaki school. Jean Dieudonné, a leading Bourbakian was at the opposite end of the spectrum in his mathematical style. Dieudonné's objective, as set out in a graduate text *Foundations of Modern Analysis* (1960) was to "train the student in the use of the most fundamental mathematical tool of our time – the axiomatic method with which he [*sic*] will have had very little contact, if any at all, during his undergraduate years". Echoing the attitude of the nineteenth century geometer Jacob Steiner, who eschewed diagrams of any kind as inimical to the property development of the subject, Dieudonné appealed only to axiomatic methods, and to make the point in his book, he deliberately avoided introducing any diagrams and appeal to "geometric intuition"[28].

What Mandelbrot advocated was the methodology of the physician's and lawyer's "casebook." He noted "this term has no counterpart in science, and I suggest we appropriate it". This was rather novel and presents a counterpoint to the image that the Bourbakian ideal of great theorems with nicely turned out proofs are what is required above all else. If mathematics is presented as a completed house built by a great mathematician, as the Bourkadians could suggest, questions are only to be found in the higher reaches of the subjects, that demand years of study to reach them.

In his autobiography, Mandelbrot claimed "in pure mathematics, my main contribution has not been to provide proofs, but to ask new questions – usually very hard ones – suggested by physics and pictures". His questions spring from elementary objects in mathematics, but ones

which start off in the world. He summarizes his long career in his autobiography:

> My whole career became one long, ardent pursuit of the concept of roughness. The roughness of clusters in the physics of disorder, of turbulent flows, of exotic noises, of chaotic dynamical systems, of the distribution of galaxies, of coastlines, of stock-prize charts, and of mathematical constructions [58].

The "monster" examples which previously existed as instances of known theorems (and therefore were of little interest in themselves from the novelty point of view) or put aside because they were not examples, were now brought into the limelight and put on to the dissecting table to be investigated.

What is a Fractal?

In his essay (1975) Mandelbrot said "I stopped short of giving a mathematical definition, because I felt this notion – like a good wine – demanded a bit of ageing before being 'bottled'." [71]. He knew of Hausdorff dimension and developed an intuitive understanding for it, but postponed a definition. A little later he adopted a working definition, and in his *Fractal Geometry of Nature* (1982) he gave his manifesto for fractals.

What is Mandelbrot's subsequent working definition of a fractal? It is simply a point set for which the Hausdorff dimension exceeds its topological dimension, $D_H > D_T$. Examples are Cantor middle third set. where $D_H = \log 2/\log 3 = 0.63\ldots$ with $D_T = 0$, and the von Koch Curve where $D_H = \log 4/\log 3 = 1.26\ldots$ with $D_T = 1$.

Mandelbrot's definition gives rise to a programme of research similar in nature to one carried out by Menger and Urysohn in the 1920s. Just as they asked "what is a curve?" and "what is dimension?" it could now be asked "what is a fractal exactly?" and "what is fractal dimension?" The whole game of topological dimension was to be played over for metric dimension. Mandelbrot's definition of a fractal fired a first shot in answering one question but, just as there were exceptions to C. Jordan's definitions of curve like Peano's famous space filling curve of the 1890s, exceptions were found to Mandelbrot's preliminary definition. It was a regret, for example, that the example of the "devil's staircase" (characterized in terms of a continuous weight function defined on the Cantor middle third set) which looked like a fractal did not conform to the working definition of a fractal since in this case $D_H = D_T = 1$.

There are, however, many examples of obvious fractals, in nature and in mathematics. The pathological curves of the late nineteenth century which suffered this fate of not conforming to definitions of "normal curves" were brought back to life. The pathologies, such as those invented by Brouwer and Cantor had been put in a cupboard and locked away. An instance of this is Brouwer's decomposition of the plane into three "countries" so that the boundary points touch each other (not just meeting at one point). Brouwer's decomposition is not unlike the shape of a fractal. Cantor's examples were equally pathological.

Mandelbrot made his own discoveries, but when he opened the cupboard of "monster curves" he liked what he saw. Otto Rössler, with justification called Mandelbrot "Cantor's time-displaced younger friend". As Freeman Dyson wrote: "Now, as Mandelbrot points out, ..., Nature has played a joke on the mathematicians. The nineteenth century mathematicians may have been lacking in imagination [in limiting themselves to Euclid and Newton], but nature has not. The same pathological structures [the gallery of monster curves] that the mathematicians invented to break loose from nineteenth century naturalism turn out to be inherent in familiar objects all around us [38]."

The Mandelbrot Set

Just how did Mandelbrot arrive at the famous Mandelbrot set? With the freedom of being funded by an IBM fellowship, he went further into the study of $p_c(z) = z^2 + c$ and experimented. The basement of the Thomas J. Watson Research Center in the early 1980s held a brand new VAX main frame computer and a Tektronix cathode ray tube for viewing results.

The quadratic map $p_c(z) = z^2 + c$ is the simplest of all the nonlinear rational transformations. The key element is to consider z, c as complex numbers. The Mandelbrot set M in the plane is defined as

$$M = \{c \in C: \text{ the Julia set } J_c \text{ is connected}\}$$

or, alternatively, in terms of the sequence $\{p_c^k(0)\}$

$$M = \{c \in C: \{p_c^k(0)\} \text{ is bounded}\} .$$

It was an exciting event when Mandelbrot first glimpsed the set, which was initially called an M-set. First observed on the Tektronix cathode ray tube, was the messy outline a fault in the primitive equipment? He and his colleagues tried again with a more powerful computer but found "the mess had failed to vanish" and even that the image showed signs of being more systematic. Mandebrot reported "we promptly took a much closer look. Many specks of dirt duly vanished after we zoomed in. But some

specks failed to vanish; in fact, they proved to resolve into complex structures endowed with 'sprouts' very similar to those of the whole set M. Peter Moldave and I could not contain our excitement." Yet concrete mathematical properties were hard to come by. One immediate result did follow quickly, when the M-set, now termed the Mandelbrot set was proved to be a connected set [29].

The Merger

When chaos theory was being discussed in the 1970s, strange attractors were regarded as weird subsets of Euclidean space. To be sure their importance was recognized but their nature was only vaguely understood. Towards the end of the 1970s, prompted by the advent of fractals, attractors could be viewed in a new light. But Chaos which arose in physical situations did not give rise to ordinary self-similar fractals, like the Koch curve, but to more complicated sets. The outcome of Chaos is apparent randomness while the key properties of ordinary fractals is regularity.

Measuring Strange Attractors

Once strange attractors were brought to light, the next stage was to measure them. This perspective was put by practitioners in the 1980s:

> Nonlinear physics presents us with a perplexing variety of complicated fractal objects and strange sets. Notable examples include strange attractors for chaotic dynamical systems. ... Naturally one wishes to characterize the objects and described the events occurring on them. For example, in dynamical systems theory one is often interested in a strange attractor ... [43].

At the same time other researchers put a similar point of view, and an explanation for making the calculations. The way to characterize strange attractors was by the natural measure of a fractal – its fractal dimension – and this became the way to proceed:

> The determination of the fractal dimension d of strange attractors has become a standard diagnostic tool in the analysis of dynamical systems. The dimension roughly speaking measures the number of degrees of freedom that are relevant to the dynamics. Most work on dimension has concerned maps, such as the Hénon map, or systems given by a few coupled ordinary differential equations, such as the Lorenz and Rössler models. For any chaotic system described by differential equations, d must be

greater than 2, but d could be much larger for a system described by partial differential equations, such as the Navier–Stokes equations [12].

As an example, the fractal dimension of the original Lorenz attractor derived from the meteorology differential equations was found to be approximately 2.05 [34].

Newer Concepts of Dimension

Just as for metric free topological dimension itself, there are a myriad of different concepts of metrically based dimension. What was wanted were measures which could be used in practice. What "dimensions" would shine light on the onset of chaotic solutions to a set of differential equations ?

One was the Lyapunov dimension drawing on the work of the Russian mathematician Alexander Lyapunov, and investigated by J. L. Kaplan and J. A. Yorke (1979). The Lyapunov dimension of an attractor A embedded in an Euclidean space of dimension n is defined by:

$$D_\mathrm{L} = k + \frac{\lambda_1 + \lambda_2 + \cdots + \lambda_k}{|\lambda_{k+1}|} \, ,$$

where $\lambda_1, \lambda_2, \ldots, \lambda_k$ are Lyapunov exponents and k is the maximum integer for which $\sum_{i=1}^{k} \lambda_i \geq 0$. In the case where $k = 1$ which occurs for two-dimensional chaotic maps, for instance,

$$D_\mathrm{L} = 1 - \frac{\lambda_1}{\lambda_2}$$

and because in this case $\lambda_2 < 0$ the value of the Lyapunov dimension will be sandwiched between 1 and the topological dimension which is 2. In the case of the Hénon mapping of the plane, the Lyapunov dimension of the attractor is approximately 1.25.

In a different direction, a generalized dimension D_q was introduced by A. Rényi in the 1950s [78]. Working in probability and information theory, he defined a spectrum of dimensions:

$$D_q = \lim_{r \to 0} \frac{1}{q-1} \frac{\log \sum_i P_i^q}{\log r}$$

depending on the value of q.

The special cases are worth noting:

- $D_0 = D_\mathrm{B}$ the box dimension
- $D_1 =$ "information dimension" related to C. Shannon's entropy [39,40]
- $D_2 = D_\mathrm{C}$ the correlation dimension.

The correlation dimension is a very practical measure and can be applied to experimental data which has been presented visually, as well as other shapes, such as photographs where the calculation involves counting points. It has been used for applications in hydrodynamics, business, lasers, astronomy, signal analysis. It can be calculated directly for time series where the "phase space" points are derived from lags or time-delays. In two dimensions this would be a sequence of points $(x(n), x(n + l))$ where l is the chosen lag. A successful computation depends on the number of points chosen. For example, it is possible to estimate the fractal dimension of the Hénon attractor to within 6% of its supposed value with 500 data points [70].

There is a sequence of inequalities between the various dimensions of a set A from the embedding space of topological dimension $D_E = n$ to the topological dimension D_T of A [90]:

$$D_T \le \ldots \le D_2 \le D_1 \le D_H \le D_0 \le n \, .$$

Historically these questions of inequalities between the various types of dimension are of the same type as had been asked about topological dimension fifty years before. Just as it was shown that the "coincidence theorem"

$$\text{ind } X = \text{Ind } X = \dim X$$

holds for well behaved spaces, such as compact metric spaces, so it is true that

$$D_L = D_B = D_1$$

for ordinary point sets such as single points and limit cycles. For the Lorenz attractor D_L agrees with the value of D_B and that of D_2 [69]. But each of the equalities fails for some fractal sets; the study of fractal dimension of attractors is covered in [34,83].

Multifractals

An ordinary fractal is one where the generalized Rényi dimension D_q is independent of q and returns a single value for the whole fractal. This occurs for the standard fractals such as the von Koch curve. A multifractal is a set in which D_q is dependent on q and a continuous spectrum of values results.

While an ordinary self-similar fractal are useful for expository purposes (and are beautifully symmetric shapes) the attractors found in physics are not uniformly self-similar. The resulting object is called a multifractal because it is *multi*-dimensional. The values of D_q is called its spectrum of the multifractal. Practical examples usually require their strange attractors to be modeled by multifractals.

Future Directions

The range of written material on Fractals and Chaos has exploded since the 1970s. A vast number of expository articles have been lodged in such journals as *Nature* and *New Scientist* and technical articles in *Physica D* and *Nonlinearity* [9].

Complexity Theory and Chaos Theory are relatively new sciences that can revolutionize the way we see our world. Stephen Hawking has said, "Complexity will be the science of the 21st century." There is even a relationship between fractals with the yet unproved Riemann hypothesis [51].

Many problems remain. Here it is sufficient to mention one representative of the genre. Fluid dynamicists are broadly in agreement that fluids are accurately modeled by the nonlinear Navier–Stokes equations. These are based on Newtonian principles and are deterministic, but theoretically the solution of these equations is largely unknown territory. A proof of the global regularity of the solutions represents a formidable challenge. The current view is that there is a dichotomy between laminar flow (like flow in the upper atmosphere) being smooth while turbulent flow (like flow near the earth's surface) is violent. Yet even this "turbulent" flow could be regular but of a complicated kind. A substantial advance in the theory will be rewarded with one of the Clay Institute's million dollar prizes. One expert is not optimistic the prize will be claimed in the near future [52].

The implication of Chaos dependent as it is on the sensitivity of initial conditions, suggests that forecasting some physical processes is theoretically impossible. Long range weather forecasting falls into this mould since it is predicated on knowing the weather conditions *exactly* at some point in time. There will inevitably be inaccuracies so exactness appears to be an impossibility.

No doubt "Chaos and (multi)Fractals" is here to stay. Rössler wrote of Chaos as the key to understanding Nature: "hairs and noodles and spaghettis and dough and taffy form an irresistible, disentanglable mess. The world of causality is thereby caricatured and, paradoxically, faithfully represented [2]". Meanwhile, the challenge for scientists and mathematicians remains [16].

Bibliography

Primary Literature

1. Abraham RH (1985) In pursuit of Birkhoff's chaotic attractor. In: Pnevmatikos SN (ed) Singularities and Dynamical Systems. North Holland, Amsterdam, pp 303–312
2. Abraham RH, Ueda Y (eds) (2000) The Chaos Avant-Garde:

memories of the early days of chaos theory. World Scientific, River Edge

3. Alexander DS (1994) A History of Complex Dynamics; from Schröder to Fatou and Julia. Vieweg, Braunschweig

4. Aubin D, Dahan Dalmedico A (2002) Writing the History of Dynamical Systems and Chaos: *Longue Durée* and revolution, Disciplines and Cultures. Historia Mathematica 29(3):235–362

5. Baker GL, Gollub JP (1996) Chaotic Dynamics. Cambridge University Press, Cambridge

6. Barrow-Green J (1997) Poincaré and the three body problem. American Mathematical Society, Providence; London Mathematical Society, London

7. Barrow-Green J (2005) Henri Poincaré, Memoir on the Three-Body Problem (1890). In: Grattan-Guinness I (ed) Landmark Writings in Western Mathematics 1640–1940. Elsevier, Amsterdam, pp 627–638

8. Batterson S (2000) Stephen Smale: the mathematician who broke the dimension barrier. American Mathematical Society, Providence

9. Berry MV, Percival IC, Weiss NO (1987) Dynamical Chaos. Proc Royal Soc London 413(1844):1–199

10. Birkhoff GD (1927) Dynamical Systems. American Mathematical Society, New York

11. Birkhoff GD (1941) Some unsolved problems of theoretical dynamics. Science 94:598–600

12. Brandstater A, Swinney HL (1986) Strange attractors in weakly turbulent Couette-Tayler flow. In: Ott E et al (eds) Coping with Chaos. Wiley Interscience, New York, pp 142–155

13. Briggs J (1992) Fractals, the patterns of chaos: discovering a new aesthetic of art, science, and nature. Thames and Hudson, London

14. Brindley J, Kapitaniak T, El Naschie MS (1991) Analytical conditions for strange chaotic and nonchaotic attractors of the quasiperiodically forced van der Pol equation. Physica D 51:28–38

15. Brolin H (1965) Invariant Sets Under Iteration of Rational Functions. Ark Mat 6:103–144

16. Campbell DK, Ecke R, Hyman JM (eds) (1992) Nonlinear science: the next decade. MIT Press, Cambridge

17. Cayley A (1879) The Newton–Fourier Imaginary Problem. Am J Math 2:97

18. Chabert J-L (1990) Un demi-siècle de fractales: 1870–1920. Historia Mathematica 17:339–365

19. Crilly AJ, Earnshaw RA, Jones H (eds) (1993) Applications of fractals and chaos: the shape of things. Springer, Berlin

20. Crilly T (1999) The emergence of topological dimension theory. In: James IM (ed) History of Topology. Elsevier, Amsterdam, pp 1–24

21. Crilly T (2006) Arthur Cayley: Mathematician Laureate of the Victorian Age. Johns Hopkins University Press, Baltimore

22. Crilly T, Moran A (2002) Commentary on Menger's Work on Curve Theory and Topology. In: Schweizer, B. et al Karl Menger, Selecta. Springer, Vienna

23. Curry J, Garnett L, Sullivan D (1983) On the iteration of rational functions: Computer experiments with Newton's method. Commun Math Phys 91:267–277

24. Dahan-Dalmedico A, Gouzevitch I (2004) Early developments in nonlinear science in Soviet Russia: The Andronov School at Gor'kiy. Sci Context 17:235–265

25. Devaney RL (1984) Julia sets and bifurcation diagrams for exponential maps. Bull Am Math Soc 11:167–171

26. Devaney RL (2003) An Introduction to Chaotic Dynamical Systems, 2nd edn. Westview Press, Boulder

27. Diacu F, Holmes P (1996) Celestial Encounters: the Origins of Chaos and Stability. Princeton University Press, Princeton

28. Dieudonné J (1960) Foundations of Modern Analysis. Academic Press, New York

29. Douady A, Hubbard J (1982) Iteration des polynomes quadratiques complexes. Comptes Rendus Acad Sci Paris Sér 1 Math 294:123–126

30. Dyson F (1997) 'Nature's Numbers' by Ian Stewart. Math Intell 19(2):65–67

31. Edgar G (1993) Classics on Fractals. Addison-Wesley, Reading

32. Falconer K (1990) Fractal Geometry: Mathematical Foundations and Applications. Wiley, Chichester

33. Falconer K (1997) Techniques in Fractal Geometry. Wiley, Chichester

34. Farmer JD, Ott E, Yorke JA (1983) The Dimension of Chaotic Attractors. Physica D 7:153–180

35. Fatou P (1919/20) Sur les equations fonctionelles. Bull Soc Math France 47:161–271; 48:33–94, 208–314

36. Feigenbaum MJ (1978) Quantitative Universality for a Class of Nonlinear Transformations. J Stat Phys 19:25–52

37. Feigenbaum MJ (1979) The Universal Metric Properties of Nonlinear Transformations. J Stat Phys 21:669–706

38. Gleick J (1988) Chaos. Sphere Books, London

39. Grassberger P (1983) Generalised dimensions of strange attractors. Phys Lett A 97:227–230

40. Grassberger P, Procaccia I (1983) Measuring the strangeness of strange attractors. Physica D 9:189–208

41. Hénon M (1976) A Two dimensional Mapping with a Strange Attractor. Commun Math Phys 50:69–77

42. Haeseler F, Peitgen H-O, Saupe D (1984) Cayley's Problem and Julia Sets. Math Intell 6:11–20

43. Halsey TC, Jensen MH, Kadanoff LP, Procaccia Shraiman I (1986) Fractal measures and their singularities: the characterization of strange sets. Phys Rev A 33:1141–1151

44. Harman PM (1998) The Natural Philosophy of James Clerk Maxwell. Cambridge University Press, Cambridge

45. Hofstadter DR (1985) Metamathematical Themas: questing for the essence of mind and pattern. Penguin, London

46. Holmes P (2005) Ninety plus years of nonlinear dynamics: More is different and less is more. Int J Bifurc Chaos Appl Sci Eng 15(9):2703–2716

47. Hurewicz W, Wallman H (1941) Dimension Theory. Princeton University Press, Princeton

48. Julia G (1918) Sur l'iteration des functions rationnelles. J Math Pures Appl 8:47–245

49. Kazim Z (2002) The Hausdorff and box dimension of fractals with disjoint projections in two dimensions. Glasgow Math J 44:117–123

50. Lapidus ML, van Frankenhuijsen M (2004) Fractal geometry and Applications: A Jubilee of Benoît Mandelbrot, Parts, 1, 2. American Mathematical Society, Providence

51. Lapidus ML, van Frankenhuysen (2000) Fractal geometry and Number Theory: Complex dimensions of fractal strings and zeros of zeta functions. Birkhäuser, Boston

52. Li YC (2007) On the True Nature of Turbulence. Math Intell 29(1):45–48

53. Lorenz EN (1963) Deterministic Nonperiodic Flow. J Atmospheric Sci 20:130–141

54. Lorenz EN (1993) The Essence of Chaos. University of Washington Press, Seattle

55. Mandelbrot BB (1975) Les objets fractals, forme, hazard et dimension. Flammarion, Paris

56. Mandelbrot BB (1980) Fractal aspects of the iteration of $z \rightarrow \lambda(1-z)$ for complex λ, z. Ann New York Acad Sci 357:249–259

57. Mandelbrot BB (1982) The Fractal Geometry of Nature. Freeman, San Francisco

58. Mandelbrot BB (2002) A maverick's apprenticeship. Imperial College Press, London

59. Mandelbrot BB, Hudson RL (2004) The (Mis)Behavior of Markets: A Fractal View of Risk, Ruin, and Reward. Basic Books, New York

60. Mattila P (1995) Geometry of Sets and Measures in Euclidean Spaces. Cambridge University Press, Cambridge

61. Mauldin RD, Williams SC (1986) On the Hausdorff dimension of some graphs. Trans Am Math Soc 298:793–803

62. May RM (1976) Simple Mathematical models with very complicated dynamics. Nature 261:459–467

63. May RM (1987) Chaos and the dynamics of biological populations. Proc Royal Soc Ser A 413(1844):27–44

64. May RM (2001) Stability and Complexity in Model Ecosystems, 2nd edn, with new introduction. Princeton University Press, Princeton

65. McMurran S, Tattersall J (1999) Mary Cartwright 1900–1998. Notices Am Math Soc 46(2):214–220

66. Menger K (1943) What is dimension? Am Math Mon 50:2–7

67. Metropolis, Stein ML, Stein P (1973) On finite limit sets for transformations on the unit interval. J Comb Theory 15:25–44

68. Morse M (1946) George David Birkhoff and his mathematical work. Bull Am Math Soc 52(5, Part 1):357–391

69. Nese JM, Dutton JA, Wells R (1987) Calculated Attractor Dimensions for low-Order Spectral Models. J Atmospheric Sci 44(15):1950–1972

70. Ott E, Sauer T, Yorke JA (1994) Coping with Chaos. Wiley Interscience, New York

71. Peitgen H-O, Richter PH (1986) The Beauty of Fractals. Springer, Berlin

72. Peitgen H-O, Jürgens H, Saupe D (1992) Chaos and fractals. Springer, New York

73. Pesin YB (1997) Dimension Theory in Dynamical Systems: contemporary views and applications. University of Chicago Press, Chicago

74. Poincaré H (1903) L'Espace et ses trois dimensions. Rev Métaphys Morale 11:281–301

75. Poincaré H, Halsted GB (tr) (1946) Science and Method. In: Cattell JM (ed) Foundations of Science. The Science Press, Lancaster

76. Przytycki F, Urbański M (1989) On Hausdorff dimension of some fractal sets. Studia Mathematica 93:155–186

77. Rössler OE (1976) An Equation for Continuous Chaos. Phys Lett A 57:397–398

78. Rényi A (1970) Probability Theory. North Holland, Amsterdam

79. Richardson LF (1993) Collected Papers, 2 vols. Cambridge University Press, Cambridge

80. Ruelle D (1980) Strange Attractors. Math Intell 2:126–137

81. Ruelle D (2006) What is a Strange Attractor? Notices Am Math Soc 53(7):764–765

82. Ruelle D, Takens F (1971) On the nature of turbulence. Commun Math Phys 20:167–192; 23:343–344

83. Russell DA, Hanson JD, Ott E (1980) Dimension of strange attractors. Phys Rev Lett 45:1175–1178

84. Shaw R (1984) The Dripping Faucet as a Model Chaotic System. Aerial Press, Santa Cruz

85. Siegel CL (1942) Iteration of Analytic Functions. Ann Math 43:607–612

86. Smale S (1967) Differentiable Dynamical Systems. Bull Am Math Soc 73:747–817

87. Smale S (1980) The Mathematics of Time: essays on dynamical systems, economic processes, and related topics. Springer, New York

88. Smale S (1998) Chaos: Finding a horseshoe on the Beaches of Rio. Math Intell 20:39–44

89. Sparrow C (1982) The Lorenz Equations: Bifurcations, Chaos, and Strange Attractors. Springer, New York

90. Sprott JC (2003) Chaos and Time-Series Analysis. Oxford University Press, Oxford

91. Takens F (1980) Detecting Strange Attractors in Turbulence. In: Rand DA, Young L-S (eds) Dynamical Systems and Turbulence. Springer Lecture Notes in Mathematics, vol 898. Springer, New York, pp 366–381

92. Theiler J (1990) Estimating fractal dimensions. J Opt Soc Am A 7(6):1055–1073

93. Tricot C (1982) Two definitions of fractional dimension. Math Proc Cambridge Philos Soc 91:57–74

94. Tucker W (1999) The Lorenz Attractor exists. Comptes Rendus Acad Sci Paris Sér 1, Mathematique 328:1197–1202

95. Viana M (2000) What's New on Lorenz Strange Attractors. Math Intell 22(3):6–19

96. Winfree AT (1983) Sudden Cardiac death- a problem for topology. Sci Am 248:118–131

97. Zhang S-Y (compiler) (1991) Bibliography on Chaos. World Scientific, Singapore

Books and Reviews

Abraham RH, Shaw CD (1992) Dynamics, The Geometry of Behavior. Addison-Wesley, Redwood City

Barnsley MF, Devaney R, Mandelbrot BB, Peitgen H-O, Saupe D, Voss R (1988) The Science of Fractal Images. Springer, Berlin

Barnsley MF, Rising H (1993) Fractals Everywhere. Academic Press, Boston

Çambel AB (1993) Applied Complexity Theory: A Paradigm for Complexity. Academic Press, San Diego

Crilly AJ, Earnshaw RA, Jones H (eds) (1991) Fractals and Chaos. Springer, Berlin

Cvitanovic P (1989) Universality in Chaos, 2nd edn. Adam Hilger, Bristol

Elliott EW, Kiel LD (eds) (1997) Chaos Theory in the Social Sciences: foundations and applications. University of Michigan Press, Ann Arbor

Gilmore R, Lefranc M (2002) The Topology of Chaos: Alice in Stretch and Squeezeland. Wiley Interscience, New York

Glass L, MacKey MM (1988) From Clocks to Chaos. Princeton University Press, Princeton

Holden A (ed) (1986) Chaos. Manchester University Press, Manchester

Kellert SH (1993) In the Wake of Chaos. University of Chicago Press, Chicago

Lauwerier H (1991) Fractals: Endlessly Repeated Geometrical Figures. Princeton University Press, Princeton

Mullin T (ed) (1994) The Nature of Chaos. Oxford University Press, Oxford

Ott E (2002) Chaos in Dynamical Systems. Cambridge University Press, Cambridge

Parker B (1996) Chaos in the Cosmos: The Stunning Complexity of the Universe. Plenum, New York, London

Peitgen H-O, Jürgens, Saupe D, Zahlten C (1990) Fractals: An animated discussion with Edward Lorenz and Benoit Mandelbrot. A VHS film in color (63 mins). Freeman, New York

Prigogine I, Stengers I (1985) Order out of Chaos: man's new dialogue with Nature. Fontana, London

Ruelle D (1993) Chance and Chaos. Penguin, London

Schroeder M (1991) Fractals, Chaos, Power Laws. Freeman, New York

Smith P (1998) Explaining Chaos. Cambridge University Press, Cambridge

Thompson JMT, Stewart HB (1988) Nonlinear Dynamics and Chaos. Wiley, New York

Fractals and Multifractals, Introduction to

DANIEL BEN-AVRAHAM[1], SHLOMO HAVLIN[2]
[1] Clarkson University, Potsdam, USA
[2] Bar-Ilan University, Ramat Gan, Israel

Fractals generalize Euclidean geometrical objects to non-integer dimensions and allow us, for the first time, to delve into the study of complex systems, disorder, and chaos. In the words of B. Mandelbrot: "Clouds are not spheres, mountains are not cones, coastlines are not circles, bark is not smooth, nor does lightning travel in a straight line," [1]. Indeed, much has changed in our perception of nature, and today it is hard to conceive of natural phenomena that are *not* fractal, in the same way that it is hard to conceive of everyday life dynamical systems that are not non-linear. The discovery of fractals over three decades ago signaled a profound shift in the way we understand and analyze the physical world around us.

Not only does fractal geometry model complex disordered objects such as clouds and mountains, coastlines and lightning, but it also finds a beautiful new symmetry in the midst of all this complexity – invariance under dilation of space – and it is this self-similarity symmetry that lends fractals their tremendous power and analytical appeal (▶ Fractals and Multifractals, Introduction to). The same symmetry plays a significant role in critical phase transitions, and one of the earliest applications of fractals has been to the study of the paradigmatic model of percolation (▶ Fractals and Percolation). Soon after their advent, researchers systematically used fractals to explain a host of anomalous phenomena (anomalous with respect to the expectations from regular Euclidean objects) in condensed matter and solid state physics (▶ Fractal Structures in Condensed Matter Physics).

One of the most influential models of nonequilibrium growth, Diffusion Limited Aggregation (DLA), produces a fractal object that is still studied today for its fascinating connections to the Laplace equation and electrical discharge, conformal mapping, patterns of bacterial growth, and viscous fingering (▶ Fractal Growth Processes). Fractals and fractal scaling also arise in regular systems with superposed disorder, for example when resistances from a random distribution are assigned to the bonds of a regular lattice (▶ Fractal and Multifractal Scaling of Electrical Conduction in Random Resistor Networks). More recently, interest has grown in the study of large, stochastic complex nets, which seem infinite-dimensional at the outset, but even so can be found to have fractal geometry (▶ Fractal and Transfractal Scale-Free Networks). Having realized that the geometry of everyday life objects around us is more likely fractal than not, it was only natural to ask about the physics of phase transitions in disordered media, for example, for the magnetization of spins in the Ising model, placed on the nodes of fractals (▶ Phase Transitions on Fractals and Networks).

Early on, interest developed in the physics of transport in fractal substrates. It was quickly discovered that fractal media diffusion, the most elementary mode of transport, does not obey Fick's law but is, rather, anomalous (▶ Anomalous Diffusion on Fractal Networks), leading to entirely new ways of viewing and analyzing transport, most notably among them Lévy flights, continuum random walks, and fractional diffusion equations (▶ Levy Statistics and Anomalous Transport: Levy Flights and Subdiffusion). Because of the anomalous diffusion of reactants in fractal disordered media, so are the kinetics of reactions among them anomalous (▶ Reaction Kinetics in Fractals). The graph Laplacian in fractals has an anomalous spectrum that displays a peculiar scaling, having a profound effect on dynamics in general (▶ Dynamics on Fractals).

Fractals refer not only to geometrical objects but, more broadly, to any kind of phenomena possessing scaling that exhibits dilation symmetry, or scale invariance, often characterized by the appearance of a power-law distribution. A whole suite of techniques has evolved to analyze this type of fractal scaling (▶ Fractal and Multifractal Time Series). Fractal phenomena of this type (and occasionally fractal objects) find applications in several diverse fields of interest, such as finances and economics (▶ Fractals and Economics), geology (▶ Fractals in Geology and Geophysics), the analysis of DNA sequences (▶ Fractals and Wavelets: What Can We Learn on Transcription and

Replication from Wavelet-Based Multifractal Analysis of DNA Sequences?), biology (► Fractals in Biology), and even in path integrals in the quantum theory of spacetime (► Fractals in the Quantum Theory of Spacetime).

Fractals and the modern study of nonlinear systems, strange attractors, and chaos have infused and enriched one another from their very incipience (► Fractals Meet Chaos). Together they provide us with a fundamentally new way of understanding the everyday world around us.

Despite our best intentions, this Section on Fractals and Multifractals remains incomplete, as it fails to include numerous important subjects such as the practical use of fractals for image compression, their ubiquity in medicine, astronomy, and their prominence in the study of real systems of porous materials, electrodeposition, molecular surfaces, colloids, and polymers, to name just a few. We hope that some, if not all of these subjects will be included in future editions.

Bibliography

1. Mandelbrot BB (1982) The Fractal Geometry of Nature. Freeman, San Francisco

Fractals and Percolation

YAKOV M. STRELNIKER[1], SHLOMO HAVLIN[1], ARMIN BUNDE[2]
[1] Department of Physics, Bar–Ilan University, Ramat–Gan, Israel
[2] Institut für Theoretische Physik III, Justus-Liebig-Universität, Giessen, Germany

Article Outline

Glossary

Percolation In the traditional meaning, percolation concerns the movement and filtering of fluids through porous materials. In this chapter, percolation is the subject of physical and mathematical models of porous media that describe the formation of a long-range connectivity in random systems and phase transitions. The most common percolation model is a lattice, where each site is occupied randomly with a probability p or empty with probability $1 - p$. At low p values, there is no connectivity between the edges of the lattice. Above some concentration p_c, the *percolation threshold*, connectivity appears between the edges. Percolation represents a geometric critical phenomena where p is the analogue of temperature in thermal phase transitions.

Fractal A fractal is a structure which can be subdivided into parts, where the shape of each part is similar to that of the original structure. This property of fractals is called self-similarity, and it was first recognized by G.C. Lichtenberg more than 200 years ago. Random fractals represent models for a large variety of structures in nature, among them porous media, colloids, aggregates, flashes, etc. The concepts of self-similarity and fractal dimensions are used to characterize percolation clusters. Self-similarity is strongly related to renormalization properties used in critical phenomena, in general, and in percolation phase transition properties.

Definition of the Subject

Percolation theory is useful for characterizing many disordered systems. Percolation is a pure random process of choosing sites to be randomly occupied or empty with certain probabilities. However, the topology obtained in such processes has a rich structure related to fractals. The structural properties of percolation clusters have become clearer thanks to the development of fractal geometry since the 1980s.

Introduction

Percolation represents the simplest model of a phase transition [1,8,13,14,26,27,30,48,49,61,64,65,68]. Assume a regular lattice (grid) where each site (or bond) is occupied with probability p or empty with probability $1 - p$. At a critical threshold, p_c, a long-range connectivity first appears: p_c is called the percolation threshold (see Fig. 1). Occupied and empty sites (or bonds) may stand for very different physical properties. For example, occupied sites may represent electrical conductors, empty sites may represent insulators, and electrical current may flow only through nearest-neighbor conducting sites. Below p_c, the grid represents an isolator since there is no conducting path between two adjacent bars of the lattice, while above p_c, conducting paths start to occur and the grid becomes a conductor. One can also consider percolation as a model

F

a $p = 0.2$ b $p = 0.59$ c $p = 0.8$
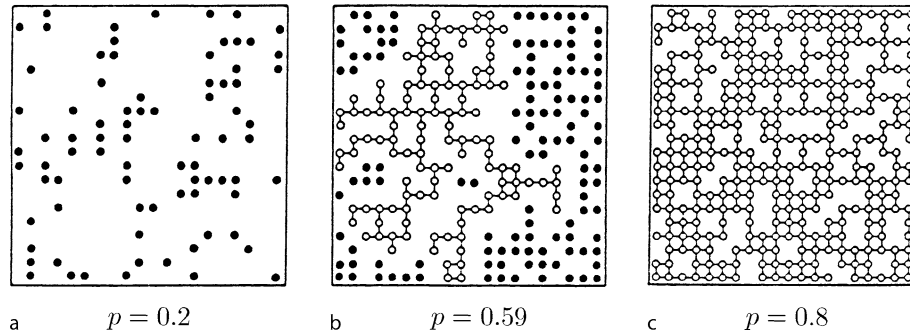
Fractals and Percolation, Figure 1
Square lattice of size 20 × 20. Sites have been randomly occupied with probability *p* (*p* = 0.20, 0.59, 0.80). Sites belonging to finite clusters are marked by *full circles*, while sites on the infinite cluster are marked by *open circles*

Fractals and Percolation, Figure 2
Invasion percolation through porous media

for liquid filtration (i. e., invasion *percolation* (see Fig. 2), which is the source of this terminology) through porous media.

A possible application of bond percolation in chemistry is the polymerization process [25,31,44], where small branching molecules can form large molecules by activating more and more bonds between them. If the activation probability *p* is above the critical concentration, a network of chemical bonds spanning the whole system can be formed, while below p_c only macromolecules of finite size can be generated. This process is called a *sol-gel* transition. An example of this *gelation* process is the boiling of an egg, which at room temperature is liquid but, upon heating, becomes a solid-like *gel*.

An example from biology concerns the spreading of an epidemic [35]. In its simplest form, an epidemic starts with one sick individual which can infect its nearest neighbors with probability *p* in one time step. After one time step, it dies, and the infected neighbors in turn can infect their (so far) uninfected neighbors, and the process is continued. Here the critical concentration separates a phase at low *p* where the epidemic always dies out after a finite number

of time steps, from a phase where the epidemic can continue forever. The same process can be used as a model for forest fires [52,60,64,71], with the infection probability replaced by the probability that a burning tree can ignite its nearest-neighbor trees in the next time step. In addition to these simple examples, percolation concepts have been found useful for describing a large number of disordered systems in physics and chemistry.

The first study introducing the concept of percolation is attributable to Flory and Stockmayer about 65 years ago, when studying the gelation process [32]. The name percolation was proposed by Broadbent and Hammersley in 1957 when they were studying the spreading of fluids in random media [15]. They also introduced relevant geometrical and probabilistic concepts. The developments of phase transition theory in the following years, in particular the series expansion method by Domb [27] and renormalization group theory by Wilson, Fisher and Kadanoff [51,65], very much stimulated research activities into the geometric percolation transition.

At the percolation threshold, the conducting (as well as insulating) clusters are self-similar (see Fig. 3) and, therefore, can be described by fractal geometry [53], where various fractal dimensions are introduced to quantify the clusters and their physical properties.

Percolation

As above (see Sect. "Introduction"), consider a square lattice, where each site is occupied randomly with probability *p* (see Fig. 1). For simplicity, let us assume that the occupied sites are electrical conductors and the empty sites represent insulators. At low concentration *p*, the occupied sites either are isolated or form small clusters (Fig. 1a). Two occupied sites belong to the same cluster if they are connected by a path of nearest-neighbor occupied sites
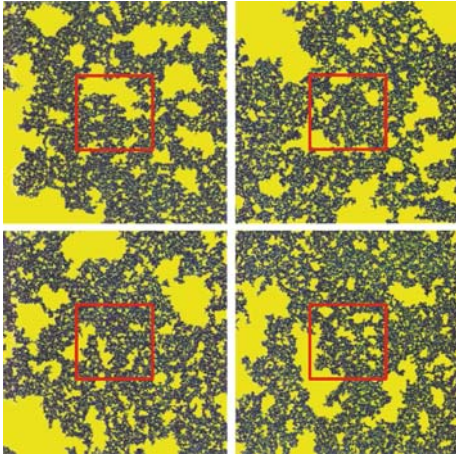
Fractals and Percolation, Figure 3
Self-similarity of the *random* percolation cluster at the critical concentration; courtesy of M. Meyer

and a current can flow between them. When p is increased, the average size of the clusters increases. At a critical concentration p_c (also called the *percolation threshold*), a large cluster appears which connects opposite edges of the lattice (Fig. 1). This cluster is called the *infinite* cluster, since its size diverges when the size of the lattice is increased to infinity. When p is increased further, the density of the infinite cluster increases, since more and more sites become part of the infinite cluster, and the average size of the *finite* clusters decreases (Fig. 1c).

The *percolation threshold* separates two different phases and, therefore, the *percolation transition* is a *geometrical phase transition*, which is characterized by the geometric features of large clusters in the neighborhood of p_c. At low values of p, only small clusters of occupied sites exist. When the concentration p is increased, the average size of the clusters increases. At the critical concentration p_c, a large cluster appears which connects opposite edges of the lattice. Accordingly, the average size of the *finite* clusters which do not belong to the infinite cluster decreases. At $p = 1$, trivially, all sites belong to the infinite cluster.

Similar to *site percolation*, it is possible to consider *bond percolation* when the bonds between sites are randomly occupied. An example of bond percolation in physics is a *random resistor network*, where the metallic wires in a regular network are cut at random. If sites are occupied with probability p and bonds are occupied with probability q, we speak of *site-bond percolation*. Two occupied sites belong to the same cluster if they are connected by a path of nearest-neighbor occupied sites with occupied bonds in between.

The definitions of site and bond percolation on a square lattice can easily be generalized to any lattice in d-dimensions. In general, in a given lattice, a bond has more nearest neighbors than a site. Thus, large clusters of bonds can be formed more effectively than large clusters of sites, and therefore, on a given lattice, the percolation threshold for bonds is smaller than the percolation threshold for sites (see Table 1).

A natural example of percolation, is *continuum percolation*, where the positions of the two components of a random mixture are not restricted to the discrete sites of a regular lattice [9,73]. As a simple example, consider a sheet of conductive material, with circular holes punched randomly in it (Swiss cheese model, see Fig. 4). The relevant quantity now is the fraction p of remaining conductive material.

Fractals and Percolation, Table 1
Percolation thresholds for the Cayley tree and several two- and three-dimensional lattices (see Refs. [8,14,41,64,75] and references therein)

Lattice	Percolation of	
	Sites	Bonds
Triangular	1/2	$2\sin(\pi/18)$
Square	0.5927460	1/2
Honeycomb	0.6962	$1 - 2\sin(\pi/18)$
Face Centered Cubic	0.198	0.119
Body Centered Cubic	0.245	0.1803
Simple Cubic (1^{st}nn)	0.31161	0.248814
Simple Cubic (2^{nd}nn)	0.137	–
Simple Cubic (3^{rd}nn)	0.097	–
Cayley Tree	$1/(z-1)$	$1/(z-1)$
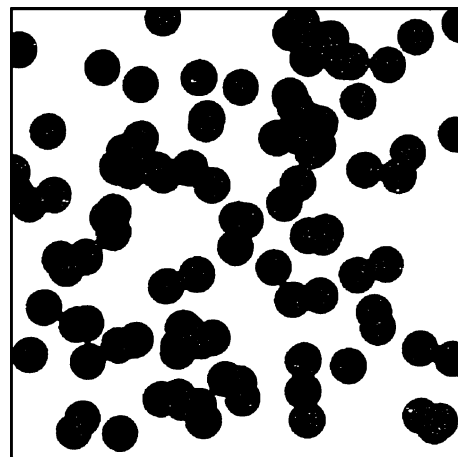
Fractals and Percolation, Figure 4
Continuum percolation: Swiss cheese model

Hopping Percolation

Above, we have discussed traditional percolation with only two values of local conductivities, 0 and 1 (insulator–conductor) or ∞ and 1 (superconductor–normal conductor). However, quantum systems should be treated by hopping conductivity, which can be described by an exponential function representing the local conductivities (between ith and jth sites): $\sigma_{ij} \sim \exp\left(-\kappa x_{ij}\right)$. Here κ can be interpreted as the dimensionless mean hopping distance or as the degree of disorder (the smaller the density of the deposited grains, the larger κ becomes), and x_{ij} is a random number taken from a uniform distribution in the range (0,1) [70].

In contrast to the traditional bond (or site) percolation model, in which the system is either a metal or an insulator, in the hopping percolation model the system always conducts some current. However, there are two regimes of such percolation [70]. A regime with many conducting paths which is not sensitive to the removal of a single bond (*weak disorder $L/\kappa^\nu > 1$*, where L is size of the system and ν is percolation critical exponent) and a regime with a single or only a few dominating conducting paths which is very sensitive to the removal of a specific single bond with the highest current (*strong disorder $L/\kappa^\nu \ll 1$*). In the strong disorder regime, the trajectories along which the highest current flows (analogous to the spanning cluster at criticality in the traditional percolation network, see Fig. 5) can be distinguished and a single bond can determine the transport properties of the entire macroscopic system.

Fractals and Percolation, Figure 5
A color density plot of the current distribution in a bond-percolating lattice for which voltage is applied in the vertical direction for strong disorder with $\kappa = 10$. The current between the sites is shown by the different colors (*orange* corresponds to the highest value, *green* to the lowest). **a** The location of the resistor, on which the value of the local current is maximal, is shown by a *circle*. **b** The current distribution after removing the above resistor. This removal results in a significant change of the current trajectories

Percolation as a Critical Phenomenon

In percolation, the concentration p of occupied sites plays the same role as temperature in thermal phase transitions. The percolation transition is a geometrical phase transition where the critical concentration p_c separates a phase of finite clusters ($p < p_c$) from a phase where an infinite cluster is present ($p > p_c$).

An important quantity is the probability P_∞ that a site (or a bond) belongs to the infinite cluster. For $p < p_c$, only finite clusters exist, and $P_\infty = 0$. For $p > p_c$, P_∞ increases with p by a power law

$$P_\infty \sim (p - p_c)^\beta \ . \tag{1}$$

P_∞ can be identified as the *order parameter* similar to magnetization, $m(T) \sim (T_c - T)^\beta$, in magnetic materials. With decreasing temperature, T, more elementary magnetic moments (spins) become aligned in the same direction, and the system becomes more ordered.

The linear size of the *finite* clusters, below and above p_c, is characterized by *correlation length* ξ. Correlation length is defined as the mean distance between two sites on the same finite cluster. When p approaches p_c, ξ increases as

$$\xi \sim |p - p_c|^{-\nu} \ , \tag{2}$$

with the same exponent ν below and above the threshold. The mean number of sites (mass) of a finite cluster also diverges,

$$S \sim |p - p_c|^{-\gamma} \ , \tag{3}$$

again with the same exponent γ above and below p_c. Analogous to S in magnetic systems is the susceptibility χ (see Fig. 6 and Table 2).

The exponents β, ν, and γ describe the critical behavior of typical quantities associated with the percolation transition, and are called *critical exponents*. The exponents
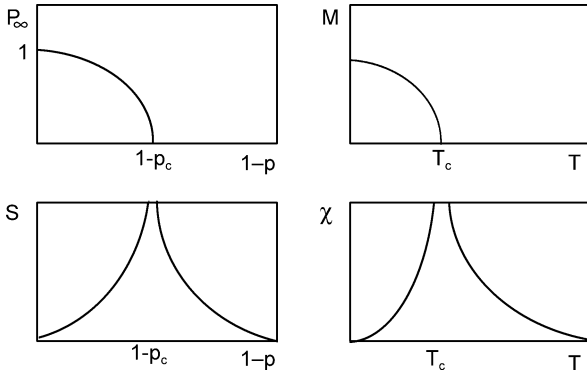
Fractals and Percolation, Figure 6
P_∞ and S compared with magnetization M and susceptibility χ

Fractals and Percolation, Table 2
Exact and best estimate values for the critical exponents for percolation (see Refs. [8,14,41,64] and references therein)

Percolation	$d = 2$	$d = 3$	$d \geq 6$
Order parameter P_∞: β	5/36	0.417 ± 0.003	1
Correlation length ξ: ν	4/3	0.875 ± 0.008	1/2
Mean cluster size S: γ	43/18	1.795 ± 0.005	1

are universal and do not depend on the structural details of the lattice (e. g., square or triangular) nor on the type of percolation (site, bond, or continuum), but depend only on the dimension d of the lattice.

This universality property is a general feature of phase transitions, where the order parameter vanishes continuously at the critical point (second order phase transition).

In Table 2, the values of the critical exponents β, ν, and γ for percolation in two, three, and six dimensions. The exponents considered here describe the geometrical properties of the percolation transition. The physical properties associated with this transition also show power-law behavior near p_c and are characterized by critical exponents. Examples include the conductivity in a random resistor or random superconducting network and the spreading velocity of an epidemic disease near the critical infection probability. It is believed that the "dynamical" exponents cannot be generally related to the geometric exponents discussed above.

Note that all quantities described above are defined in the thermodynamic limit of large systems. In a finite system, for example, P_∞, is not strictly zero below p_c.

Percolation Clusters as Fractals

As first noticed by Stanley [66], the structure of percolation clusters (when the length scale is smaller than ξ) can

be well described by the fractal concept [53]. Fractal geometry is a mathematical tool for dealing with complex structures that have no characteristic length scale. Scale-invariant systems are usually characterized by noninteger ("fractal") dimensions. This terminology is associated with B. Mandelbrot [53] (though some notion of noninteger dimensions and several basic properties of fractal objects were studied earlier by G. Cantor, G. Peano, D. Hilbert, H. von Koch, W. Sierpinski, G. Julia, F. Hausdorff, C. F. Gauss, and A. Dürer).

Fractal Dimension d_f

In regular systems (with uniform density) such as long wires, large thin plates, or large filled cubes, the dimension d characterizes how the mass $M(L)$ changes with the linear size L of the system. If we consider a smaller part of a system of linear size bL ($b < 1$), then $M(bL)$ is decreased by a factor of b^d, i. e.,

$$M(bL) = b^d M(L) . \tag{4}$$

The solution of the functional Eq. (4) is simply $M(L) = AL^d$. For a long wire, mass changes linearly with b, i. e., $d = 1$. For the thin plates, we obtain $d = 2$, and for cubes $d = 3$; see Fig. 7.

Mandelbrot coined the name "fractal dimension", and those objects described by a fractal dimension are called fractals. Thus, to include fractal structures, (4) we can generalize

$$M(bL) = b^{d_f} M(L) , \tag{5}$$

and

$$M(L) = AL^{d_f} , \tag{6}$$

where d_f is the fractal dimension and can be a noninteger. Below, we present two examples of dealing with d_f: (i) the deterministic Sierpinski Gasket and (ii) random percolation clusters and criticality.
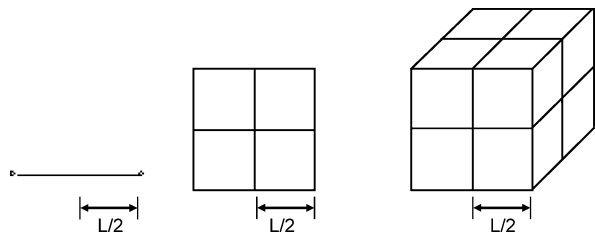
Fractals and Percolation, Figure 7
Examples of regular systems with dimensions $d = 1$, $d = 2$, and $d = 3$
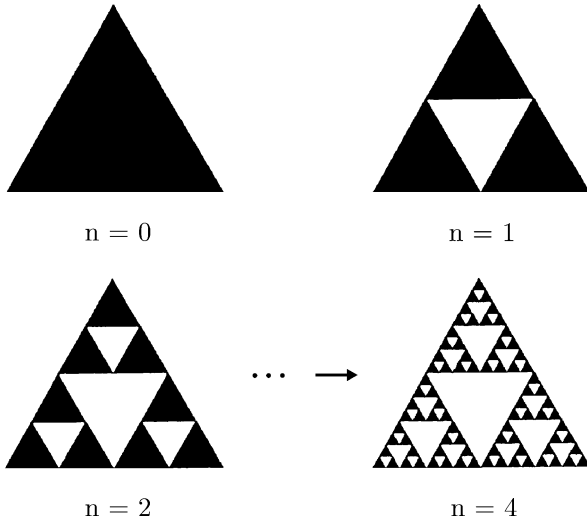
Fractals and Percolation, Figure 8
2D Sierpinski gasket. Generation and self-similarity

Fractals and Percolation, Figure 9
Lattice composed of Sierpinski gasket cells of size ξ
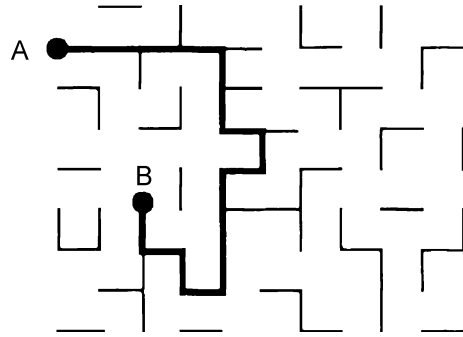
Fractals and Percolation, Figure 10
Shortest path between two sites A and B on a percolation cluster

Sierpinski Gasket   This fractal is generated by dividing a full triangle into four smaller triangles and removing the central triangle (see Fig. 8). In subsequent iterations, this procedure is repeated by dividing each of the remaining triangles into four smaller triangles and removing the central triangles. To obtain the fractal dimension, we consider the mass of the gasket within a linear size L and compare it with the mass within $\frac{1}{2}L$. Since $M(\frac{1}{2}L) = \frac{1}{3}M(L)$, we have $d_f = \log 3/\log 2 \cong 1.585$.

Percolation Fractal

We assume that at p_c ($\xi = \infty$) the clusters are fractals. Thus for $p > p_c$, we expect length scales smaller than ξ to have critical properties and therefore a fractal structure. For length scales larger than ξ, one expects a homogeneous system which is composed of many unit cells of size ξ:

$$M(r) \sim \begin{cases} r^{d_f}, & r \ll \xi, \\ r^d, & r \gg \xi. \end{cases} \qquad (7)$$

For a demonstration of this feature see Fig. 9.

One can relate the fractal dimension d_f of percolation clusters to the exponents β and ν. The probability that an arbitrary site within a circle of radius r smaller than ξ belongs to the infinite cluster, is the ratio between the number of sites on the infinite cluster and the total number of sites,

$$P_\infty \sim r^{d_f}/r^d, \qquad r < \xi. \qquad (8)$$

This equation is certainly correct for $r = a\xi$, where a is an arbitrary constant smaller than 1. Substituting $r = a\xi$

in (8) yields $P_\infty \sim \xi^{d_f}/\xi^d$. Both sides are powers of $p - p_c$. Substituting Eqs. (1) and (2) into the latter one obtains [8,28,41,49,64],

$$d_f = d - \beta/\nu . \qquad (9)$$

Thus, the fractal dimension of the infinite cluster at p_c is not a new independent exponent but depends on β and ν. Since β and ν are universal exponents, d_f is also universal.

Shortest Path Dimensions, $d_{\min}$ and d_ℓ

The fractal dimension, however, is not sufficient to fully characterize a percolation cluster, since two clusters with very different topologies may have the same fractal dimension d_f. As an additional characterization of a fractal, one can consider, e. g., the shortest path between two arbitrary sites A and B on the cluster (see Figs. 10, 11) [3,16,35, 42,55,58].

The structure formed by the sites of this path is also self-similar and is described by a fractal dimension
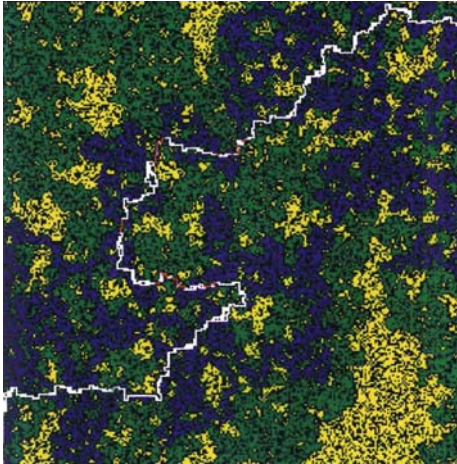
Fractals and Percolation, Figure 11
Percolation system at critical concentration in a 510 × 510 square lattice. The finite clusters are in *yellow*. Substructures of the infinite percolation cluster are shown in different colors: the shortest path between two points at opposite sites of the system is shown in *white*, the single connected sites ("red" sites) in *red*, the loops in blue and the dangling ends in *green*; courtesy of S. Schwarzer

$d_{\min}$ [46,67]. Accordingly, the length ℓ of the path, which is often called the "chemical distance", scales with the "Euclidean distance" r between A and B as

$$\ell \sim r^{d_{\min}} . \tag{10}$$

The inverse relation

$$r \sim \ell^{1/d_{\min}} \equiv \ell^{\tilde{\nu}} \tag{11}$$

tells how r scales with ℓ.

Closely related to $d_{\min}$ and d_f is the "chemical" dimension d_ℓ, which describes how the cluster mass M within chemical distance ℓ from a given site scales with ℓ,

$$M(\ell) \sim \ell^{d_\ell} . \tag{12}$$

While the fractal dimension d_f characterizes how the mass of a cluster scales with the Euclidean distance r, the graph dimension d_ℓ characterizes how the mass scales with the chemical distance ℓ. Combining Eqs. (7), (10) and (12) we obtain the relation between $d_{\min}$, d_ℓ and d_f

$$d_\ell = d_f/d_{\min} . \tag{13}$$

To measure d_f, an arbitrary site is chosen on the cluster and one determines the number $M(r)$ of sites within a distance r from this site. To measure d_ℓ, an arbitrary site is chosen on the cluster at criticality and one determines the number $M(\ell)$ of sites which are connected to

this site by a shortest path with length less than or equal to ℓ. Finally, to measure $d_{\min}$, two arbitrary sites are chosen on the cluster and one determines the length $\ell(r)$ of the shortest path connecting them. As for $M(r)$, averages must be performed for $M(\ell)$ and $r(\ell)$ over many realizations. In regular Euclidean lattices, both d_ℓ and d_f coincide with the Euclidean space dimension d and $d_{\min} = 1$.

The chemical dimension d_ℓ (or $d_{\min} = 1/\tilde{\nu}$) is an important tool for distinguishing between different fractal structures which may have a similar fractal dimension. In $d = 3$, for example, DLA (diffusion limited aggregation) clusters and percolation clusters have approximately the same fractal dimension $d_f \cong 2.5$, but have different $\tilde{\nu}$: $\tilde{\nu} = 1$ for DLA [54] but $\tilde{\nu} \cong 0.73$ for percolation [36,46,63,69].

While d_f has been related to the (known) critical exponents, (9), no such relation has been found for d_{min} or d_ℓ. The values of d_ℓ or $d_{\min}$ are known only from approximate methods, mainly numerical simulations (see also Refs. [17,37,57,76]).

Fractal Substructures

The fractal dimensions d_f and d_ℓ are not the only exponents characterizing a percolation cluster at p_c. A percolation cluster is composed of several fractal sub-structures, which are described by other exponents. Imagine applying a voltage difference between two sites at opposite edges of a metallic percolation cluster: The *backbone* of the cluster consists of those sites (or bonds) which carry the electric current. The *dangling ends* are those parts of the cluster which carry no current and are connected to the backbone by a single site only. The *red bonds* (or singly connected bonds) [22,66] are those bonds that carry the total current; when they are cut the current flow stops. In analogy to red bonds we can define anti-red bonds [34]. If an anti-red bond is added to a nonconducting percolation system below p_c, the current will be able to flow in the system. The *blobs*, finally, are those parts of the backbone that remain after the red bonds have been removed.

Further fractal substructures of the cluster are the *external perimeter* (which is also called the *hull*), the *skeleton* and the *elastic backbone*. The hull consists of those sites of the cluster which are adjacent to empty sites and are connected with infinity via empty sites. In contrast, the *total* perimeter also includes the holes in the cluster. The external perimeter is an important model for random fractal interfaces. The skeleton is defined as the union of *all* shortest paths from a given site to all sites at a chemical distance ℓ [43]. The elastic backbone is the union of all shortest paths between two sites [47].
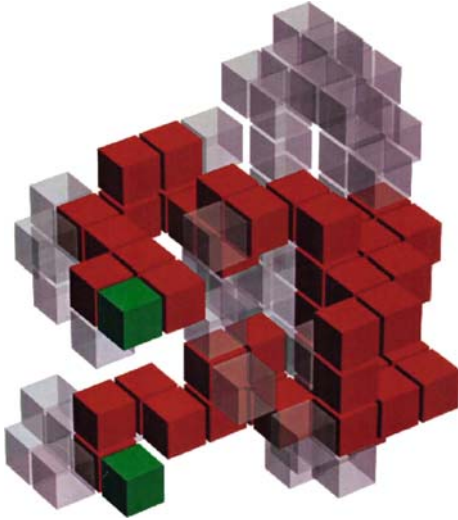
Fractals and Percolation, Figure 12
Percolation cluster at the critical concentration in a simple cubic lattice. The backbone between two (*green*) cluster sites is shown in *red*, the *gray* sites represent the dangling ends; courtesy of M. Porto

The fractal dimension d_B of the backbone is smaller than the fractal dimension d_f of the cluster (see Table 3). This reflects the fact that most of the mass of the cluster is concentrated in the dangling ends, which is seen clearly in Fig. 12. The value of the fractal dimension of the backbone is known only from numerical simulations [45,59]. Note, also, that the graph dimension d_ℓ^B of the backbone is smaller than that of percolation. In contrast, $\tilde{\nu}$ is the *same* for both backbone and percolation cluster, indicating the more universal nature of $\tilde{\nu}$. This can be understood by recalling that every two sites on a percolation cluster are located on the corresponding backbone.

The fractal dimensions of the red bonds d_{red} and the hull d_h are known from exact analytical arguments. It has been proven by Coniglio [22,23] that the mean number of red bonds varies with p as

$$n_{red} \sim (p - p_c)^{-1} \sim \xi^{1/\nu} \sim r^{1/\nu} , \tag{14}$$

and the fractal dimension of the red bonds is therefore $d_{red} = 1/\nu$. The fractal dimension of the skeleton is very close to $d_{min} = 1/\tilde{\nu}$, supporting the assumption that percolation clusters at criticality are finitely ramified [43].

The hull of the cluster in $d = 2$ has the fractal dimension $d_h = 7/4$, which was first found numerically by Sapoval, Rosso, and Gouyet [63] and proven rigorously by Saleur and Duplantier [62]. If the hull is defined slightly differently and next-nearest neighbors of the perimeter are regarded as connected, many "fjords" are removed from

Fractals and Percolation, Table 3
Fractal dimensions of the substructures composing percolation clusters (see Ref. [8,14,41,59,64] and references therein). For fractal dimensions in $d = 4$ and $d = 5$ see Ref. [57]

Fractal dimensions	Space dimension		
	$d = 2$	$d = 3$	$d \geq 6$
d_f	91/48	2.524 ± 0.008	4
d_ℓ	1.678 ± 0.005	1.84 ± 0.02	2
d_{min}	1.13 ± 0.004	1.374 ± 0.004	2
d_{red}	3/4	1.143 ± 0.01	2
d_h	7/4	2.548 ± 0.014	4
d_B	1.64 ± 0.02	1.87 ± 0.04	2
d_ℓ^B	1.43 ± 0.02	1.34 ± 0.03	1

the hull. According to Grossmann and Aharony [38], the fractal dimension of this modified hull is close to 4/3, the fractal dimension of self-avoiding random walks in $d = 2$. In three dimensions, in contrast, the mass of the hull seems to be proportional to the mass of the cluster, and both have the same fractal dimension.

In Table 3 we summarize the values of the fractal dimension d_f and the graph dimension d_ℓ of the percolation cluster and its fractal substructures.

Anomalous Transport on Percolation Clusters: Diffusion and Conductivity

Due to self-similarity, transport quantities are significantly modified on fractal substrates. This can be seen in two representative examples:

(1) total resistance or the conductivity,
(2) mean square displacement and the probability density of random walks.

Consider a metallic network of size L^d. At opposite faces of the network are metallic bars with a voltage difference between them.

If we vary the linear size L of the system, the total resistance R varies as

$$R \sim \sigma^{-1} L / L^{d-1} , \tag{15}$$

where $\sigma \sim L^0 = $ const is the conductivity (inverse to the resistivity, $\sigma = \rho^{-1}$) of the metal. Since σ does not depend on L, (15) states that the total resistance of the network depends on its linear size L via the power law $R \sim L^{2-d} \equiv L^{\tilde{\zeta}}$, which defines the resistance exponent $\tilde{\zeta}$, here $\tilde{\zeta} = 2 - d$.

The idea that transport properties of percolation systems can be efficiently studied by means of diffusion was suggested by de Gennes [24] (see also Kopelman [50]).

The diffusion process can be modeled by random walkers, which can jump randomly between nearest-neighbor occupied sites in the lattice. For such a random walker moving in a disordered environment, including bottlenecks, loops, and dead ends, de Gennes coined the term *ant in the labyrinth*.

By calculating the mean square displacement of the walker, one obtains the diffusion constant, which according to Einstein is proportional to dc conductivity. Not only are the conductivity and diffusion exponents above p_c related; also related are the exponents characterizing the size dependence of the dc conductivity and the time dependence of the mean square displacement of the random walker. Since it is numerically more efficient to calculate the relevant transport quantities by simulating random walks than to determine the conductivity directly from Kirchhoff's equations, the study of random walks has improved our knowledge not only of diffusion but also of the transport process in percolation in general [5,6,10,12, 18,39,40,56,74].

Due to the presence of large holes, bottlenecks, and dangling ends in the fractal, the motion of a random walker is slowed down. Fick's law for the mean square displacement ($\langle r^2(t) \rangle = a^2 t$) is no longer valid. Instead, the mean square displacement is described by a more general power law,

$$\langle r^2(t) \rangle \sim t^{2/d_w} , \tag{16}$$

where the new exponent d_w ("diffusion exponent" or "fractal dimension of the random walk") is always greater than 2.

Both the resistance exponent $\tilde{\zeta}$ and the exponent d_w can be related by the Einstein equation

$$\sigma = (e^2 n / k_B T) D , \tag{17}$$

which relates the dc conductivity σ of the system to the diffusion constant $D = \lim_{t \to \infty} \langle r^2(t) \rangle / 2dt$ of the random walk. In Eq. (17), e and n denote charge and density of the mobile particles, respectively.

Simple scaling arguments can now be used to relate d_w to $\tilde{\zeta}$ and $\tilde{\mu}$. Since n is proportional to the density of the substrate, $n \sim L^{d_f - d}$, the right-hand side of Eq. (17) is proportional to $L^{d_f - d} t^{2/d_w - 1}$. The left-hand side of Eq. (17) is proportional to $L^{-\tilde{\mu}}$. Since the time a random walker takes to travel a distance L scales as L^{d_w}, we find $L^{-\tilde{\mu}} \sim L^{d_f - d + 2 - d_w}$, from which the Einstein relation [4]

$$d_w = d_f - d + 2 + \tilde{\mu} = d_f + \tilde{\zeta} \tag{18}$$

follows. For example, for the Sierpinski gasket $d_f = \log 3/\log 2$, and $\tilde{\zeta} = \log(5/3)/\log(2/4)$, therefore $d_w = \log 5/\log 2$.

In general, determining d_w for random fractals is not easy. An exception is topologically linear fractal structures ($d_\ell = 1$), which can be considered as nonintersecting paths. Along a path (in ℓ-space), diffusion is normal and $\langle \ell^2(t) \rangle = t$. Since $\ell \sim r^{d_f}$, the mean square displacement in r-space scales as $\langle r^2 \rangle \sim t^{1/d_f}$, leading to $d_w = 2d_f$ in this case. In percolation, d_w cannot be calculated exactly, but upper and lower bounds can be derived which are very close to each other in $d \geq 3$ dimensions. A good estimate is $d_w \cong 3d_f/2$ (Alexander–Orbach conjecture [4]).

The long-term behavior of the mean square displacement of a random walker on an infinite percolation cluster is characterized by the diffusion constant D. It is easy to see that D is related to the diffusion constant D' of the whole percolation system: above p_c, the dc conductivity of the percolation system increases as $\sigma \sim (p - p_c)^\mu$, so due to the Einstein relation, Eq. (17), the diffusion constant D' must also increase in this way. The mean square displacement (and hence D') is obtained by averaging over all possible starting points of a particle in the percolation system. It is clear that only those particles which start on the infinite cluster can travel from one side of the system to the other, and thus contribute to D'. Particles that start on a finite cluster cannot leave the cluster, and thus do not contribute to D'. Hence D' is related to D by $D' = DP_\infty$, implying

$$D \sim (p - p_c)^{\mu - \beta} \sim \xi^{-(\mu - \beta)/\nu} . \tag{19}$$

Combining (16) and (19), the mean square displacement on the infinite cluster can be written as [7,33,72]

$$\langle r^2(t) \rangle \sim \begin{cases} t^{2/d_w} & \text{if } t \ll t_\xi , \\ (p - p_c)^{\mu - \beta} t & \text{if } t \gg t_\xi , \end{cases} \tag{20}$$

where

$$t_\xi \sim \xi^{d_w} \tag{21}$$

describes the time scale the random walker needs, on average, to explore the fractal regime in the cluster. As $\xi \sim (p - p_c)^{-\nu}$ is the only length scale here, t_ξ is the only relevant time scale, and we can bridge the short time regime and the long time regime by a scaling function $f(t/t_\xi)$,

$$\langle r^2(t) \rangle = t^{2/d_w} f(t/t_\xi) . \tag{22}$$

To satisfy (20)–(21), we require $f(x) \sim x^0$ for $x \ll 1$ and $f(x) \sim x^{1 - 2/d_w}$ for $x \gg 1$. The first relation trivially satisfies (20)–(21). The second relation gives $D = \lim_{t \to \infty} \langle r^2(t) \rangle / 2dt \sim t_\xi^{2/d_w - 1}$, which in connection

with (19) and (21) yields a relation between d_w and μ [7, 33,72],

$$d_w = 2 + (\mu - \beta)/\nu . \tag{23}$$

Comparing (18) and (23) one can express the exponent $\tilde{\mu}$ by μ,

$$\tilde{\mu} = \mu/\nu . \tag{24}$$

Networks

Networks are defined as nodes connected by links, called graphs in mathematics. Many real world system can be describe as networks. Perhaps the best known example of a network is the Internet, where computers (nodes) around the globe are connected by cables (links) in such a way that an email message can travel from one computer to another by traveling along only a few links. Social relations between people can be represented by a social network [2]; nodes represent people and links represent their relations. One important property of a network is the "small world" phenomena. The shortest path (minimum number of hops) between any two nodes is very small, of the order of $\log N$ or smaller, where N is the number of nodes in the network [11,19,29]. The lattices discussed in Sect. "Percolation" are also networks, where the sites of the lattice are the nodes and the bonds represent the links. In this case, the number of links per node is fixed but, in general, the number of links per node can be taken from any degree distribution, $P(k)$. In lattices, due to spatial constraints, the distances between nodes is large and scales as $N^{1/d}$, where d is the dimension of the lattice. Since many networks have no spatial constraints, it follows that such networks can be regarded as embedded in infinite dimension, $d = \infty$, justifying a very small distance, of order $\log N$. In networks which are not constrained to geographical space, there is no Euclidean distance and the distance metric is only the shortest path ℓ defined in Sect. "Percolation Clusters as Fractals".

We will show in this chapter that ideas from percolation and fractals can be applied to obtain useful results in networks which are not embedded in space. The main difference compared to lattices is that the condition for percolation is no longer the spanning property, but rather the property of having a cluster containing number of nodes of order N, where N is the total original number of nodes in the network. Such a component, if it exists, is termed the *giant component*. The condition for the existence of a giant component above the percolation threshold, and its absence below the threshold, applies also to lattices, and therefore can be regarded as more general than the spanning property. An interesting property of percolation, called *universality*, is that the behavior at and close to the critical point depends only on the dimensionality of the lattice, and not on the microscopic connection details of the lattice. This behavior is characterized by a set of critical exponents that are the same for all two-dimensional lattices, square, triangular or hexagonal, and for either site or bond percolation. However, a different set of critical exponents will be obtained for a lattice of another dimension. Furthermore, above some critical dimension ($d_c = 6$ for percolation in d-dimensional lattices), known as the upper critical dimension, the critical behavior remains the same for all $d \geq d_c$. This is due to the insignificance of loops in high dimensions, and thus usually allows for easy determination of the critical exponents for high dimensions, using the "infinite dimensional" or "mean field" approach. Erdős and Rényi (ER) studied an ensemble of networks with N nodes and $2M$ links that randomly connect pairs of nodes. They found that $p_c = 1/\langle k \rangle = N/2M$. Percolation on ER networks or on infinite dimensional lattices, as well as on Cayley trees, has the same critical exponents, due to the fact that their topology is the same and no spatial constraint is imposed on the networks. For ER networks, as for lattices, in $d \geq 6$ the size S of the percolation cluster at p_c, scales with N as, $S \sim N^{2/3}$ [11,20].

The value of 2/3 can be related to the upper critical dimension, $d_c = 6$, and to the fractal dimension of percolation clusters $d_f = 4$ for $d = 6$. Since $N = L^6$ and $S = L^4$, it follows that $S \sim N^{2/3}$.

In recent years it was realized [2] that $P(k)$ for many real networks is very broad and, in many cases, is best represented by a power law, $P(k) \sim k^{-\gamma}$. Networks with a power law degree distribution are called *scale free (SF) network*. Heterogeneity of the degrees may affect critical behavior, even above the upper critical dimension. The heterogeneity of the degrees can be regarded as a breakdown of translational symmetry that exists in lattices, ER networks and Cayley trees. In these cases, each node has a typical number of neighbors, while in scale free networks the variation between node degrees is very large. A general result for p_c for any random network with a given degree distribution is [21]

$$p_c = \frac{1}{\kappa - 1} , \quad \kappa \equiv \frac{\langle k^2 \rangle}{\langle k \rangle} .$$

This result yields that for $\gamma < 3$, $\langle k^2 \rangle \to \infty$ and therefore $p_c \to 0$. That is, no finite percolation threshold exists. Thus, even if most of the nodes of the Internet are removed, those which are left can still communicate. This explains the puzzle of why viruses and worms stay a long time in the Internet even if many people use antivirus soft-

ware. It also explains why in order to effectively immunize populations, one needs to immunize most of the people.

The percolation critical exponents for SF networks are still mean-field or infinite-dimensional in the sense of the insignificance of loops. However, they are different from those of the standard mean field percolation. Indeed, for scale free networks [2], the size of the spanning percolation cluster is [20],

$$S \sim \begin{cases} N^{(\gamma-2)/(\gamma-1)}, & 3 < \gamma < 4 \\ N^{2/3}, & \gamma > 4 \end{cases} \qquad (25)$$

As shown above, for $\gamma < 3$, there is no percolation threshold, and therefore no spanning percolation cluster. Note that SF networks can be regarded as a generalization of ER networks, since for $\gamma > 4$ one obtains the ER network results.

Summary and Future Directions

The percolation problem and its numerous modifications can be useful in describing several physical, chemical, and biological processes, such as the spreading of epidemics or forest fires, gelation processes, and the invasion of water into oil in porous media, which is relevant for the process of recovering oil from porous rocks. In some cases, modification changes the universality class of a percolation. We begin with an example in which the universality class does not change. We showed that a random process such as percolation can lead naturally to fractal structures. This may be one of the reasons why fractals occur frequently in nature.

Bibliography

Primary Literature

1. Aharony A (1986) In: Grinstein G, Mazenko G (eds) Directions in condensed matter physics. World Scientific, Singapore
2. Albert R, Barabasi A-L (2002) Statistical mechanics of complex networks. Rev Mod Phys 74:47
3. Alexandrowicz Z (1980) Phys Lett A 80:284
4. Alexander S, Orbach R (1982) J Phys Lett 43:L625
5. Alexander S, Bernasconi J, Schneider WR, Orbach R (1981) Rev Mod Phys 53:175; Alexander S (1983) In: Deutscher G, Zallen R, Adler J (eds) Percolation Structures and Processes. Adam Hilger, Bristol, p 149
6. Avnir D (ed) (1989) The fractal approach to heterogeneos chemistry. Wiley, Chichester
7. Ben-Avraham D, Havlin S (1982) J Phys A 15:L691; Havlin S, Ben-Avraham D, Sompolinsky H (1983) Phys Rev A 27: 1730
8. Ben-Avraham D, Havlin S (2000) Diffusion and reactions in fractals and disordered systems. Cambridge University Press, Cambridge
9. Benguigui L (1984) Phys Rev Lett 53:2028
10. Blumen A, Klafter J, Zumofen G (1986) In: Zschokke I (ed) Optical spectroscopy of glasses. Reidel, Dordrecht, pp 199–265
11. Bollobás B (1985) Random graphs. Academic Press, London
12. Bouchaud JP, Georges A (1990) Phys Rep 195:127
13. Bunde A (1986) Adv Solid State Phys 26:113
14. Bunde A, Havlin S (eds) (1996) Fractals and disordered systems, 2nd edn. Springer, Berlin; Bunde A, Havlin S (eds) (1995) Fractals in Science, 2nd edn. Springer, Berlin
15. Broadbent SR, Hammersley JM (1957) Proc Camb Phil Soc 53:629
16. Cardey JL, Grassberger P (1985) J Phys A 18:L267
17. Cardy J (1998) J Phys A 31:L105
18. Clerc JP, Giraud G, Laugier JM, Luck JM (1990) Adv Phys 39:191
19. Cohen R, Havlin S (2003) Scale-free networks are ultrasmall. Phys Rev Lett 90:058701
20. Cohen R, Havlin S (2008) Complex networks: Structure, stability and function. Cambridge University Press, Cambridge
21. Cohen R, Erez K, Ben-Avraham D, Havlin S (2000) Resilience of the internet to random breakdowns. Phys Rev Lett 85:4626
22. Coniglio A (1982) J Phys A 15:3829
23. Coniglio A (1982) Phys Rev Lett 46:250
24. de Gennes PG (1976) La Recherche 7:919
25. de Gennes PG (1979) Scaling concepts in polymer physics. Cornell University Press, Ithaca
26. Deutscher G, Zallen R, Adler J (eds) (1983) A collection of review articles: percolation structures and processes. Adam Hilger, Bristol
27. Domb C (1983) In: Deutscher G, Zallen R, Adler J (eds) Percolation structures and processes. Adam Hilger, Bristol; Domb C, Stoll E, Schneider T (1980) Contemp Phys 21: 577
28. Elam WT, Kerstein AR, Rehr JJ (1984) Phys Rev Lett 52:1515
29. Erdős P, Rényi A (1959) On random graphs. Publicationes Mathematicae 6:290; (1960) Publ Math Inst Hung Acad Sci 5:17
30. Essam JW (1980) Rep Prog Phys 43:843
31. Family F, Landau D (eds) (1984) Kinetics of aggregation and gelation. North Holland, Amsterdam; For a review on gelation see: Kolb M, Axelos MAV (1990) In: Stanley HE, Ostrowsky N (eds) Correlatios and Connectivity: Geometric Aspects of Physics, Chemistry and Biology. Kluwer, Dordrecht, p 225
32. Flory PJ (1971) Principles of polymer chemistry. Cornell University, New York; Flory PJ (1941) J Am Chem Soc 63:3083–3091–3096; Stockmayer WH (1943) J Chem Phys 11:45
33. Gefen Y, Aharony A, Alexander S (1983) Phys Rev Lett 50:77
34. Gouyet JF (1992) Phys A 191:301
35. Grassberger P (1986) Math Biosci 62:157; (1985) J Phys A 18:L215; (1986) J Phys A 19:1681
36. Grassberger P (1992) J Phys A 25:5867
37. Grassberger P (1999) J Phys A 32:6233
38. Grossman T, Aharony A (1987) J Phys A 20:L1193
39. Haus JW, Kehr KW (1987) Phys Rep 150:263
40. Havlin S, Ben-Avraham D (1987) Adv Phys 36:695
41. Havlin S, Ben-Avraham D (1987) Diffusion in random media. Adv Phys 36:659
42. Havlin S, Nossal R (1984) Topological properties of percolation clusters. J Phys A 17:L427
43. Havlin S, Nossal R, Trus B, Weiss GH (1984) J Stat Phys A 17:L957
44. Herrmann HJ (1986) Phys Rep 136:153
45. Herrmann HJ, Stanley HE (1984) Phys Rev Lett 53:1121; Hong DC, Stanley HE (1984) J Phys A 16:L475
46. Herrmann HJ, Stanley HE (1988) J Phys A 21:L829
47. Herrmann HJ, Hong DC, Stanley HE (1984) J Phys A 17:L261

48. Kesten H (1982) Percolation theory for mathematicians. Birkhauser, Boston (A mathematical approach); Grimmet GR (1989) Percolation. Springer, New York

49. Kirkpatrick S (1979) In: Maynard R, Toulouse G (eds) Le Houches Summer School on Ill Condensed Matter. North Holland, Amsterdam

50. Kopelman R (1976) In: Fong FK (ed) Topics in applied physics, vol 15. Springer, Heidelberg

51. Ma SK (1976) Modern theory of critical phenomena. Benjamin, Reading

52. Mackay G, Jan N (1984) J Phys A 17:L757

53. Mandelbrot BB (1982) The fractal geometry of nature. Freeman, San Francisco; Mandelbrot BB (1977) Fractals: Form, Chance and Dimension. Freeman, San Francisco

54. Meakin P, Majid I, Havlin S, Stanley HE (1984) J Phys A 17:L975

55. Middlemiss KM, Whittington SG, Gaunt DC (1980) J Phys A 13:1835

56. Montroll EW, Shlesinger MF (1984) In: Lebowitz JL, Montroll EW (eds) Nonequilibrium phenomena II: from stochastics to hydrodynamics. Studies in Statistical Mechanics, vol 2. North-Holland, Amsterdam

57. Paul G, Ziff RM, Stanley HE (2001) Phys Rev E 64:26115

58. Pike R, Stanley HE (1981) J Phys A 14:L169

59. Porto M, Bunde A, Havlin S, Roman HE (1997) Phys Rev E 56:1667

60. Ritzenberg AL, Cohen RI (1984) Phys Rev B 30:4036

61. Sahimi M (1993) Application of percolation theory. Taylor Francis, London

62. Saleur H, Duplantier B (1987) Phys Rev Lett 58:2325

63. Sapoval B, Rosso M, Gouyet JF (1985) J Phys Lett 46:L149

64. Stauffer D, Aharony A (1994) Introduction to percolation theory, 2nd edn. Taylor Francis, London

65. Stanley HE (1971) Introduction to phase transition and critical phrenomena. Oxford University, Oxford

66. Stanley HE (1977) J Phys A 10:L211

67. Stanley HE (1984) J Stat Phys 36:843

68. Turcotte DL (1992) Fractals and chaos. In: Geology and geophysics. Cambridge University Press, Cambridge

69. Toulouse G (1974) Nuovo Cimento B 23:234

70. Tyc S, Halperin BI (1989) Phys Rev B 39:R877; Strelniker YM, Berkovits R, Frydman A, Havlin S (2004) Phys Rev E 69:R065105; Strelniker YM, Havlin S, Berkovits R, Frydman A (2005) Phys Rev E 72:016121; Strelniker YM (2006) Phys Rev B 73:153407

71. von Niessen W, Blumen A (1988) Canadian J For Res 18:805

72. Webman I (1991) Phys Rev Lett 47:1496

73. Webman I, Jortner J, Cohen MH (1976) Phys Rev B 14:4737

74. Weiss GH, Rubin RJ (1983) Adv Chem Phys 52:363; Weiss GH (1994) Aspects and applications of the random walk. North Holland, Amsterdam

75. Ziff RM (1992) Phys Rev Lett 69:2670

76. Ziff RM (1999) J Phys A 32:L457

Books and Reviews

Bak P (1996) How nature works: The science of self organized criticality. Copernicus, New York

Barabási A-L (2003) Linked: How everything is connected to everything else and what it means for business, science and everyday life. Plume

Bergman DJ, Stroud D (1992) Solid State Phys 46:147–269

Dorogovtsev SN, Mendes JFF (2003) Evolution of networks: From biological nets to the internet and www (physics). Oxford University Press, Oxford

Eglash R (1999) African fractals: Modern computing and indigenous design. Rutgers University Press, New Brunswick, NJ

Feder J (1988) Fractals. Plenum, New York

Gleick J (1997) Chaos. Penguin Books, New York

Gould H, Tobochnik J (1988) An introduction to computer simulation methods. In: Application to physical systems. Addison-Wesley, Reading, MA

Meakin P (1998) Fractals, scaling and growth far from equilibrium. Cambridge University Press, Cambridge

Pastor-Satorras R, Vespignani A (2004) Evolution and structure of the internet: A statistical physics approach. Cambridge University Press, Cambridge

Peitgen HO, Jurgens H, Saupe D (1992) Chaos and fractals. Springer, New York

Peng G, Decheng T (1990) The fractal nature of a fracture surface. J Physics A 14:3257–3261

Pikovsky A, Rosenblum M, Kurths J, Chirikov B, Cvitanovic P, Moss F, Swinney H (2003) Synchronization: A universal concept in nonlinear sciences. Cambridge University Press, Cambridge

Vicsek T (1992) Fractal growth phenomena. World Scientific, Singapore

Fractals in the Quantum Theory of Spacetime

LAURENT NOTTALE
CNRS, Paris Observatory and Paris Diderot University, Paris, France

Article Outline

Glossary

Fractality In the context of the present article, the geometric property of being structured over all (or many) scales, involving explicit scale dependence which may go up to scale divergence.

Spacetime Inter-relational level of description of the set of all positions and instants (events) and of their transformations. The events are defined with respect to a given reference system (i. e., in a relative way), but a spacetime is characterized by invariant relations which are valid in all reference systems, such as, e. g., the metric invariant. In the generalization to a fractal space-time, the events become explicitly dependent on resolution.

Relativity The property of physical quantities according to which they can be defined only in terms of relationships, not in an abolute way. These quantities depend on the state of the reference system, itself defined in a relative way, i. e., with respect to other coordinate systems.

Covariance Invariance of the form of equations under general coordinate transformations.

Geodesics Curves in a space (more generally in a spacetime) which minimize the proper time. In a geometric spacetime theory, the motion equation is given by a geodesic equation.

Quantum Mechanics Fundamental axiomatic theory of elementary particle, nuclear, atomic, molecular, etc. physical phenomena, according to which the state of a physical system is described by a wave function whose square modulus yields the probability density of the variables, and which is solution of a Schrödinger equation constructed from a correspondence principle (among other postulates).

Definition of the Subject

The question of the foundation of quantum mechanics from first principles remains one of the main open problems of modern physics. In its current form, it is an axiomatic theory of an algebraic nature founded upon a set of postulates, rules and derived principles. This is to be compared with Einstein's theory of gravitation, which is founded on the principle of relativity and, as such, is of an essentially geometric nature. In its framework, gravitation is understood as a very manifestation of the curvature of a Riemannian space-time.

It is therefore relevant to question the nature of the quantum space-time and to ask for a possible refoundation of the quantum theory upon geometric first principles. In this context, it has been suggested that the quantum laws and properties could actually be manifestations of a fractal and nondifferentiable geometry of space-time [52,69,71], coming under the principle of scale relativity [53,54]. This principle extends, to scale transformations of the reference system, the theories of relativity (which have been, up to

now, applied to transformations of position, orientation and motion).

Such an approach allows one to recover the main tools and equations of standard quantum mechanics, but also to suggest generalizations, in particular toward high energies, since it leads to the conclusion that the Planck length scale could be a minimum scale in nature, unreachable and invariant under dilations [53].

But it has another radical consequence. Namely, it allows the possibility of a new form of macroscopic quantum-type behavior for a large class of complex systems, namely those whose behavior is characterized by Newtonian dynamics, fractal stochastic fluctuations over a large range of scales, and small scale irreversibility. In this case the equation of motion may take the form of a Schrödinger equation, which yields peaks of probability density according to the symmetry, field and limit conditions. These peaks may be interpreted as a tendency for the system to form structures [57], in terms of a macroscopic constant which is no longer $\hbar$, therefore possibly leading to a new theory of self-organization.

It is remarkable that, under such a relativistic view, the question of complexity may be posed in an original way. Namely, in a fully relativistic theory there is no intrinsic complexity, since the various physical properties of an 'object' are expected to vanish in the proper system of coordinates linked to the object. Therefore, in such a framework the apparent complexity of a system comes from the complexity of the change of reference frames from the proper frame to the observer (or measurement apparatus) reference frame. This does not mean that the complexity can always be reduced, since this change can itself be infinitely complex, as it is the case in the situation described here of a fractal and nondifferentiable space-time.

Introduction

There have been many attempts during the 20th century at understanding the quantum behavior in terms of differentiable manifolds. The failure of these attempts indicates that a possible 'quantum geometry' should be of a completely new nature. Moreover, following the lessons of Einstein's construction of a geometric theory of gravitation, it seems clear that any geometric property to be attributed to space-time itself, and not only to particular objects or systems embedded in space-time, must necessarily be universal.

Fortunately, the founders of quantum theory have brought to light a universal and fundamental behavior of the quantum realm, in opposition to the classical world; namely, the explicit dependence of the measurement re-

sults on the apparatus resolution described by the Heisenberg uncertainty relations. This leads one to contemplate the possibility that the space-time underlying the quantum mechanical laws can be explicitly dependent on the scale of observation [52,69,71]. Now the concept of a scale-dependent geometry (at the level of objects and media) has already been introduced and developed by Benoit Mandelbrot, who coined the word 'fractal' in 1975 to describe it. But here we consider a complementary program that uses fractal geometry, not only for describing 'objects' (that remain embedded in an Euclidean space), but also for intrinsically describing the geometry of space-time itself.

A preliminary work toward such a goal may consist of introducing the fractal geometry in Einstein's equations of general relativity at the level of the source terms. This would amount to giving a better description of the density of matter in the Universe accounting for its hierarchical organization and fractality over many scales (although possibly not all scales), then to solve Einstein's field equations for such a scale dependent momentum-energy tensor. A full implementation of this approach remains a challenge to cosmology.

But a more direct connection of the fractal geometry with fundamental physics comes from its use in describing not only the distribution of matter in space, but also the geometry of space-time itself. Such a goal may be considered as the continuation of Einstein's program of generalization of the geometric description of space-time. In the new fractal space-time theory, [27,52,54,69,71], the essence of quantum physics is a manifestation of the nondifferentiable and fractal geometry of space-time.

Another line of thought leading to the same suggestion comes, not from relativity and space-time theories, but from quantum mechanics itself. Indeed, it has been discovered by Feynman [30] that the typical quantum mechanical paths (i. e., those that contribute in a main way to the path integral) are nondifferentiable and fractal. Namely, Feynman has proved that, although a mean velocity can be defined for them, no mean-square velocity exists at any point, since it is given by $\langle v^2 \rangle \propto \delta t^{-1}$. One now recognizes in this expression the behavior of a curve of fractal dimension $D_F = 2$ [1].

Based on these premises, the reverse proposal, according to which the laws of quantum physics find their very origin in the fractal geometry of space-time, has been developed along three different and complementary approaches.

Ord and co-workers [50,71,72,73], extending the Feynman chessboard model, have worked in terms of probabilistic models in the framework of the statistical mechanics of binary random walks.

El Naschie has suggested to give up not only the differentiability, but also the continuity of space-time. This leads him to work in terms of a 'Cantorian' space-time [27,28,29], and to therefore use in a preferential way the mathematical tool of number theory (see a more detailed review of these two approaches in Ref. [45]).

The scale relativity approach [52,54,56,63,69] which is the subject of the present article, is, on the contrary, founded on a fundamentally continuous geometry of space-time which therefore includes the differentiable and nondifferentiable cases, constrained by the principle of relativity applied to both motion and scale.

Other applications of fractals to the quantum theory of space-time have been proposed in the framework of a possible quantum gravity theory. They are of another nature than those considered in the present article, since they are applicable only in the framework of the quantum theory instead of deriving it from the geometry, and they concern only very small scales on the magnitude of the Planck scale. We send the interested reader to Kröger's review paper on "Fractal geometry in quantum mechanics, field theory and spin systems", and to references therein [45].

In the present article, we summarize the steps by which one recovers, in the scale relativity and fractal space-time framework, the main tools and postulates of quantum mechanics and of gauge field theories. A more detailed account can be found in Refs. [22,23,54,56,63,67,68], including possible applications of the theory to various sciences.

Foundations of Scale Relativity Theory

The theory of scale relativity is based on giving up the hypothesis of manifold differentiability. In this framework, the coordinate transformations are continuous but can be nondifferentiable. This has several consequences [54], leading to the following preliminary steps of construction of the theory:

(1) One can prove the following theorem [5,17,18,54,56]: a continuous and nondifferentiable curve is fractal in a general meaning, namely, its length is explicitly dependent on a scale variable ε, i. e., $\mathcal{L} = \mathcal{L}(\varepsilon)$, and it diverges, $\mathcal{L} \to \infty$, when $\varepsilon \to 0$. This theorem can be readily extended to a continuous and nondifferentiable manifold, which is therefore fractal not as an hypothesis, but as a consequence of giving up an hypothesis (that of differentiability).

(2) The fractality of space-time [52,54,69,71] involves the scale dependence of the reference frames. One therefore adds a new variable ε which characterizes the 'state of scale' to the usual variables defining the coordinate system.

In particular, the coordinates themselves become functions of these scale variables, i. e., $X = X(\varepsilon)$.

(3) The scale variables ε can never be defined in an absolute way, but only in a relative way. Namely, only their ratio $\rho = \varepsilon'/\varepsilon$ has a physical meaning. In experimental situations, these scale variables amount to the resolution of the measurement apparatus (it may be defined as standard errors, intervals, pixel size, etc.). In a theoretical analysis, they are the space and time differential elements themselves. This universal behavior extends to the principle of relativity in such a way that it also applies to the transformations (dilations and contractions) of these resolution variables [52,53,54].

Scale Laws

Fractal Coordinate and Differential Dilation Operator

Consider a variable length measured on a fractal curve, and (more generally) a non-differentiable (fractal) curvilinear coordinate $\mathcal{L}(s, \varepsilon)$, that depends on some parameter s which characterizes the position on the curve (it may be, e. g., a time coordinate), and on the resolution ε. Such a coordinate generalizes the concept of curvilinear coordinates introduced for curved Riemannian space-times in Einstein's general relativity [54] to nondifferentiable and fractal space-times.

Such a scale-dependent fractal length $\mathcal{L}(s, \varepsilon)$ remains finite and differentiable when $\varepsilon \neq 0$; namely, one can define a slope for any resolution ε, being aware that this slope is itself a scale-dependent fractal function. It is only at the limit $\varepsilon \to 0$ that the length is infinite and the slope undefined, i. e., that nondifferentiability manifests itself.

Therefore the laws of dependence of this length upon position and scale may be written in terms of a double differential calculus, i. e., it can be the solution of differential equations involving the derivatives of $\mathcal{L}$ with respect to both s and ε.

As a preliminary step, one needs to establish the relevant form of the scale variables and the way they intervene in scale differential equations. For this purpose, let us apply an infinitesimal dilation $\mathrm{d}\rho$ to the resolution, which is therefore transformed as $\varepsilon \to \varepsilon' = \varepsilon(1 + \mathrm{d}\rho)$. The dependence on position is omitted at this stage in order to simplify the notation. By applying this transformation to a fractal coordinate $\mathcal{L}$, one obtains, to the first order in the differential element,

$$\mathcal{L}(\varepsilon') = \mathcal{L}(\varepsilon + \varepsilon\,\mathrm{d}\rho) = \mathcal{L}(\varepsilon) + \frac{\partial\mathcal{L}(\varepsilon)}{\partial\varepsilon}\,\varepsilon\,\mathrm{d}\rho$$
$$= (1 + \tilde{D}\mathrm{d}\rho)\,\mathcal{L}(\varepsilon)\,, \tag{1}$$

where $\tilde{D}$ is, by definition, the dilation operator.

Since $\mathrm{d}\varepsilon/\varepsilon = \mathrm{d}\ln\varepsilon$, the identification of the two last members of Eq. (1) yields

$$\tilde{D} = \varepsilon\,\frac{\partial}{\partial\varepsilon} = \frac{\partial}{\partial\ln\varepsilon}\,. \tag{2}$$

This form of the infinitesimal dilation operator shows that the natural variable for the resolution is $\ln\varepsilon$, and that the expected new differential equations will indeed involve quantities such as $\partial\mathcal{L}(s, \varepsilon)/\partial\ln\varepsilon$. This theoretical result agrees and explains the current knowledge according to which most measurement devices (of light, sound, etc.), including their physiological counterparts (eye, ear, etc.) respond according to the logarithm of the intensity (e. g., magnitudes, decibels, etc.).

Self-similar Fractals as Solutions of a First Order Scale Differential Equation

Let us start by writing the simplest possible differential equation of scale, and then solve it. We shall subsequently verify that the solutions obtained comply with the principle of relativity. As we shall see, this very simple approach already yields a fundamental result: it gives a foundation and an understanding from first principles for self-similar fractal laws, which have been shown by Mandelbrot and many others to be a general description of a large number of natural phenomena, in particular biological ones (see, e. g., [48,49,70], other volumes of these series and references therein). In addition, the obtained laws, which combine fractal and scale-independent behaviours, are the equivalent for scales of what inertial laws are for motion [49]. Since they serve as a fundamental basis of description for all the subsequent theoretical constructions, we shall now describe their derivation in detail.

The simplest differential equation of explicit scale dependence which one can write is of first order and states that the variation of $\mathcal{L}$ under an infinitesimal scale transformation $\mathrm{d}\ln\varepsilon$ depends only on $\mathcal{L}$ itself. Basing ourselves on the previous derivation of the form of the dilation operator, we thus write

$$\frac{\partial\mathcal{L}(s, \varepsilon)}{\partial\ln\varepsilon} = \beta(\mathcal{L})\,. \tag{3}$$

The function β is a priori unknown. However, still looking for the simplest form of such an equation, we expand $\beta(\mathcal{L})$ in powers of $\mathcal{L}$, namely we write $\beta(\mathcal{L}) = a + b\mathcal{L} + \ldots$. Disregarding for the moment the s dependence, we obtain, to the first order, the following linear equation in which a and b are constants:

$$\frac{\mathrm{d}\mathcal{L}}{\mathrm{d}\ln\varepsilon} = a + b\mathcal{L}\,. \tag{4}$$

variation of the length

variation of the fractal dimension
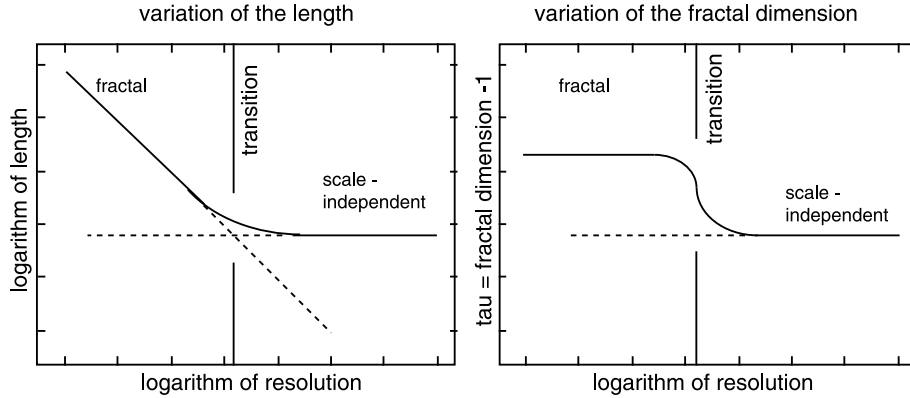
Fractals in the Quantum Theory of Spacetime, Figure 1
Scale dependence of the length (*left*) and of the effective fractal dimension $D_F = \tau_F + 1$ (*right*) in the case of "inertial" scale laws (which are solutions of the simplest, first order scale differential equation). Toward the small scale one gets a scale-invariant law with constant fractal dimension, while the explicit scale-dependence is lost at scales larger than a transition scale λ

In order to find the solution of this equation, let us change the names of the constants as $\tau_F = -b$ and $\mathcal{L}_0 = a/\tau_F$, so that $a + b\mathcal{L} = -\tau_F(\mathcal{L} - \mathcal{L}_0)$. We obtain the equation

$$\frac{d\mathcal{L}}{\mathcal{L} - \mathcal{L}_0} = -\tau_F \, d\ln\varepsilon \,. \tag{5}$$

Its solution is (see Fig. 1)

$$\mathcal{L}(\varepsilon) = \mathcal{L}_0 \left\{ 1 + \left(\frac{\lambda}{\varepsilon}\right)^{\tau_F} \right\} \,, \tag{6}$$

where λ is an integration constant. This solution corresponds to a length measured on a fractal curve up to a given point. One can now generalize it to a variable length that also depends on the position characterized by the parameter s. One obtains

$$\mathcal{L}(s, \varepsilon) = \mathcal{L}_0(s) \left\{ 1 + \zeta(s)\left(\frac{\lambda}{\varepsilon}\right)^{\tau_F} \right\} \,, \tag{7}$$

in which, in the most general case, the exponent τ_F may itself be a variable depending on the position.

The same kind of result is obtained for the projections on a given axis of such a fractal length [54]. Let $X(s, \varepsilon)$ be one of these projections, it reads

$$X(s, \varepsilon) = x(s) \left\{ 1 + \zeta_x(s)\left(\frac{\lambda}{\varepsilon}\right)^{\tau_F} \right\} \,. \tag{8}$$

In this case $\zeta_x(s)$ becomes a highly fluctuating function which may be described by a stochastic variable, as can be seen in Fig. 2.

The important point here and for what follows is that the solution obtained is the sum of two terms (a classical-like, "differentiable part" and a nondifferentiable "frac-
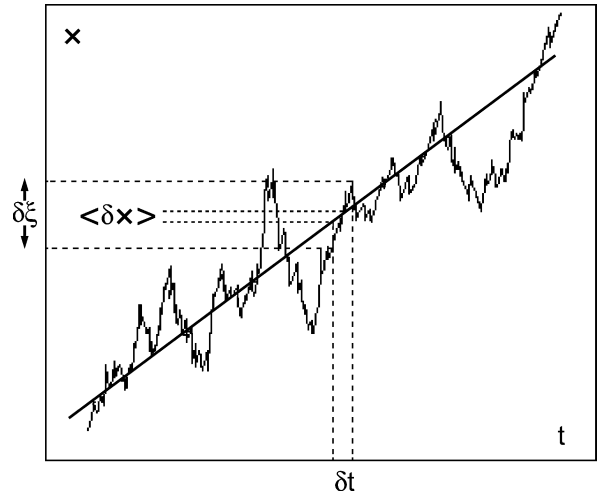
Fractals in the Quantum Theory of Spacetime, Figure 2
A fractal function. An example of such a fractal function is given by the projections of a fractal curve on Cartesian coordinates, as a function of a continuous and monotonous parameter (here the time t) which marks the position on the curve. The figure also exhibits the relation between space and time differential elements for such a fractal function, and compares the differentiable part δx and nondifferentiable part $\delta \xi$ of the space elementary displacement $\delta X = \delta x + \delta \xi$. While the differentiable coordinate variation $\delta x = \langle \delta X \rangle$ is of the same order as the time differential δt, the fractal fluctuation becomes much larger than δt when $\delta t \ll T$, where T is a transition time scale, and it depends on the fractal dimension D_F as $\delta \xi \propto \delta t^{1/D_F}$. Therefore the two contributions to the full differential displacement are related by the fractal law $\delta \xi^{D_F} \propto \delta x$, since δx and δt are differential elements of the same order

tal part") which is explicitly scale-dependent and tends to infinity when $\varepsilon \to 0$ [22,54]. By differentiating these two parts in the above projection, we obtain the differential

formulation of this essential result,

$$dX = dx + d\xi , \qquad (9)$$

where dx is a classical differential element, while $d\xi$ is a differential element of fractional order (see Fig. 2, in which the parameter s that characterizes the position on the fractal curve has been taken to be the time t). This relation plays a fundamental role in the subsequent developments of the theory.

Consider the case when τ_F is constant. In the asymptotic small scale regime, $\varepsilon \ll \lambda$, one obtains a power-law dependence on resolution which reads

$$\mathcal{L}(s, \varepsilon) = \mathcal{L}_0(s) \left(\frac{\lambda}{\varepsilon} \right)^{\tau_F} . \qquad (10)$$

In this expression we recognize the standard form of a self-similar fractal behavior with constant fractal dimension $D_F = 1 + \tau_F$, which has already been found to yield a fair description of many physical and biological systems [49]. Here the topological dimension is $D_T = 1$, since we deal with a length, but this can be easily generalized to surfaces ($D_T = 2$), volumes ($D_T = 3$), etc., according to the general relation $D_F = D_T + \tau_F$. The new feature here is that this result has been derived from a theoretical analysis based on first principles, instead of being postulated or deduced from a fit of observational data.

It should be noted that in the above expressions, the resolution is a length interval, $\varepsilon = \delta X$ defined along the fractal curve (or one of its projected coordinate). But one may also travel on the curve and measure its length on constant time intervals, then change the time scale. In this case the resolution ε is a time interval, $\varepsilon = \delta t$. Since they are related by the fundamental relation (see Fig. 2)

$$\delta X^{D_F} \sim \delta t , \qquad (11)$$

the fractal length depends on the time resolution as

$$X(s, \delta t) = X_0(s) \times \left(\frac{T}{\delta t} \right)^{1 - 1/D_F} . \qquad (12)$$

An example of the use of such a relation is Feynman's result according to which the mean square value of the velocity of a quantum mechanical particle is proportional to δt^{-1} (see p. 176 in [30]), which corresponds to a fractal dimension $D_F = 2$, as later recovered by Abbott and Wise [1] by using a space resolution.

More generally (in the usual case when $\varepsilon = \delta X$), following Mandelbrot, the scale exponent $\tau_F = D_F - D_T$ can be defined as the slope of the $(\ln \varepsilon, \ln \mathcal{L})$ curve, namely

$$\tau_F = \frac{d \ln \mathcal{L}}{d \ln(\lambda/\varepsilon)} . \qquad (13)$$

For a self-similar fractal such as that described by the fractal part of the above solution, this definition yields a constant value which is the exponent in Eq. (10). However, one can anticipate on the following, and use this definition to compute an "effective" or "local" fractal dimension, now variable, from the complete solution that includes the differentiable and the nondifferentiable parts, and therefore a transition to effective scale independence. Differentiating the logarithm of Eq. (16) yields an effective exponent given by

$$\tau_{\text{eff}} = \frac{\tau_F}{1 + (\varepsilon/\lambda)^{\tau_F}} . \qquad (14)$$

The effective fractal dimension $D_F = 1 + \tau_F$ therefore jumps from the nonfractal value $D_F = D_T = 1$ to its constant asymptotic value at the transition scale λ (see right part of Fig. 1).

Galilean Relativity of Scales

We can now check that the fractal part of such a law is compatible with the principle of relativity extended to scale transformations of the resolutions (i. e., with the principle of scale relativity). It reads $\mathcal{L} = \mathcal{L}_0(\lambda/\varepsilon)^{\tau_F}$ (Eq. 10), and it is therefore a law involving two variables ($\ln \mathcal{L}$ and τ_F) as a function of one parameter (ε) which, according to the relativistic view, characterizes the state of scale of the system (its relativity is apparent in the fact that we need another scale λ to define it by their ratio). More generally, all the following statements remain true for the complete scale law including the transition to scale-independence, by making the replacement of $\mathcal{L}$ by $\mathcal{L} - \mathcal{L}_0$. Note that, to be complete, we anticipate on what follows and consider a priori τ_F to be a variable, even if, in the simple law first considered here, it takes a constant value.

Let us take the logarithm of Eq. (10). It yields $\ln(\mathcal{L}/\mathcal{L}_0) = \tau_F \ln(\lambda/\varepsilon)$. The two quantities $\ln \mathcal{L}$ and τ_F then transform under a finite scale transformation $\varepsilon \to \varepsilon' = \rho \varepsilon$ as

$$\ln \frac{\mathcal{L}(\varepsilon')}{\mathcal{L}_0} = \ln \frac{\mathcal{L}(\varepsilon)}{\mathcal{L}_0} - \tau_F \ln \rho , \qquad (15)$$

and to be complete,

$$\tau_F' = \tau_F . \qquad (16)$$

These transformations have exactly the same mathematical structure as the Galilean group of motion transformation (applied here to scale rather than motion), which reads

$$x' = x - t\, v, \quad t' = t . \qquad (17)$$

This is confirmed by the dilation composition law, $\varepsilon \to \varepsilon' \to \varepsilon''$, which reads

$$\ln \frac{\varepsilon''}{\varepsilon} = \ln \frac{\varepsilon'}{\varepsilon} + \ln \frac{\varepsilon''}{\varepsilon'} , \qquad (18)$$

and is therefore similar to the law of composition of velocities between three reference systems K, K' and K'',

$$V''(K''/K) = V(K'/K) + V'(K''/K') . \qquad (19)$$

Since the Galileo group of motion transformations is known to be the simplest group that implements the principle of relativity, the same is true for scale transformations.

It is important to realize that this is more than a simple analogy: the same physical problem is set in both cases, and is therefore solved under similar mathematical structures (since the logarithm transforms what would have been a multiplicative group into an additive group). Indeed, in both cases, it is equivalent to finding the transformation law of a position variable (X for motion in a Cartesian system of coordinates, $\ln \mathcal{L}$ for scales in a fractal system of coordinates) under a change of the state of the coordinate system (change of velocity V for motion and of resolution $\ln \rho$ for scale), knowing that these state variables are defined only in a relative way. Namely, V is the relative velocity between the reference systems K and K', and ρ is the relative scale: note that ε and ε' have indeed disappeared in the transformation law, only their ratio remains. This remark establishes the status of resolutions as (relative) "scale velocities" and of the scale exponent τ_F as a "scale time".

Recall finally that since the Galilean group of motion is only a limiting case of the more general Lorentz group, a similar generalization is expected in the case of scale transformations, which we shall briefly consider in Sect. "Special Scale-Relativity".

Breaking of Scale Invariance

The standard self-similar fractal laws can be derived from the scale relativity approach. However, it is important to note that Eq. (16) provides us with another fundamental result, as shown in Fig. 1. Namely, it also contains a spontaneous breaking of the scale symmetry. Indeed, it is characterized by the existence of a transition from a fractal to a non-fractal behavior at scales larger than some transition scale λ. The existence of such a breaking of scale invariance is also a fundamental feature of many natural systems, which remains, in most cases, misunderstood.

The advantage of the way it is derived here is that it appears as a natural, spontaneous, but only effective symmetry breaking, since it does not affect the underlying scale

symmetry. Indeed, the obtained solution is the sum of two terms, the scale-independent contribution (differentiable part), and the explicitly scale-dependent and divergent contribution (fractal part). At large scales the scaling part becomes dominated by the classical part, but it is still underlying even though it is hidden. There is therefore an apparent symmetry breaking (see Fig. 1), though the underlying scale symmetry actually remains unbroken.

The origin of this transition is, once again, to be found in relativity (namely, in the relativity of position and motion). Indeed, if one starts from a strictly scale-invariant law without any transition, $\mathcal{L} = \mathcal{L}_0(\lambda/\varepsilon)^{\tau_F}$, then adds a translation in standard position space ($\mathcal{L} \to \mathcal{L} + \mathcal{L}_1$), one obtains

$$\mathcal{L}' = \mathcal{L}_1 + \mathcal{L}_0 \left(\frac{\lambda}{\varepsilon}\right)^{\tau_F} = \mathcal{L}_1 \left\{1 + \left(\frac{\lambda_1}{\varepsilon}\right)^{\tau_F}\right\} . \qquad (20)$$

Therefore one recovers the broken solution (that corresponds to the constant $a \neq 0$ in the initial scale differential equation). This solution is now asymptotically scale-dependent (in a scale-invariant way) only at small scales, and becomes independent of scale at large scales, beyond some relative transition λ_1 which is partly determined by the translation itself.

Multiple Scale Transitions

Multiple transitions can be obtained by a simple generalization of the above result [58]. Still considering a perturbative approach and taking the Taylor expansion of the differential equation $\mathrm{d}\mathcal{L}/\mathrm{d}\ln\varepsilon = \beta(\mathcal{L})$, but now to the second order of the expansion, one obtains the equation

$$\frac{\mathrm{d}\mathcal{L}}{\mathrm{d}\ln\varepsilon} = a + b\mathcal{L} + c\mathcal{L}^2 + \cdots . \qquad (21)$$

One of its solutions, which generalizes that of Eq. (5), describes a scaling behavior which is broken toward both the small and large scales, as observed in most real fractal systems,

$$\mathcal{L} = \mathcal{L}_0 \left(\frac{1 + (\lambda_0/\varepsilon)^{\tau_F}}{1 + (\lambda_1/\varepsilon)^{\tau_F}}\right) . \qquad (22)$$

Due to the non-linearity of the β function, there are now two transition scales in such a law. Indeed,

*when $\varepsilon < \lambda_1 < \lambda_0$, one has $(\lambda_0/\varepsilon) \gg 1$ and $(\lambda_1/\varepsilon) \gg 1$, so that $\mathcal{L} = \mathcal{L}_0(\lambda_0/\lambda_1)^{\tau_F} \approx$ cst, independent of scale;

*when $\lambda_1 < \varepsilon < \lambda_0$, one has $(\lambda_0/\varepsilon) \gg 1$ but $(\lambda_1/\varepsilon) \ll 1$, so that the denominator disappears, and one recovers the previous pure scaling law $\mathcal{L} = \mathcal{L}_0 (\lambda_0/\varepsilon)^{\tau_F}$;

*when $\lambda_1 < \lambda_0 < \varepsilon$, one has $(\lambda_0/\varepsilon) \ll 1$ and $(\lambda_1/\varepsilon) \ll 1$, so that $\mathcal{L} = \mathcal{L}_0 =$ cst, independent of scale.

Scale Relativity Versus Scale Invariance

Let us briefly be more specific about the way in which the scale relativity viewpoint differs from scaling or simple scale invariance. In the standard concept of scale invariance, one considers scale transformations of the coordinate,

$$X \to X' = q \times X, \tag{23}$$

then one looks for the effect of such a transformation on some function $f(X)$. It is scaling when

$$f(qX) = q^\alpha \times f(X). \tag{24}$$

The scale relativity approach involves a more profound level of description, since the coordinate X is now explicitly resolution-dependent, i. e. $X = X(\varepsilon)$. Therefore we now look for a scale transformation of the resolution,

$$\varepsilon \to \varepsilon' = \rho \, \varepsilon, \tag{25}$$

which implies a scale transformation of the position variable that takes in the self-similar case the form

$$X(\rho \, \varepsilon) = \rho^{-\tau_F} X(\varepsilon). \tag{26}$$

But now the scale factor on the variable has a physical meaning which goes beyond a trivial change of units. It corresponds to a coordinate measured at two different resolutions on a fractal curve of fractal dimension $D_F = 1 + \tau_F$, and one can obtain a scaling function of a fractal coordinate,

$$f(\rho^{-\tau_F} X) = \rho^{-\alpha \tau_F} \times f(X). \tag{27}$$

In other words, there are now three levels of transformation in the scale relativity framework (the resolution, the variable, and its function) instead of only two in the usual conception of scaling.

Generalized Scale Laws

Discrete Scale Invariance, Complex Dimension, and Log-Periodic Behavior Fluctuations with respect to pure scale invariance are potentially important, namely the log-periodic correction to power laws that is provided, e. g., by complex exponents or complex fractal dimensions. It has been shown that such a behavior provides a very satisfactory and possibly predictive model of the time evolution of many critical systems, including earthquakes and market crashes ([79] and references therein). More recently, it has been applied to the analysis of major event chronology of the evolutionary tree of life [14,65,66], of

human development [11] and of the main economic crisis of western and precolumbian civilizations [36,37,40,65].

One can recover log-periodic corrections to self-similar power laws through the requirement of covariance (i. e., of form invariance of equations) applied to scale differential equations [58]. Consider a scale-dependent function $\mathcal{L}(\varepsilon)$. In the applications to temporal evolution quoted above, the scale variable is identified with the time interval $|t - t_c|$, where t_c is the date of a crisis. Assume that $\mathcal{L}$ satisfies a first order differential equation,

$$\frac{d\mathcal{L}}{d \ln \varepsilon} - \nu \mathcal{L} = 0, \tag{28}$$

whose solution is a pure power law $\mathcal{L}(\varepsilon) \propto \varepsilon^\nu$ (cf Sect. "Self-Similar Fractals as Solutions of a First Order Scale Differential Equation"). Now looking for corrections to this law, one remarks that simply incorporating a complex value of the exponent ν would lead to large log-periodic fluctuations rather than to a controllable correction to the power law. So let us assume that the right-hand side of Eq. (28) actually differs from zero:

$$\frac{d\mathcal{L}}{d \ln \varepsilon} - \nu \mathcal{L} = \chi. \tag{29}$$

We can now apply the scale covariance principle and require that the new function χ be a solution of an equation which keeps the same form as the initial equation,

$$\frac{d\chi}{d \ln \varepsilon} - \nu' \chi = 0. \tag{30}$$

Setting $\nu' = \nu + \eta$, we find that $\mathcal{L}$ must be a solution of a second-order equation:

$$\frac{d^2 \mathcal{L}}{(d \ln \varepsilon)^2} - (2\nu + \eta)\frac{d\mathcal{L}}{d \ln \varepsilon} + \nu(\nu + \eta)\mathcal{L} = 0. \tag{31}$$

The solution reads $\mathcal{L}(\varepsilon) = a\varepsilon^\nu(1 + b\varepsilon^\eta)$, and finally, the choice of an imaginary exponent $\eta = i\omega$ yields a solution whose real part includes a log-periodic correction:

$$\mathcal{L}(\varepsilon) = a\varepsilon^\nu [1 + b \cos(\omega \ln \varepsilon)]. \tag{32}$$

As previously recalled in Sect. "Breaking of Scale Invariance," adding a constant term (a translation) provides a transition to scale independence at large scales (see Fig. 3).

Lagrangian Approach to Scale Laws In order to obtain physically relevant generalizations of the above simplest (scale-invariant) laws, a Lagrangian approach can be used in scale space analogously to using it to derive the laws of
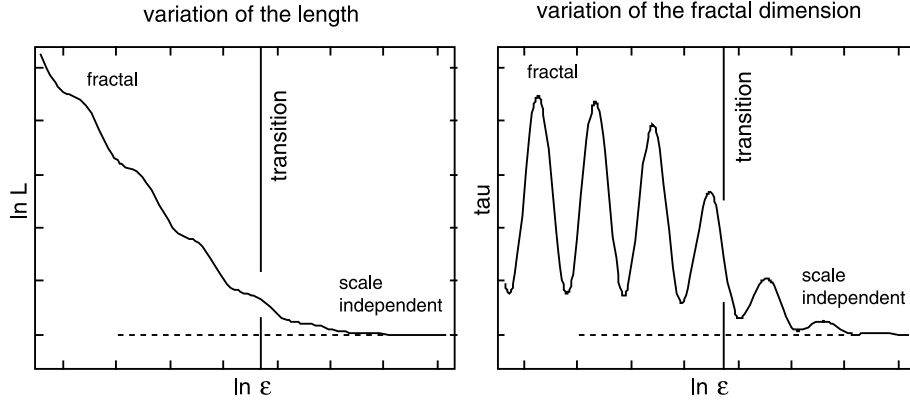
Fractals in the Quantum Theory of Spacetime, Figure 3
Scale dependence of the length $\mathcal{L}$ of a fractal curve and of the effective "scale time" $\tau_F = D_F - D_T$ (fractal dimension minus topological dimension) in the case of a log-periodic behavior with fractal/non-fractal transition at scale λ, which reads $\mathcal{L}(\varepsilon) = \mathcal{L}_0\{1 + (\lambda/\varepsilon)^\nu \exp[b\cos(\omega\ln(\varepsilon/\lambda))]\}$

motion, leading to reversal of the definition and meaning of the variables [58].

This reversal is analogous to that achieved by Galileo concerning the laws of motion. Indeed, from the Aristotle viewpoint, "time is the measure of motion". In the same way, the fractal dimension, in its standard (Mandelbrot's) acception, is defined from the topological measure of the fractal object (length of a curve, area of a surface, etc.) and resolution, namely (see Eq. 13)

$$t = \frac{x}{v} \quad \leftrightarrow \quad \tau_F = D_F - D_T = \frac{d\ln\mathcal{L}}{d\ln(\lambda/\varepsilon)} \,. \tag{33}$$

In the case mainly considered here, when $\mathcal{L}$ represents a length (i. e., more generally, a fractal coordinate), the topological dimension is $D_T = 1$ so that $\tau_F = D_F - 1$. With Galileo, time becomes a primary variable, and the velocity is deduced from space and time, which are therefore treated on the same footing, in terms of a space-time (even though the Galilean space-time remains degenerate because of the implicitly assumed infinite velocity of light).

In analogy, the scale exponent $\tau_F = D_F - 1$ becomes in this new representation a primary variable that plays, for scale laws, the same role as played by time in motion laws (it is called "djinn" in some publications which therefore introduce a five-dimensional 'space-time-djinn' combining the four fractal fluctuations and the scale time).

Carrying on the analogy, in the same way that the velocity is the derivative of position with respect to time, $v = dx/dt$, we expect the derivative of $\ln\mathcal{L}$ with respect to scale time τ_F to be a "scale velocity". Consider as reference the self-similar case, that reads $\ln\mathcal{L} = \tau_F\ln(\lambda/\varepsilon)$. Deriving with respect to τ_F, now considered as a variable,

yields $d\ln\mathcal{L}/d\tau_F = \ln(\lambda/\varepsilon)$, i. e., the logarithm of resolution. By extension, one assumes that this scale velocity provides a new general definition of resolution even in more general situations, namely,

$$\mathbb{V} = \ln\left(\frac{\lambda}{\varepsilon}\right) = \frac{d\ln\mathcal{L}}{d\tau_F} \,. \tag{34}$$

One can now introduce a scale Lagrange function $\widetilde{L}(\ln\mathcal{L}, \mathbb{V}, \tau_F)$, from which a scale action is constructed:

$$\widetilde{S} = \int_{\tau_1}^{\tau_2} \widetilde{L}(\ln\mathcal{L}, \mathbb{V}, \tau_F)d\tau_F \,. \tag{35}$$

The application of the action principle yields a scale Euler–Lagrange equation which reads

$$\frac{d}{d\tau_F}\frac{\partial\widetilde{L}}{\partial\mathbb{V}} = \frac{\partial\widetilde{L}}{\partial\ln\mathcal{L}} \,. \tag{36}$$

One can now verify that in the free case, i. e., in the absence of any "scale force" (i. e., $\partial\tilde{L}/\partial\ln\mathcal{L} = 0$), one recovers the standard fractal laws derived hereabove. Indeed, in this case the Euler–Lagrange equation becomes

$$\partial\tilde{L}/\partial\mathbb{V} = \text{const} \quad \Rightarrow \quad \mathbb{V} = \text{const.} \tag{37}$$

which is the equivalent for scale of what inertia is for motion. Still in analogy with motion laws, the simplest possible form for the Lagrange function is a quadratic dependence on the scale velocity, (i. e., $\tilde{L} \propto \mathbb{V}^2$). The constancy of $\mathbb{V} = \ln(\lambda/\varepsilon)$ means that it is independent of the scale time τ_F. Equation (34) can therefore be integrated to give the usual power law behavior, $\mathcal{L} = \mathcal{L}_0(\lambda/\varepsilon)^{\tau_F}$, as expected.

But this reversed viewpoint also has several advantages which allow a full implementation of the principle of scale relativity:

(i) The scale time τ_F is given the status of a fifth dimension and the logarithm of the resolution $\mathbb{V} = \ln(\lambda/\varepsilon)$, is given the status of a scale velocity (see Eq. 34). This is in accordance with its scale-relativistic definition, in which it characterizes the state of scale of the reference system, in the same way as the velocity $v = dx/dt$ characterizes its state of motion.

(ii) This allows one to generalize the formalism to the case of four independent space-time resolutions, $\mathbb{V}^\mu = \ln(\lambda^\mu/\varepsilon^\mu) = d \ln \mathcal{L}^\mu/d\tau_F$.

(iii) Scale laws more general than the simplest self-similar ones can be derived from more general scale Lagrangians [57,58] involving "scale accelerations" $\Gamma = d^2 \ln \mathcal{L}/d\tau_F^2 = d \ln(\lambda/\varepsilon)/d\tau_F$, as we shall see in what follows.

Note however that there is also a shortcoming in this approach. Contrary to the case of motion laws, in which time is always flowing toward the future (except possibly in elementary particle physics at very small time scales), the variation of the scale time may be non-monotonic, as exemplified by the previous case of log-periodicity. Therefore this Lagrangian approach is restricted to monotonous variations of the fractal dimension, or, more generally, to scale intervals on which it varies in a monotonous way.

Scale Dynamics The previous discussion indicates that the scale invariant behavior corresponds to freedom (i. e. scale force-free behavior) in the framework of a scale physics. However, in the same way as there are forces in nature that imply departure from inertial, rectilinear uniform motion, we expect most natural fractal systems to also present distortions in their scale behavior with respect to pure scale invariance. This implies taking non-linearity in the scale space into account. Such distorsions may be, as a first step, attributed to the effect of the dynamics of scale ("scale dynamics"), i. e., of a "scale field", but it must be clear from the very beginning of the description that they are of geometric nature (in analogy with the Newtonian interpretation of gravitation as the result of a force, which has later been understood from Einstein's general relativity theory as a manifestation of the curved geometry of space-time).

In this case the Lagrange scale-equation takes the form of Newton's equation of dynamics,

$$F = \mu \, \frac{d^2 \ln \mathcal{L}}{d\tau_F^2} \, , \tag{38}$$

where μ is a "scale mass", which measures how the system resists to the scale force, and where $\Gamma = d^2 \ln \mathcal{L}/d\tau_F^2 = d \ln(\lambda/\varepsilon)/d\tau_F$ is the scale acceleration.

In this framework one can therefore attempt to define generic, scale-dynamical behaviours which could be common to very different systems, as corresponding to a given form of the scale force.

Constant Scale Force A typical example is the case of a constant scale force. Setting $G = F/\mu$, the potential reads $\varphi = G \ln \mathcal{L}$, analogous to the potential of a constant force f in space, which is $\varphi = -fx$, since the force is $-\partial\varphi/\partial x = f$. The scale differential equation is

$$\frac{d^2 \ln \mathcal{L}}{d\tau_F^2} = G \, . \tag{39}$$

It can be easily integrated. A first integration yields $d \ln \mathcal{L}/d\tau_F = G\tau_F + \mathbb{V}_0$, where $\mathbb{V}_0$ is a constant. Then a second integration yields a parabolic solution (which is the equivalent for scale laws of parabolic motion in a constant field),

$$\mathbb{V} = \mathbb{V}_0 + G\tau_F \, ; \quad \ln \mathcal{L} = \ln \mathcal{L}_0 + \mathbb{V}_0\tau_F + \frac{1}{2}G \, \tau_F^2 \, , \tag{40}$$

where $\mathbb{V} = d \ln \mathcal{L}/d\tau_F = \ln(\lambda/\varepsilon)$.

However the physical meaning of this result is not clear under this form. This is due to the fact that, while in the case of motion laws we search for the evolution of the system with time, in the case of scale laws we search for the dependence of the system on resolution, which is the directly measured observable. Since the reference scale λ is arbitrary, the variables can be re-defined in such a way that $\mathbb{V}_0 = 0$, i. e., $\lambda = \lambda_0$. Indeed, from Eq. (40) one gets $\tau_F = (\mathbb{V} - \mathbb{V}_0)/G = [\ln(\lambda/\varepsilon) - \ln(\lambda/\lambda_0)]/G = \ln(\lambda_0/\varepsilon)/G$. Then one obtains

$$\tau_F = \frac{1}{G} \ln\left(\frac{\lambda_0}{\varepsilon}\right), \quad \ln\left(\frac{\mathcal{L}}{\mathcal{L}_0}\right) = \frac{1}{2G} \ln^2\left(\frac{\lambda_0}{\varepsilon}\right) \, . \tag{41}$$

The scale time τ_F becomes a linear function of resolution (the same being true, as a consequence, of the fractal dimension $D_F = 1 + \tau_F$), and the $(\ln \mathcal{L}, \ln \varepsilon)$ relation is now parabolic instead of linear (see Fig. 4 and compare to Fig. 1). Note that, as in previous cases, we have considered here only the small scale asymptotic behavior, and that we can once again easily generalize this result by including a transition to scale-independence at large scale. This is simply achieved by replacing $\mathcal{L}$ by $(\mathcal{L} - \mathcal{L}_0)$ in every equations.

There are several physical situations where, after careful examination of the data, the power-law models were clearly rejected since no constant slope could be defined in

variation of the length | variation of the fractal dimension
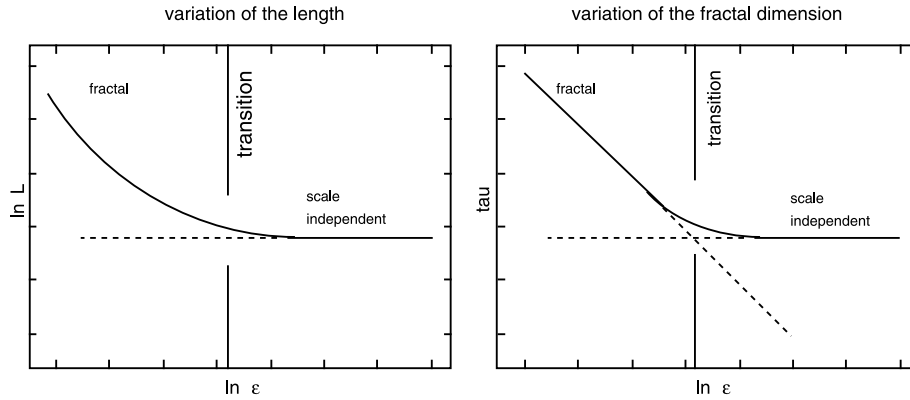
Fractals in the Quantum Theory of Spacetime, Figure 4
Scale dependence of the length of a fractal curve $\ln \mathcal{L}$ and of its effective fractal dimension ($D_F = D_T + \tau_F$, where D_T is the topological dimension) in the case of a constant scale force, with an additional fractal to non-fractal transition

the $(\log \mathcal{L}, \log \varepsilon)$ plane. In the several cases where a clear curvature appears in this plane, e. g., turbulence [26], sandpiles [9], fractured surfaces in solid mechanics [10], the physics could come under such a scale-dynamical description. In these cases it might be of interest to identify and study the scale force responsible for the scale distortion (i. e., for the deviation from standard scaling).

Special Scale-Relativity

Let us close this section about the derivation of scale laws of increasing complexity by coming back to the question of finding the general laws of scale transformations that meet the principle of scale relativity [53]. It has been shown in Sect. "Galilean Relativity of Scales" that the standard self-similar fractal laws come under a Galilean group of scale transformations. However, the Galilean relativity group is known, for motion laws, to be only a degenerate form of the Lorentz group. It has been proved that a similar result holds for scale laws [53,54].

The problem of finding the linear transformation laws of fields in a scale transformation $\mathbb{V} = \ln \rho \ (\varepsilon \to \varepsilon')$ amounts to finding four quantities, $a(\mathbb{V}), b(\mathbb{V}), c(\mathbb{V})$, and $d(\mathbb{V})$, such that

$$
\begin{aligned}
\ln \frac{\mathcal{L}'}{\mathcal{L}_0} &= a(\mathbb{V}) \ln \frac{\mathcal{L}}{\mathcal{L}_0} + b(\mathbb{V}) \, \tau_F \,, \\
\tau_{F'} &= c(\mathbb{V}) \ln \frac{\mathcal{L}}{\mathcal{L}_0} + d(\mathbb{V}) \tau_F \,.
\end{aligned}
\tag{42}
$$

Set in this way, it immediately appears that the current 'scale-invariant' scale transformation law of the standard form (Eq. 8), given by $a = 1, b = \mathbb{V}, c = 0$ and $d = 1$, corresponds to a Galilean group.

This is also clear from the law of composition of dilatations, $\varepsilon \to \varepsilon' \to \varepsilon''$, which has a simple additive form,

$$
\mathbb{V}'' = \mathbb{V} + \mathbb{V}' \,.
\tag{43}
$$

However the general solution to the 'special relativity problem' (namely, find a, b, c and d from the principle of relativity) is the Lorentz group [47,53]. This result has led to the suggestion of replacing the standard law of dilatation, $\varepsilon \to \varepsilon' = \varrho \times \varepsilon$ by a new Lorentzian relation, namely, for $\varepsilon < \lambda_0$ and $\varepsilon' < \lambda_0$

$$
\ln \frac{\varepsilon'}{\lambda_0} = \frac{\ln(\varepsilon/\lambda_0) + \ln \varrho}{1 + \ln \varrho \, \ln(\varepsilon/\lambda_0)/\ln^2(\Lambda/\lambda_0)} \,.
\tag{44}
$$

This relation introduces a fundamental length scale Λ, which is naturally identified, towards the small scales, with the Planck length (currently $1.6160(11) \times 10^{-35}$ m) [53],

$$
\Lambda = l_{\mathbb{P}} = (\hbar G/c^3)^{1/2} \,,
\tag{45}
$$

and toward the large scales (for $\varepsilon > \lambda_0$ and $\varepsilon' > \lambda_0$) with the scale of the cosmological constant, $\mathbb{L} = \Lambda^{-1/2}$ (see Chap. 7.1 in [54]).

As one can see from Eq. (44), if one starts from the scale $\varepsilon = \Lambda$ and apply any dilatation or contraction ϱ, one obtains again the scale $\varepsilon' = \Lambda$, whatever the initial value of λ_0. In other words, Λ can be interpreted as a limiting lower (or upper) length-scale, which is impassable and invariant under dilatations and contractions.

Concerning the length measured along a fractal coordinate which was previously scale-dependent as $\ln(\mathcal{L}/\mathcal{L}_0) = \tau_0 \ln(\lambda_0/\varepsilon)$ for $\varepsilon < \lambda_0$, it becomes in the new framework, in the simplified case when one starts from the
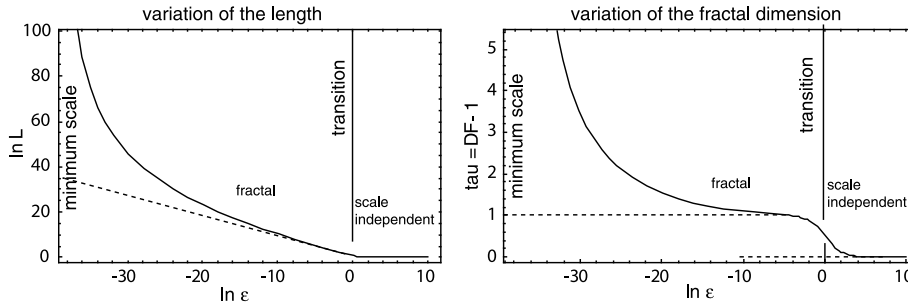
Fractals in the Quantum Theory of Spacetime, Figure 5
Scale dependence of the logarithm of the length and of the effective fractal dimension, $D_F = 1 + \tau_F$, in the case of scale-relativistic Lorentzian scale laws including a transition to scale independence toward large scales. The constant $\mathbb{C}$ has been taken to be $\mathbb{C} = 4\pi^2 \approx 39.478$, which is a fundamental value for scale ratios in elementary particle physics in the scale relativity framework [53,54,63], while the effective fractal dimension jumps from $D_F = 1$ to $D_F = 2$ at the transition, then increases without any limit toward small scales

reference scale $\mathcal{L}_0$ (see Fig. 5)

$$\ln \frac{\mathcal{L}}{\mathcal{L}_0} = \frac{\tau_0 \, \ln(\lambda_0/\varepsilon)}{\sqrt{1 - \ln^2(\lambda_0/\varepsilon)/\ln^2(\lambda_0/\Lambda)}} \ . \tag{46}$$

The main new feature of scale relativity with respect to the previous fractal or scale-invariant approaches is that the scale exponent τ_F and the fractal dimension $D_F = 1 + \tau_F$, which were previously constant ($D_F = 2$, $\tau_F = 1$), are now explicitly varying with scale (see Fig. 5), following the law (given once again in the simplified case when we start from the reference scale $\mathcal{L}_0$):

$$\tau_F(\varepsilon) = \frac{\tau_0}{\sqrt{1 - \ln^2(\lambda_0/\varepsilon)/\ln^2(\lambda_0/\Lambda)}} \ . \tag{47}$$

Under this form, the scale covariance is explicit, since one keeps a power law form for the length variation, $\mathcal{L} = \mathcal{L}_0(\lambda/\varepsilon)^{\tau_F(\varepsilon)}$, but now in terms of a variable fractal dimension.

For a more complete development of special relativity, including its implications regarding new conservative quantities and applications in elementary particle physics and cosmology, see [53,54,56,63].

The question of the nature of space-time geometry at the Planck scale is a subject of intense work (see, e. g., [3,46] and references therein). This is a central question for practically all theoretical attempts, including noncommutative geometry [15,16], supersymmetry, and superstring theories [35,75] – for which the compactification scale is close to the Planck scale – and particularly for the theory of quantum gravity. Indeed, the development of loop quantum gravity by Rovelli and Smolin [76] led to the conclusion that the Planck scale could be a quantized minimal scale in Nature, involving also a quantization of surfaces and volumes [77].

Over the last years, there has also been significant research effort aimed at the development of a 'Doubly-Special-Relativity' [2] (see a review in [3]), according to which the laws of physics involve a fundamental velocity scale c and a fundamental minimum length scale L_p, identified with the Planck length.

The concept of a new relativity in which the Planck length-scale would become a minimum invariant length is exactly the founding idea of the special scale relativity theory [53], which has been incorporated in other attempts of extended relativity theories [12,13]. But, despite the similarity of aim and analysis, the main difference between the 'Doubly-Special-Relativity' approach and the scale relativity one is that the question of defining an invariant length-scale is considered in the scale relativity/fractal space-time theory as coming under a relativity of scales. Therefore the new group to be constructed is a multiplicative group, that becomes additive only when working with the logarithms of scale ratios, which are definitely the physically relevant scale variables, as one can show by applying the Gell-Mann-Levy method to the construction of the dilation operator (see Sect. "Fractal Coordinate and Differential Dilation Operator").

From Fractal Space to Nonrelativistic Quantum Mechanics

The first step in the construction of a theory of the quantum space-time from fractal and nondifferentiable geometry, which has been described in the previous sections, has consisted of finding the laws of explicit scale dependence at a given "point" or "instant" (under their new fractal definition).

The next step, which will now be considered, amounts to writing the equation of motion in such a fractal space(-

time) in terms of a geodesic equation. As we shall see, after integration this equation takes the form of a Schrödinger equation (and of the Klein-Gordon and Dirac equations in the relativistic case). This result, first obtained in Ref. [54], has later been confirmed by many subsequent physical [22,25,56,57] and mathematical works, in particular by Cresson and Ben Adda [6,7,17,19] and Jumarie [41,42,43,44], including attempts of generalizations using the tool of the fractional integro-differential calculus [7,21,44].

In what follows, we consider only the simplest case of fractal laws, namely, those characterized by a constant fractal dimension. The various generalized scale laws considered in the previous section lead to new possible generalizations of quantum mechanics [56,63].

Critical Fractal Dimension 2

Moreover, we simplify again the description by considering only the case $D_F = 2$. Indeed, the nondifferentiability and fractality of space implies that the paths are random walks of the Markovian type, which corresponds to such a fractal dimension. This choice is also justified by Feynman's result [30], according to which the typical paths of quantum particles (those which contribute mainly to the path integral) are nondifferentiable and of fractal dimension $D_F = 2$ [1]. The case $D_F \neq 2$, which yields generalizations to standard quantum mechanics has also been studied in detail (see [56,63] and references therein). This study shows that $D_F = 2$ plays a critical role in the theory, since it suppresses the explicit scale dependence in the motion (Schrödinger) equation – but this dependence remains hidden and reappears through, e. g., the Heisenberg relations and the explicit dependence of measurement results on the resolution of the measurement apparatus.

Let us start from the result of the previous section, according to which the solution of a first order scale differential equation reads for $D_F = 2$, after differentiation and reintroduction of the indices,

$$dX^\mu = dx^\mu + d\xi^\mu = v^\mu ds + \zeta^\mu \sqrt{\lambda_c \, ds} \, , \qquad (48)$$

where λ_c is a length scale which must be introduced for dimensional reasons and which, as we shall see, generalizes the Compton length. The ζ^μ are dimensionless highly fluctuating functions. Due to their highly erratic character, we can replace them by stochastic variables such that $\langle \zeta^\mu \rangle = 0$, $\langle (\zeta^0)^2 \rangle = -1$ and $\langle (\zeta^k)^2 \rangle = 1$ ($k = 1$ to 3). The mean is taken here on a purely mathematic probability law which can be fully general, since the final result does not depend on its choice.

Metric of a Fractal Space-Time

Now one can also write the fractal fluctuations in terms of the coordinate differentials, $d\xi^\mu = \zeta^\mu \sqrt{\lambda^\mu \, dx^\mu}$. The identification of this expression with that of Eq. (3) leads one to recover the Einstein-de Broglie length and time scales,

$$\lambda_x = \frac{\lambda_c}{dx/ds} = \frac{\hbar}{p_x} \, , \quad \tau = \frac{\lambda_c}{dt/ds} = \frac{\hbar}{E} \, . \qquad (49)$$

Let us now assume that the large scale (classical) behavior is given by the Riemannian metric potentials $g_{\mu\nu}(x, y, z, t)$. The invariant proper time dS along a geodesic in terms of the complete differential elements $dX^\mu = dx^\mu + d\xi^\mu$

$$dS^2 = g_{\mu\nu} dX^\mu dX^\nu = g_{\mu\nu}(dx^\mu + d\xi^\mu)(dx^\nu + d\xi^\nu). \quad (50)$$

Now replacing the $d\xi$'s by their expression, one obtains a fractal metric [54,68]. Its two-dimensional and diagonal expression, neglecting the terms of the zero mean (in order to simplify its writing) reads

$$dS^2 = g_{00}(x, t) \left(1 + \zeta_0^2 \frac{\tau_F}{dt} \right) c^2 dt^2$$
$$- g_{11}(x, t) \left(1 + \zeta_1^2 \frac{\lambda_x}{dx} \right) dx^2 \, . \qquad (51)$$

We therefore obtain generalized fractal metric potentials which are divergent and explicitly dependent on the coordinate differential elements [52,54]. Another equivalent way to understand this metric consists in remarking that it is no longer only quadratic in the space-time differental elements, but that it also contains them in a linear way.

As a consequence, the curvature is also explicitly scale-dependent and divergent when the scale intervals tend to zero. This property ensures the fundamentally non-Riemannian character of a fractal space-time, as well as the possibility to characterize it in an intrinsic way. Indeed, such a characterization, which is a necessary condition for defining a space in a genuine way, can be easily made by measuring the curvature at smaller and smaller scales. While the curvature vanishes by definition toward the small scales in Gauss-Riemann geometry, a fractal space can be characterized from the interior by the verification of the divergence toward small scales of curvature, and therefore of physical quantities like energy and momentum.

Now the expression of this divergence is nothing but the Heisenberg relations themselves, which therefore acquire in this framework the status of a fundamental geometric test of the fractality of space-time [52,54,69].

Geodesics of a Fractal Space-Time

The next step in such a geometric approach consists of the identifying the wave-particles with fractal space-time geodesics. Any measurement is interpreted as a selection of the geodesics bundle linked to the interaction with the measurement apparatus (that depends on its resolution) and/or to the information known about it (for example, the which-way-information in a two-slit experiment [56].

The three main consequences of nondifferentiability are:

(i) The number of fractal geodesics is infinite. This leads one to adopt a generalized statistical fluid-like description where the velocity $V^\mu(s)$ is replaced by a scale-dependent velocity field $V^\mu[X^\mu(s, ds), s, ds]$.

(ii) There is a breaking of the reflexion invariance of the differential element ds. Indeed, in terms of fractal functions $f(s, ds)$, two derivatives are defined,

$$
\begin{aligned}
X'_+(s, ds) &= \frac{X(s + ds, ds) - X(s, ds)}{ds} , \\
X'_-(s, ds) &= \frac{X(s, ds) - X(s - ds, ds)}{ds} ,
\end{aligned}
\tag{52}
$$

which transform into each other under the reflection ($ds \leftrightarrow -ds$), and which have a priori no reason to be equal. This leads to a fundamental two-valuedness of the velocity field.

(iii) The geodesics are themselves fractal curves of fractal dimension $D_F = 2$ [30].

This means that one defines two divergent fractal velocity fields, $V_+[x(s, ds), s, ds] = v_+[x(s), s] + w_+[x(s, ds), s, ds]$ and $V_-[x(s, ds), s, ds] = v_-[x(s), s] + w_-[x(s, ds), s, ds]$, which can be decomposed in terms of differentiable parts v_+ and v_-, and of fractal parts w_+ and w_-. Note that, contrary to other attempts such as Nelson's stochastic quantum mechanics which introduces forward and backward velocities [51] (and which has been later disproved [34,80]), the two velocities are here both forward, since they do not correspond to a reversal of the time coordinate, but of the time differential element now considered as an independent variable.

More generally, we define two differentiable parts of derivatives d_+/ds and d_-/ds, which, when they are applied to x^μ, yield the differential parts of the velocity fields, $v_+^\mu = d_+ x^\mu/ds$ and $v_-^\mu = d_- x^\mu/ds$.

Covariant Total Derivative

Let us first consider the non-relativistic case. It corresponds to a three-dimensional fractal space, without fractal time, in which the invariant ds is therefore identified with the time differential element dt. One describes the elementary displacements dX^k, where $k = 1, 2, 3$, on the geodesics of a nondifferentiable fractal space in terms of the sum of the two terms (omitting the indices for simplicity) $dX_\pm = d_\pm x + d\xi_\pm$, where dx represents the differentiable part and $d\xi$ the fractal (nondifferentiable) part, defined as

$$
d_\pm x = v_\pm dt , \quad d\xi_\pm = \zeta_\pm \sqrt{2\mathcal{D}} \, dt^{1/2} .
\tag{53}
$$

Here $\zeta_\pm$ are stochastic dimensionless variables such that $\langle \zeta_\pm \rangle = 0$ and $\langle \zeta_\pm^2 \rangle = 1$, and $\mathcal{D}$ is a parameter that generalizes the Compton scale (namely, $\mathcal{D} = \hbar/2m$ in the case of standard quantum mechanics) up to the fundamental constant $c/2$. The two time derivatives are then combined in terms of a complex total time derivative operator [54],

$$
\frac{\widehat{d}}{dt} = \frac{1}{2} \left(\frac{d_+}{dt} + \frac{d_-}{dt} \right) - \frac{i}{2} \left(\frac{d_+}{dt} - \frac{d_-}{dt} \right).
\tag{54}
$$

Applying this operator to the differentiable part of the position vector yields a complex velocity

$$
\mathcal{V} = \frac{\widehat{d}}{dt} x(t) = V - iU = \frac{v_+ + v_-}{2} - i \frac{v_+ - v_-}{2} .
\tag{55}
$$

In order to find the expression for the complex time derivative operator, let us first calculate the derivative of a scalar function f. Since the fractal dimension is 2, one needs to go to the second order of expansion. For one variable it reads

$$
\frac{df}{dt} = \frac{\partial f}{\partial t} + \frac{\partial f}{\partial X} \frac{dX}{dt} + \frac{1}{2} \frac{\partial^2 f}{\partial X^2} \frac{dX^2}{dt} .
\tag{56}
$$

Generalizing this process to three dimensions is straightforward.

Let us now take the stochastic mean of this expression, i. e., we take the mean on the stochastic variables $\zeta_\pm$ which appear in the definition of the fractal fluctuation $d\xi_\pm$. By definition, since $dX = dx + d\xi$ and $\langle d\xi \rangle = 0$, we have $\langle dX \rangle = dx$, so that the second term is reduced (in 3 dimensions) to $v.\nabla f$. Now concerning the term dX^2/dt, it is infinitesimal and therefore it would not be taken into account in the standard differentiable case. But in the nondifferentiable case considered here, the mean square fluctuation is non-vanishing and of order dt, namely, $\langle d\xi^2 \rangle = 2\mathcal{D}dt$, so that the last term of Eq. (56) amounts in three dimensions to a Laplacian operator. One obtains, respectively for the $(+)$ and $(-)$ processes,

$$
\frac{d_\pm f}{dt} = \left(\frac{\partial}{\partial t} + v_\pm.\nabla \pm \mathcal{D}\Delta \right) f .
\tag{57}
$$

Finally, by combining these two derivatives in terms of the complex derivative of Eq. (54), it reads [54]

$$\frac{\widehat{\mathrm{d}}}{\mathrm{d}t} = \frac{\partial}{\partial t} + \mathcal{V}.\nabla - i\mathcal{D}\Delta . \tag{58}$$

Under this form, this expression is not fully covariant [74], since it involves derivatives of the second order, so that its Leibniz rule is a linear combination of the first and second order Leibniz rules. By introducing the velocity operator [61]

$$\widehat{\mathcal{V}} = \mathcal{V} - i\mathcal{D}\nabla , \tag{59}$$

it may be given a fully covariant expression,

$$\frac{\widehat{\mathrm{d}}}{\mathrm{d}t} = \frac{\partial}{\partial t} + \widehat{\mathcal{V}}.\nabla . \tag{60}$$

Under this form it satisfies the first order Leibniz rule for partial derivatives.

We shall now see that $\widehat{\mathrm{d}}/\mathrm{d}t$ plays the role of a "covariant derivative operator" (in analogy with the covariant derivative of general relativity). Namely, one may write in its terms the equation of physics in a nondifferentiable space under a strongly covariant form identical to the differentiable case.

Complex Action and Momentum

The steps of construction of classical mechanics can now be followed, but in terms of complex and scale dependent quantities. One defines a Lagrange function that keeps its usual form, $\mathcal{L}(x, \mathcal{V}, t)$, but which is now complex. Then one defines a generalized complex action

$$S = \int_{t_1}^{t_2} \mathcal{L}(x, \mathcal{V}, t)\mathrm{d}t . \tag{61}$$

Generalized Euler–Lagrange equations that keep their standard form in terms of the new complex variables can be derived from this action [22,54], namely

$$\frac{\widehat{\mathrm{d}}}{\mathrm{d}t}\frac{\partial L}{\partial \mathcal{V}} - \frac{\partial L}{\partial x} = 0 . \tag{62}$$

From the homogeneity of space and Noether's theorem, one defines a generalized complex momentum given by the same form as in classical mechanics as

$$\mathcal{P} = \frac{\partial \mathcal{L}}{\partial \mathcal{V}} . \tag{63}$$

If the action is now considered as a function of the upper limit of integration in Eq. (61), the variation of the action from a trajectory to another nearby trajectory yields a generalization of another well-known relation of classical mechanics,

$$\mathcal{P} = \nabla S . \tag{64}$$

Motion Equation

As an example, consider the case of a single particle in an external scalar field with potential energy ϕ (but the method can be applied to any situation described by a Lagrange function). The Lagrange function, $L = \frac{1}{2}mv^2 - \phi$, is generalized as $\mathcal{L}(x, \mathcal{V}, t) = \frac{1}{2}m\mathcal{V}^2 - \phi$. The Euler–Lagrange equations then keep the form of Newton's fundamental equation of dynamics $F = m\,\mathrm{d}v/\mathrm{d}t$, namely,

$$m\frac{\widehat{\mathrm{d}}}{\mathrm{d}t}\mathcal{V} = -\nabla\phi , \tag{65}$$

which is now written in terms of complex variables and complex operators.

In the case when there is no external field ($\phi = 0$), the covariance is explicit, since Eq. (65) takes the free form of the equation of inertial motion, i. e., of a geodesic equation,

$$\frac{\widehat{\mathrm{d}}}{\mathrm{d}t}\mathcal{V} = 0 . \tag{66}$$

This is analogous to Einstein's general relativity, where the equivalence principle leads to the covariant equation of the motion of a free particle in the form of an inertial motion (geodesic) equation $Du_\mu/\mathrm{d}s = 0$, written in terms of the general-relativistic covariant derivative D, of the four-vector u_μ and of the proper time differential $\mathrm{d}s$.

The covariance induced by the effects of the nondifferentiable geometry leads to an analogous transformation of the equation of motions, which, as we show below, become after integration the Schrödinger equation, which can therefore be considered as the integral of a geodesic equation in a fractal space.

In the one-particle case the complex momentum $\mathcal{P}$ reads

$$\mathcal{P} = m\mathcal{V} , \tag{67}$$

so that from Eq. (64) the complex velocity $\mathcal{V}$ appears as a gradient, namely the gradient of the complex action

$$\mathcal{V} = \nabla S/m . \tag{68}$$

Wave Function

Up to now the various concepts and variables used were of a classical type (space, geodesics, velocity fields), even if

they were generalized to the fractal and nondifferentiable, explicitly scale-dependent case whose essence is fundamentally not classical.

We shall now make essential changes of variable, that transform this apparently classical-like tool to quantum mechanical tools (without any hidden parameter or new degree of freedom). The complex wave function ψ is introduced as simply another expression for the complex action S, by making the transformation

$$\psi = e^{iS/S_0} . \tag{69}$$

Note that, despite its apparent form, this expression involves a phase and a modulus since S is complex. The factor S_0 has the dimension of an action (i. e., an angular momentum) and must be introduced because S is dimensioned while the phase should be dimensionless. When this formalism is applied to standard quantum mechanics, S_0 is nothing but the fundamental constant $\hbar$. As a consequence, since

$$S = -iS_0 \ln \psi , \tag{70}$$

one finds that the function ψ is related to the complex velocity appearing in Eq. (68) as follows

$$\mathcal{V} = -i \frac{S_0}{m} \nabla \ln \psi . \tag{71}$$

This expression is the fundamental relation that connects the two description tools while giving the meaning of the wave function in the new framework. Namely, it is defined here as a velocity potential for the velocity field of the infinite family of geodesics of the fractal space. Because of nondifferentiability, the set of geodesics that defines a 'particle' in this framework is fundamentally non-local. It can easily be generalized to a multiple particle situation (in particular to entangled states) which are described by a single wave function ψ, from which the various velocity fields of the subsets of the geodesic bundle are derived as $\mathcal{V}_k = -i (S_0/m_k) \nabla_k \ln \psi$, where k is an index for each particle. The indistinguishability of identical particles naturally follows from the fact that the 'particles' are identified with the geodesics themselves, i. e., with an infinite ensemble of purely geometric curves. In this description there is no longer any point-mass with 'internal' properties which would follow a 'trajectory', since the various properties of the particle – energy, momentum, mass, spin, charge (see next sections) – can be derived from the geometric properties of the geodesic fluid itself.

Correspondence Principle

Since we have $\mathcal{P} = -iS_0 \nabla \ln \psi = -iS_0(\nabla \psi)/\psi$, we obtain the equality [54]

$$\mathcal{P}\psi = -i\hbar \nabla \psi \tag{72}$$

in the standard quantum mechanical case $S_0 = \hbar$, which establishes a correspondence between the classical momentum p, which is the real part of the complex momentum in the classical limit, and the operator $-i\hbar\nabla$.

This result is generalizable to other variables, in particular to the Hamiltonian. Indeed, a strongly covariant form of the Hamiltonian can be obtained by using the fully covariant form of the covariant derivative operator given by Eq. (60). With this tool, the expression of the relation between the complex action and the complex Lagrange function reads

$$\mathcal{L} = \frac{\widehat{\mathrm{d}} S}{\mathrm{d}t} = \frac{\partial S}{\partial t} + \widehat{\mathcal{V}}.\nabla S . \tag{73}$$

Since $\mathcal{P} = \nabla S$ and $\mathcal{H} = -\partial S/\partial t$, one obtains for the generalized complex Hamilton function the same form it has in classical mechanics, namely [63,67],

$$\mathcal{H} = \widehat{\mathcal{V}}.\mathcal{P} - \mathcal{L} . \tag{74}$$

After expanding the velocity operator, one obtains $\mathcal{H} = \mathcal{V}.\mathcal{P} - i\mathcal{D}\nabla.\mathcal{P} - \mathcal{L}$, which includes an additional term [74], whose origin is now understood as an expression of nondifferentiability and strong covariance.

Schrödinger Equation and Compton Relation

The next step of the construction amounts to writing the fundamental equation of dynamics Eq. (65) in terms of the function ψ. It takes the form

$$iS_0 \frac{\widehat{\mathrm{d}}}{\mathrm{d}t}(\nabla \ln \psi) = \nabla \phi . \tag{75}$$

As we shall now see, this equation can be integrated in a general way in the form of a Schrödinger equation. Replacing $\widehat{\mathrm{d}}/\mathrm{d}t$ and $\mathcal{V}$ by their expressions yields

$$\nabla \Phi = iS_0 \left[\frac{\partial}{\partial t} \nabla \ln \psi - i \left\{ \frac{S_0}{m} (\nabla \ln \psi.\nabla)(\nabla \ln \psi) + \mathcal{D}\Delta(\nabla \ln \psi) \right\} \right] . \tag{76}$$

This equation may be simplified thanks to the identity [54],

$$\nabla \left(\frac{\Delta \psi}{\psi} \right) = 2(\nabla \ln \psi.\nabla)(\nabla \ln \psi) + \Delta(\nabla \ln \psi) . \tag{77}$$

We recognize, in the right-hand side of Eq. (77), the two terms of Eq. (76), which were respectively a factor of S_0/m and $\mathcal{D}$. This leads to the definition of the wave function as

$$\psi = e^{iS/2m\mathcal{D}} , \qquad (78)$$

which means that the arbitrary parameter S_0 (which is identified with the constant $\hbar$ in standard QM) is now linked to the fractal fluctuation parameter by the relation

$$S_0 = 2m\mathcal{D} . \qquad (79)$$

This relation (which can actually be proved instead of simply being set as a simplifying choice, see [62,67]) is actually a generalization of the Compton relation, since the geometric parameter $\mathcal{D} = \langle d\xi^2\rangle/2dt$ can be written in terms of a length scale as $\mathcal{D} = \lambda c/2$, so that, when $S_0 = \hbar$, it becomes $\lambda = \hbar/mc$. But a geometric meaning is now given to the Compton length (and therefore to the inertial mass of the particle) in the fractal space-time framework.

The fundamental equation of dynamics now reads

$$\nabla\phi = 2im\mathcal{D}\left[\frac{\partial}{\partial t}\nabla\ln\psi - i\{2\mathcal{D}(\nabla\ln\psi.\nabla)(\nabla\ln\psi) \right.$$
$$\left. + \mathcal{D}\Delta(\nabla\ln\psi)\}\right] . \quad (80)$$

Using the above remarkable identity and the fact that $\partial/\partial t$ and ∇ commute, it becomes

$$-\frac{\nabla\phi}{m} = -2\mathcal{D}\nabla\left\{i\frac{\partial}{\partial t}\ln\psi + \mathcal{D}\frac{\Delta\psi}{\psi}\right\} . \qquad (81)$$

The full equation becomes a gradient,

$$\nabla\left\{\frac{\phi}{m} - 2\mathcal{D}\nabla\left(\frac{i\,\partial\psi/\partial t + \mathcal{D}\Delta\psi}{\psi}\right)\right\} = 0 \qquad (82)$$

and it can be easily integrated to finally obtain a generalized Schrödinger equation [54]

$$\mathcal{D}^2\Delta\psi + i\mathcal{D}\frac{\partial}{\partial t}\psi - \frac{\phi}{2m}\psi = 0 , \qquad (83)$$

up to an arbitrary phase factor which may be set to zero by a suitable choice of the ψ phase. One recovers the standard Schrödinger equation of quantum mechanics for the particular case when $\mathcal{D} = \hbar/2m$.

Von Neumann's and Born's Postulates

In the framework described here, "particles" are identified with the various geometric properties of fractal space(-time) geodesics. In such an interpretation, a measurement

(and more generally any knowledge about the system) amounts to selecting the sub-set of the geodesics family which only contains the geodesics with the geometric properties corresponding to the measurement result. Therefore, just after the measurement, the system is in the state given by the measurement result, which is precisely the von Neumann postulate of quantum mechanics.

The Born postulate can also be inferred from the scale-relativity construction [22,62,67]. Indeed, the probability for the particle to be found at a given position must be proportional to the density of the geodesics fluid at this point. The velocity and the density of the fluid are expected to be solutions of a Euler and continuity system of four equations, with four unknowns, (ρ, V_x, V_y, V_z).

Now, by separating the real and imaginary parts of the Schrödinger equation, setting $\psi = \sqrt{P}\times e^{i\theta}$ and using a mixed representation (P, V), where $V = \{V_x, V_y, V_z\}$, one precisely obtains such a standard system of fluid dynamics equations, namely,

$$\left(\frac{\partial}{\partial t} + V\cdot\nabla\right)V = -\nabla\left(\phi - 2\mathcal{D}^2\frac{\Delta\sqrt{P}}{\sqrt{P}}\right) ,$$
$$\frac{\partial P}{\partial t} + \text{div}(PV) = 0 . \qquad (84)$$

This allows one to unequivocally identify $P = |\psi|^2$ with the probability density of the geodesics and therefore with the probability of presence of the 'particle'. Moreover,

$$Q = -2\mathcal{D}^2\frac{\Delta\sqrt{P}}{\sqrt{P}} \qquad (85)$$

can be interpreted as the new potential which is expected to emerge from the fractal geometry, in analogy with the identification of the gravitational field as a manifestation of the curved geometry in Einstein's general relativity. This result is supported by numerical simulations, in which the probability density is obtained directly from the distribution of geodesics without writing the Schrödinger equation [39,63].

Nondifferentiable Wave Function

In more recent works, instead of taking only the differentiable part of the velocity field into account, one constructs the covariant derivative and the wave function in terms of the full velocity field, including its divergent nondifferentiable part of zero mean [59,62]. As we shall briefly see now, this still leads to the standard form of the Schrödinger equation. This means that, in the scale relativity framework, one expects the Schrödinger equation to have fractal and nondifferentiable solutions. This result

agrees with a similar conclusion by Berry [8] and Hall [38], but it is considered here as a direct manifestation of the nondifferentiability of space itself.

Consider the full complex velocity field, including its differentiable and nondifferentiable parts,

$$\tilde{\mathcal{V}} = \mathcal{V} + \mathcal{W} = \left(\frac{v_+ + v_-}{2} - i \frac{v_+ - v_-}{2} \right)$$
$$+ \left(\frac{w_+ + w_-}{2} - i \frac{w_+ - w_-}{2} \right). \quad (86)$$

It is related to a full complex action $\tilde{S}$ and a full wave function $\tilde{\psi}$ as

$$\tilde{\mathcal{V}} = \mathcal{V} + \mathcal{W} = \nabla \tilde{S}/m = -2i\mathcal{D}\nabla \ln \tilde{\psi} . \quad (87)$$

Under the standard point of view, the complex fluctuation $\mathcal{W}$ is infinite and therefore $\nabla \ln \tilde{\psi}$ is undefined, so that Eq. (87) would be meaningless. In the scale relativity approach, on the contrary, this equation keeps a mathematical and physical meaning, in terms of fractal functions, which are explicitly dependent on the scale interval dt and divergent when $dt \to 0$.

After some calculations [59,62], one finds that the covariant derivative built from the total process (including the differentiable and nondifferentiable divergent terms) is finally

$$\frac{\widehat{d}}{dt} = \frac{\partial}{\partial t} + \tilde{\mathcal{V}}.\nabla - i\mathcal{D}\Delta . \quad (88)$$

The subsequent steps of the derivation of the Schrödinger equation are unchanged (now in terms of scale-dependent fractal functions), so that one obtains

$$\mathcal{D}^2\Delta\tilde{\psi} + i\mathcal{D}\frac{\partial\tilde{\psi}}{\partial t} - \frac{\phi}{2m}\tilde{\psi} = 0 , \quad (89)$$

where $\tilde{\psi}$ can now be a nondifferentiable and fractal function. The research of such a behavior in laboratory experiments is an interesting new challenge for quantum physics.

One may finally stress the fact that this result is obtained by accounting for all the terms, differentiable (dx) and nondifferentiable ($d\xi$), in the description of the elementary displacements in a nondifferentiable space, and that it does not depend at all on the probability distribution of the stochastic variables $d\xi$, about which no hypothesis is needed. This means that the description is fully general, and that the effect on motion of a nondifferentiable space(-time) amounts to a fundamental indeterminism, i. e., to a total loss of information about the past path which will in all cases lead to the QM description.

From Fractal Space-Time to Relativistic Quantum Mechanics

All these results can be generalized to relativistic quantum mechanics, which corresponds in the scale relativity framework to a full fractal space-time. This yields, as a first step, the Klein–Gordon equation [22,55,56].

Then an account of the new two-valuedness of the velocity allows one to suggest a geometric origin for the spin and to obtain the Dirac equation [22]. Indeed, the total derivative of a physical quantity also involves partial derivatives with respect to the space variables, $\partial/\partial x^\mu$. From the very definition of derivatives, the discrete symmetry under the reflection $dx^\mu \leftrightarrow -dx^\mu$ is also broken. Since, at this level of description, one should also account for parity as in the standard quantum theory, this leads to introduce a bi-quaternionic velocity field [22], in terms of which the Dirac bispinor wave function can be constructed.

The successive steps that lead to the Dirac equation naturally generalize the Schrödinger case. One introduces a biquaternionic generalization of the covariant derivative that keeps the same form as in the complex case, namely,

$$\frac{\widehat{d}}{ds} = \mathcal{V}^\nu\partial_\nu + i\frac{\lambda}{2}\partial^\nu\partial_\nu , \quad (90)$$

where $\lambda = 2\mathcal{D}/c$. The biquaternionic velocity field is related to the biquaternionic (i. e., bispinorial) wave function, by

$$\mathcal{V}_\mu = i\frac{S_0}{m}\psi^{-1}\partial_\mu\psi . \quad (91)$$

This is the relativistic expression of the fundamental relation between the scale relativity tools and the quantum mechanical tools of description. It gives a geometric interpretation to the wave function, which is, in this framework, a manifestation of the geodesic fluid and of its associated fractal velocity field. The covariance principle allows us to write the equation of motion under the form of a geodesic differential equation,

$$\frac{\widehat{d}\,\mathcal{V}_\mu}{ds} = 0 . \quad (92)$$

After some calculations, this equation can be integrated and factorized, and one finally derives the Dirac equation [22],

$$\frac{1}{c}\frac{\partial\psi}{\partial t} = -\alpha^k\frac{\partial\psi}{\partial x^k} - i\frac{mc}{\hbar}\beta\psi . \quad (93)$$

Finally it is easy to recover the Pauli equation and Pauli spinors as nonrelativistic approximations of the Dirac equation and Dirac bispinors [23].

Gauge Fields as Manifestations of Fractal Geometry

General Scale Transformations and Gauge Fields

Finally, let us briefly recall the main steps of applying of the scale relativity principles to the foundation of gauge theories, in the Abelian [55,56] and non-Abelian [63,68] cases.

This application is based on a general description of the internal fractal structures of the "particle" (identified with the geodesics of a nondifferentiable space-time) in terms of scale variables $\eta_{\alpha\beta}(x, y, z, t) = \varrho_{\alpha\beta} \, \varepsilon_\alpha \, \varepsilon_\beta$ whose true nature is tensorial, since it involves resolutions that may be different for the four space-time coordinates and may be correlated. This resolution tensor (similar to a covariance error matrix) generalizes the single resolution variable ε. Moreover, one considers a more profound level of description in which the scale variables may now be functions of the coordinates. Namely, the internal structures of the geodesics may vary from place to place and during the time evolution, in agreement with the non-absolute character of the scale space. This generalization amounts to the construction of a 'general scale relativity' theory.

We assume, for simplicity of the writing, that the two tensorial indices can be gathered under one common index. We therefore write the scale variables under the simplified form $\eta_{\alpha_1\alpha_2} = \eta_\alpha$, $\alpha = 1$ to $N = n(n + 1)/2$, where n is the number of space-time dimensions ($n = 3$, $N = 6$ for fractal space, $n = 4$, $N = 10$ for fractal space-time and $n = 5$, $N = 15$ in the special scale relativity case where one treats the djinn (scale-time τ_F) as a fifth dimension [53]).

Let us consider infinitesimal scale transformations. The transformation law on the η_α can be written in a linear way as

$$\eta'_\alpha = \eta_\alpha + \delta\eta_\alpha = (\delta_{\alpha\beta} + \delta\theta_{\alpha\beta}) \, \eta^\beta \,, \tag{94}$$

where $\delta_{\alpha\beta}$ is the Kronecker symbol. Let us now assume that the η_α's are functions of the standard space-time coordinates. This leads one to define a new scale-covariant derivative by writing the total variation of the resolution variables as the sum of the inertial variation, described by the covariant derivative, and of the new geometric contribution, namely,

$$d\eta_\alpha = D\eta_\alpha - \eta^\beta \, \delta\theta_{\alpha\beta} = D\eta_\alpha - \eta^\beta \, W^\mu_{\alpha\beta} \, dx_\mu \,. \tag{95}$$

This covariant derivative is similar to that of general relativity, i. e., it amounts to the subtraction of the new geometric part in order to only keep the inertial part (for which the motion equation will therefore take a geodesical,

free-like form). This is different from the case of the previous quantum-covariant derivative, which includes the effects of nondifferentiability by adding new terms in the total derivative.

In this new situation we are led to introduce "gauge field potentials" $W^\mu_{\alpha\beta}$ that enter naturally in the geometrical framework of Eq. (95). These potentials are linked to the scale transformations as follows:

$$\delta\theta_{\alpha\beta} = W^\mu_{\alpha\beta} \, dx_\mu \,. \tag{96}$$

But one should keep in mind, when using this expression, that these potentials find their origin in a covariant derivative process and are therefore not gradients.

Generalized Charges

After having written the transformation law of the basic variables (the η_α's), one now needs to describe how various physical quantities transform under these η_α transformations. These new transformation laws are expected to depend on the nature of the objects they transform (e. g., vectors, tensors, spinors, etc.), which implies a jump to group representations.

We anticipate the existence of charges (which are fully constructed herebelow) by generalizing the relation (91) to multiplets between the velocity field and the wave function. In this case the multivalued velocity becomes a biquaternionic matrix,

$$\mathcal{V}^\mu_{jk} = i\lambda \, \psi_j^{-1} \partial^\mu \psi_k \,. \tag{97}$$

The biquaternionic, and therefore non-commutative, nature of the wave function [15], which is equivalent to Dirac bispinors, plays an essential role here. Indeed, the general structure of Yang–Mills theories and the correct construction of non-Abelian charges can be obtained thanks to this result [68].

The action also becomes a tensorial biquaternionic quantity,

$$dS_{jk} = dS_{jk}\big(x^\mu, \mathcal{V}^\mu_{jk}, \eta_\alpha\big) \,, \tag{98}$$

and, in the absence of a field (free particle) it is linked to the generalized velocity (and therefore to the spinor multiplet) by the relation

$$\partial^\mu S_{jk} = -mc \, \mathcal{V}^\mu_{jk} = -i\hbar \, \psi_j^{-1} \partial^\mu \psi_k \,. \tag{99}$$

Now, in the presence of a field (i. e., when the second-order effects of the fractal geometry appearing in the right hand side of Eq. (95) are included), using the complete expression for $\partial^\mu \eta_\alpha$,

$$\partial^\mu \eta_\alpha = D^\mu \eta_\alpha - W^\mu_{\alpha\beta} \, \eta^\beta \,, \tag{100}$$

one obtains a non-Abelian relation,

$$\partial^\mu S_{jk} = D^\mu S_{jk} - \eta^\beta \frac{\partial S_{jk}}{\partial \eta_\alpha} W^\mu_{\alpha\beta} . \tag{101}$$

This finally leads to the definition of a general group of scale transformations whose generators are

$$T^{\alpha\beta} = \eta^\beta \partial^\alpha \tag{102}$$

(where we use the compact notation $\partial^\alpha = \partial/\partial\eta_\alpha$), yielding the generalized charges,

$$\frac{\tilde{g}}{c} t^{\alpha\beta}_{jk} = \eta^\beta \frac{\partial S_{jk}}{\partial \eta_\alpha} . \tag{103}$$

This unified group is submitted to a unitarity condition, since, when it is applied to the wave functions, $\psi\psi^\dagger$ must be conserved. Knowing that α, β represent two indices each, this is a large group – at least SO(10) – that contains the electroweak theory [33,78,81] and the standard model U(1) × SU(2) × SU(3) and its simplest grand unified extension SU(5) [31,32] as a subset (see [53,54] for solutions in the special scale relativity framework to the problems encountered by SU(5) GUTs).

As it is shown in more detail in Ref. [68], the various ingredients of Yang–Mills theories (gauge covariant derivative, gauge invariance, charges, potentials, fields, etc.) may subsequently be recovered in such a framework, but they now have a first principle and geometric scale-relativistic foundation.

Future Directions

In this contribution, we have recalled the main steps that lead to a new foundation of quantum mechanics and of gauge fields on the principle of relativity itself (once it includes scale transformations of the reference system), and on the generalized geometry of space-time which is naturally associated with such a principle, namely, nondifferentiability and fractality.

For this purpose, two covariant derivatives have been constructed, which account for the nondifferentiable and fractal geometry of space-time, and which allow one to write the equations of motion as geodesics equations. After a change of variable, these equations finally take the form of the quantum mechanical and quantum field equations.

Let us conclude by listing some original features of the scale relativity approach which could lead to experimental tests of the theory and/or to new experiments in the future [63,67]: (i) nondifferentiable and fractal solutions of the Schrödinger equation; (ii) zero particle interference in a Young slit experiment; (iii) possible breaking of the Born postulate for a possible effective kinetic energy operator $\widehat{T} \neq -(\hbar^2/2m)\Delta$ [67]; (iv) underlying quantum behavior in the classical domain, at scales far larger than the de Broglie scale [67]; (v) macroscopic systems described by a Schrödinger-type mechanics based on a generalized macroscopic parameter $\mathcal{D} \neq \hbar/2m$ (see Chap. 7.2 in [54] and [24,57]); (vi) applications to cosmology [60]; (vii) applications to life sciences and other sciences [4,64,65], etc.

Bibliography

Primary Literature

1. Abbott LF, Wise MB (1981) Am J Phys 49:37
2. Amelino-Camelia G (2001) Phys Lett (B)510:255
3. Amelino-Camelia G (2002) Int J Mod Phys (D)11:1643
4. Auffray C, Nottale L (2007) Progr Biophys Mol Bio 97:79
5. Ben Adda F, Cresson J (2000) CR Acad Sci Paris 330:261
6. Ben Adda F, Cresson J (2004) Chaos Solit Fractals 19:1323
7. Ben Adda F, Cresson J (2005) Appl Math Comput 161:323
8. Berry MV (1996) J Phys A: Math Gen 29:6617
9. Cafiero R, Loreto V, Pietronero L, Vespignani A, Zapperi S (1995) Europhys Lett 29:111
10. Carpinteri A, Chiaia B (1996) Chaos Solit Fractals 7:1343
11. Cash R, Chaline J, Nottale L, Grou P (2002) CR Biologies 325:585
12. Castro C (1997) Found Ph ys Lett 10:273
13. Castro C, Granik A (2000) Chaos Solit Fractals 11:2167
14. Chaline J, Nottale L, Grou P (1999) C R Acad Sci Paris 328:717
15. Connes A (1994) Noncommutative Geometry. Academic Press, New York
16. Connes A, Douglas MR, Schwarz A J High Energy Phys 02:003 (hep-th/9711162)
17. Cresson J (2001) Mémoire d'habilitation à diriger des recherches. Université de Franche-Comté, Besançon
18. Cresson J (2002) Chaos Solit Fractals 14:553
19. Cresson J (2003) J Math Phys 44:4907
20. Cresson J (2006) Int J Geometric Methods in Mod Phys 3(7)
21. Cresson J (2007) J Math Phys 48:033504
22. Célérier MN, Nottale L (2004) J Phys A: Math Gen 37:931
23. Célérier MN, Nottale L (2006) J Phys A: Math Gen 39:12565
24. da Rocha D, Nottale L (2003) Chaos Solit Fractals 16:565
25. Dubois D (2000) In: Proceedings of CASYS'1999, 3rd International Conference on Computing Anticipatory Systems, Liège, Belgium, Am. Institute of Physics Conference Proceedings 517:417
26. Dubrulle B, Graner F, Sornette D (eds) (1997) In: Dubrulle B, Graner F, Sornette D (eds) Scale invariance and beyond, Proceedings of Les Houches school, EDP Sciences, Les Ullis/Springer, Berlin, New York, p 275
27. El Naschie MS (1992) Chaos Solit Fractals 2:211
28. El Naschie MS Chaos Solit Fractals 11:2391
29. El Naschie MS, Rössler O, Prigogine I (eds) (1995) Quantum mechanics, diffusion and chaotic fractals. Pergamon, New York
30. Feynman RP, Hibbs AR (1965) Quantum mechanics and path integrals. MacGraw-Hill, New York
31. Georgi H, Glashow SL (1974) Phys Rev Lett 32:438
32. Georgi H, Quinn HR, Weinberg S (1974) Phys Rev Lett 33:451
33. Glashow SL (1961) Nucl Phys 22:579
34. Grabert H, Hänggi P, Talkner P (1979) Phys Rev A(19):2440

35. Green MB, Schwarz JH, Witten E (1987) Superstring Theory, vol 2. Cambridge University Press,
36. Grou P (1987) L'aventure économique. L'Harmattan, Paris
37. Grou P, Nottale L, Chaline J (2004) In: Zona Arqueologica, Miscelanea en homenaje a Emiliano Aguirre, IV Arqueologia, 230, Museo Arquelogico Regional, Madrid
38. Hall MJW (2004) J Phys A: Math Gen 37:9549
39. Hermann R (1997) J Phys A: Math Gen 30:3967
40. Johansen A, Sornette D (2001) Physica A(294):465
41. Jumarie G (2001) Int J Mod Phys A(16):5061
42. Jumarie G (2006) Chaos Solit Fractals 28:1285
43. Jumarie G (2006) Comput Math 51:1367
44. Jumarie G (2007) Phys Lett A 363:5
45. Kröger H (2000) Phys Rep 323:81
46. Laperashvili LV, Ryzhikh DA (2001) arXiv: hep-th/0110127 (Institute for Theoretical and Experimental Physics, Moscow)
47. Levy-Leblond JM (1976) Am J Phys 44:271
48. Losa G, Merlini D, Nonnenmacher T, Weibel E (eds) Fractals in biology and medicine. vol 3. Proceedings of Fractal 2000 Third International Symposium, Birkhäuser
49. Mandelbrot B (1982) The fractal geometry of nature. Freeman, San Francisco
50. McKeon DGC, Ord GN (1992) Phys Rev Lett 69:3
51. Nelson E (1966) Phys Rev 150:1079
52. Nottale L (1989) Int J Mod Phys A(4):5047
53. Nottale L (1992) Int J Mod Phys A(7):4899
54. Nottale L (1993) Fractal space-time and microphysics: Towards a theory of scale relativity. World Scientific, Singapore
55. Nottale L (1994) In: Relativity in general, (Spanish Relativity Meeting (1993)), edited Alonso JD, Paramo ML (eds), Editions Frontières, Paris, p 121
56. Nottale L (1996) Chaos Solit Fractals 7:877
57. Nottale L (1997) Astron Astrophys 327:867
58. Nottale L (1997) In: Scale invariance and beyond, Proceedings of Les Houches school, Dubrulle B, Graner F, Sornette D (eds) EDP Sciences, Les Ullis/Springer, Berlin, New York, p 249
59. Nottale L (1999) Chaos Solit Fractals 10:459
60. Nottale L (2003) Chaos Solit Fractals 16:539
61. Nottale L (2004) American Institute of Physics Conference Proceedings 718:68
62. Nottale L (2008) Proceedings of 7th International Colloquium on Clifford Algebra and their applications, 19–29 May 2005, Toulouse, Advances in Applied Clifford Algebras (in press)
63. Nottale L (2008) The theory of scale relativity. (submitted)
64. Nottale L, Auffray C (2007) Progr Biophys Mol Bio 97:115
65. Nottale L, Chaline J, Grou P (2000) Les arbres de l'évolution: Univers, Vie, Sociétés. Hachette, Paris, 379 pp
66. Nottale L, Chaline J, Grou P (2002) In: Fractals in biology and medicine, vol 3. Proceedings of Fractal (2000) Third International Symposium, Losa G, Merlini D, Nonnenmacher T, Weibel E (eds), Birkhäuser, p 247
67. Nottale L, Célérier MN (2008) J Phys A 40:14471
68. Nottale L, Célérier MN, Lehner T (2006) J Math Phys 47:032303
69. Nottale L, Schneider J (1984) J Math Phys 25:1296
70. Novak M (ed) (1998) Fractals and beyond: Complexities in the sciences, Proceedings of the Fractal 98 conference, World Scientific
71. Ord GN (1983) J Phys A: Math Gen 16:1869
72. Ord GN (1996) Ann Phys 250:51
73. Ord GN, Galtieri JA (2002) Phys Rev Lett 1989:250403
74. Pissondes JC (1999) J Phys A: Math Gen 32:2871
75. Polchinski J (1998) String theories. Cambridge University Press, Cambridge
76. Rovelli C, Smolin L (1988) Phys Rev Lett 61:1155
77. Rovelli C, Smolin L (1995) Phys Rev D(52):5743
78. Salam A (1968) Elementary particle theory. Svartholm N (ed). Almquist & Wiksells, Stockholm
79. Sornette D (1998) Phys Rep 297:239
80. Wang MS, Liang WK (1993) Phys Rev D(48):1875
81. Weinberg S (1967) Phys Rev Lett 19:1264

Books and Reviews

Georgi H (1999) Lie Algebras in particle physics. Perseus books, Reading, Massachusetts

Landau L, Lifchitz E (1970) Theoretical physics, 10 volumes, Mir, Moscow

Lichtenberg AJ, Lieberman MA (1983) Regular and stochastic motion. Springer, New York

Lorentz HA, Einstein A, Minkowski H, Weyl H (1923) The principle of relativity. Dover, New York

Mandelbrot B (1975) Les objets fractals. Flammarion, Paris

Misner CW, Thorne KS, Wheeler JA (1973) Gravitation. Freeman, San Francisco

Peebles J (1980) The large-scale structure of the universe. Princeton University Press, Princeton

Rovelli C (2004) Quantum gravity. Cambridge Universty Press, Cambridge

Weinberg S (1972) Gravitation and cosmology. Wiley, New York

Fractal Structures in Condensed Matter Physics

Tsuneyoshi Nakayama
Toyota Physical and Chemical Research Institute,
Nagakute, Japan

Article Outline

Glossary

Anomalous diffusion It is well known that the mean-square displacement $\langle r^2(t) \rangle$ of a diffusing particle on a uniform system is proportional to the time t such

as $\langle r^2(t) \rangle \sim t$. This is called *normal diffusion*. Particles on fractal networks diffuse more slowly compared with the case of normal diffusion. This slow diffusion called *anomalous diffusion* follows the relation given by $\langle r^2(t) \rangle \sim t^a$, where the condition $0 < a < 1$ always holds.

Brownian motion Einstein published the important paper in 1905 opening the way to investigate the movement of small particles suspended in a stationary liquid, the so-called Brownian motion, which stimulated J. Perrin in 1909 to pursue his experimental work confirming the atomic nature of matter. The trail of a random walker provides an instructive example for understanding the meaning of random fractal structures.

Fractons Fractons, excitations on fractal elastic-networks, were named by S. Alexander and R. Orbach in 1982. Fractons manifest not only static properties of fractal structures but also their dynamic properties. These modes show unique characteristics such as strongly localized nature with the localization length of the order of wavelength.

Spectral density of states The spectral density of states of ordinary elastic networks are expressed by the Debye spectral density of states given by $D(\omega) \sim \omega^{d-1}$, where d is the Euclidean dimensionality. The spectral density of states of fractal networks is given by $D(\omega) \sim \omega^{d_s - 1}$, where d_s is called the spectral or fracton dimension of the system.

Spectral dimension This exponent characterizes the spectral density of states for vibrational modes excited on fractal networks. The spectral dimension constitutes the dynamic exponent of fractal networks together with the conductivity exponent and the exponent of anomalous diffusion.

Definition of the Subject

The idea of fractals is based on *self-similarity*, which is a symmetry property of a system characterized by invariance under an isotropic scale-transformation on certain length scales. The term *scale-invariance* has the implication that objects look the same on different scales of observations. While the underlying concept of fractals is quite simple, the concept is used for an extremely broad range of topics, providing a simple description of highly complex structures found in nature. The term *fractal* was first introduced by Benoit B. Mandelbrot in 1975, who gave a definition on fractals in a simple manner "*A fractal is a shape made of parts similar to the whole in some way*". Thus far, the concept of fractals has been extensively used to understand the behaviors of many complex systems or

has been applied from physics, chemistry, and biology for applied sciences and technological purposes. Examples of fractal structures in condensed matter physics are numerous such as polymers, colloidal aggregations, porous media, rough surfaces, crystal growth, spin configurations of diluted magnets, and others. The critical phenomena of phase transitions are another example where self-similarity plays a crucial role. Several books have been published on fractals and reviews concerned with special topics on fractals have appeared.

Length, area, and volume are special cases of ordinary Euclidean *measures*. For example, length is the measure of a one-dimensional (1d) object, area the measure of a two-dimensional (2d) object, and volume the measure of a three-dimensional (3d) object. Let us employ a physical quantity (observable) as the measure to define dimensions for Euclidean systems, for example, a total mass $M(r)$ of a fractal object of the size r. For this, the following relation should hold

$$r \propto M(r)^{1/d} , \qquad (1)$$

where d is the Euclidean dimensionality. Note that Euclidean spaces are the simplest scale-invariant systems. We extend this idea to introduce dimensions for self-similar fractal structures. Consider a set of particles with unit mass m randomly distributed on a d-dimensional Euclidean space called the *embedding space* of the system. Draw a sphere of radius r and denote the total mass of particles included in the sphere by $M(r)$. Provided that the following relation holds *in the meaning of statistical average* such as

$$r \propto \langle M(r) \rangle^{1/D_f} , \qquad (2)$$

where $\langle \ldots \rangle$ denotes the ensemble-average over different spheres of radius r, we call D_f the *similarity dimension*. It is necessary, of course, that D_f is smaller than the embedding Euclidean dimension d. The definition of dimension as a *statistical quantity* is quite useful to specify the characteristic of a self-similar object if we could choose a suitable measure.

There are many definitions to allocate dimensions. Sometimes these take the same value as each other and sometimes not. The *capacity dimension* is based on the coverage procedure. As an example, the length of a curved line L is given by the product of the number N of straight-line segment of length r needed to step along the curve from one end to the other such as $L(r) = N(r)r$. While, the area $S(r)$ or the volume $V(r)$ of arbitrary objects can be measured by covering it with squares or cubes of linear

size r. The identical relation,

$$M(r) \propto N(r)r^d \tag{3}$$

should hold for the total mass $M(r)$ as measure, for example. If this relation does not change as $r \to 0$, we have the relation $N(r) \propto r^{-d}$. We can extend the idea to define the dimensions of fractal structures such as

$$N(r) \propto r^{-D_f} , \tag{4}$$

from which the *capacity dimension* D_f is given by

$$D_f := \lim_{r \to 0} \frac{\ln N(r)}{\ln(1/r)} . \tag{5}$$

The definition of D_f can be rendered in the following implicit form

$$\lim_{r \to 0} N(r)r^{D_f} = \text{const} . \tag{6}$$

Equation (5) brings out a key property of the *Hausdorff dimension* [10], where the product $N(r)r^{D_f}$ remains finite as $r \to 0$. If D_f is altered even by an infinitesimal amount, this product will diverge either to zero or to infinity. The Hausdorff dimension coincides with the capacity dimension for many fractal structures, although the Hausdorff dimension is defined less than or equal to the capacity dimension. Hereafter, we refer to the capacity dimension or the Hausdorff dimension mentioned above as the *fractal dimension*.

Introduction

Fractal structures are classified into two categories; *deterministic* fractals and *random* fractals. In condensed matter physics, we encounter many examples of random fractals. The most important characteristic of random fractals is the spatial and/or sample-to-sample fluctuations in

their properties. We must discuss their characteristics by averaging over a large ensemble. The nature of deterministic fractals can be easily understood from some examples. An instructive example is the *Mandelbrot–Given fractal* [12], which can be constructed by starting with a structure with eight line segments as shown in Fig. 1a (the first stage of the Mandelbrot–Given fractal). In the second stage, each line segment of the initial structure is replaced by the initial structure itself (Fig. 1b). This process is repeated indefinitely. The Mandelbrot–Given fractal possesses an obvious dilatational symmetry, as seen from Fig. 1c, i. e., when we magnify a part of the structure, the enlarged portion looks just like the original one. Let us apply (5) to determine D_f of the Mandelbrot–Given fractal. The Mandelbrot–Given fractal is composed of 8 parts of size 1/3, hence, $N(1/3) = 8$, $N((1/3)^2) = 8^2$, and so on. We thus have a relation of the form $N(r) \propto r^{-\ln_3 8}$, which gives the fractal dimension $D_f = \ln_3 8 = 1.89278\dots$. The Mandelbrot–Given fractal has many analogous features with *percolation networks* (see Sect. "Dynamical Properties of Fractal Structures"), a typical random fractal, such that the fractal dimension of a 2d percolation network is $D_f = 91/48 = 1.895833\dots$, which is very close to that of the Mandelbrot–Given fractal.

The geometric characteristics of random fractals can be understood by considering two extreme cases of random structures. Figure 2a represents the case in which particles are randomly but *homogeneously* distributed in a d-dimensional box of size L, where d represents ordinary Euclidean dimensionality of the embedding space. If we divide this box into smaller boxes of size l, the *mass density* of the ith box is

$$\rho_i(l) = \frac{M_i(l)}{l^d} , \tag{7}$$
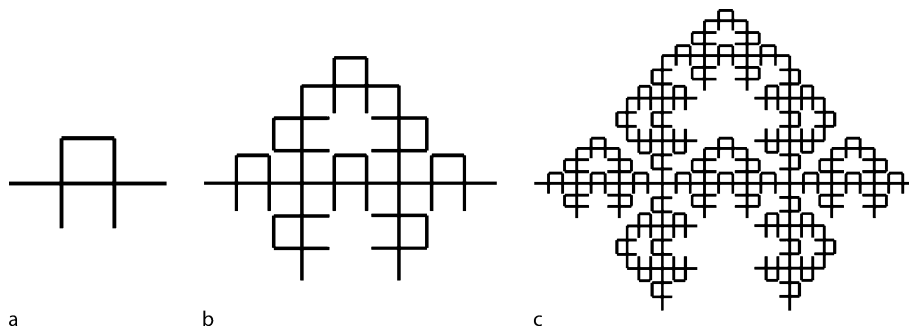
Fractal Structures in Condensed Matter Physics, Figure 1
Mandelbrot–Given fractal. a The initial structure with eight line segments, **b** the object obtained by replacing each line segment of the initial structure by the initial structure itself (the second stage), and **c** the third stage of the Mandelbrot–Given fractal obtained by replacing each line segment of the second-sage structure by the initial structure
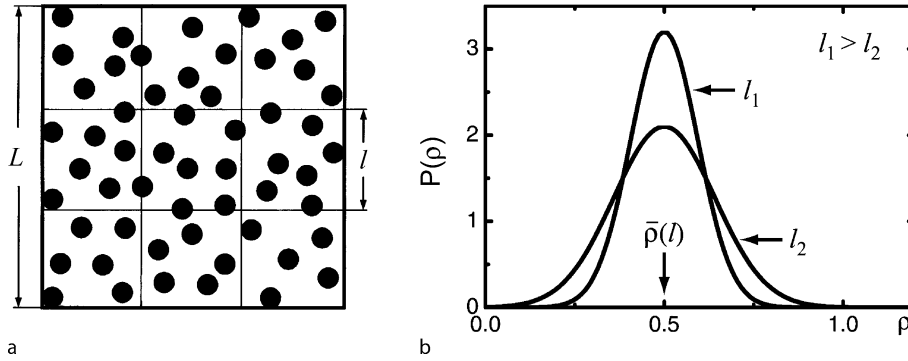
Fractal Structures in Condensed Matter Physics, Figure 2
a Homogeneous random structure in which particles are randomly but homogeneously distributed, and **b** the distribution functions of local densities ρ, where $\overline{\rho}(l)$ is the average mass density independent of l

where $M_i(l)$ represents the total mass (measure) inside box i. Since this quantity depends on the box i, we plot the distribution function $P(\rho)$, from which curves like those in Fig. 2b may be obtained for two box sizes l_1 and ($l_2 < l_2$). We see that the central peak position of the distribution function $P(\rho)$ is the same for each case. This means that the average mass density yields

$$\overline{\rho}(l) = \frac{\langle M_i(l) \rangle_i}{l^d}$$

becomes constant, indicating that $\langle M_i(l) \rangle_i \propto l^d$. The above is equivalent to

$$\overline{\rho} = \frac{m}{\overline{a}^d} \, , \tag{8}$$

where $\overline{a}$ is the average distance (characteristic length-scale) between particles and the mass of a single particle. This indicates that there exists a single length scale $\overline{a}$ characterizing the random system given in Fig. 2a.

The other type of random structure is shown in Fig. 3a, where particle positions are correlated with each other and $\rho_i(l)$ greatly fluctuates from box to box, as shown in Fig. 3b. The relation $\langle M_i(l) \rangle_i \propto l^d$ may not hold at all for this type of structure. Assuming the fractality for this system, namely, if the power law $\langle M_i(l) \rangle_i \propto l_f^D$ holds, the average mass density becomes

$$\overline{\rho}(l) = \frac{\langle M_i(l) \rangle_i}{l^d} \propto l^{D_f - d} \, , \tag{9}$$

where $\rho_i(l) = 0$ is excluded. In the case $D_f < d$, $\overline{\rho}(l)$ depends on l and decreases with increasing l. Thus, there is no characteristic length scale for the type of random structure shown in Fig. 3a. If (9) holds with $D_f < d$, so that $\langle M_i(l) \rangle_i$ is proportional to l_f^D, the structure is said to be *fractal*. It is important to note that there is no *characteristic length scale* for the type of random fractal structure shown in Fig. 3b. Thus, we can extend the idea of self-similarity not only for deterministic self-similar structures, but also
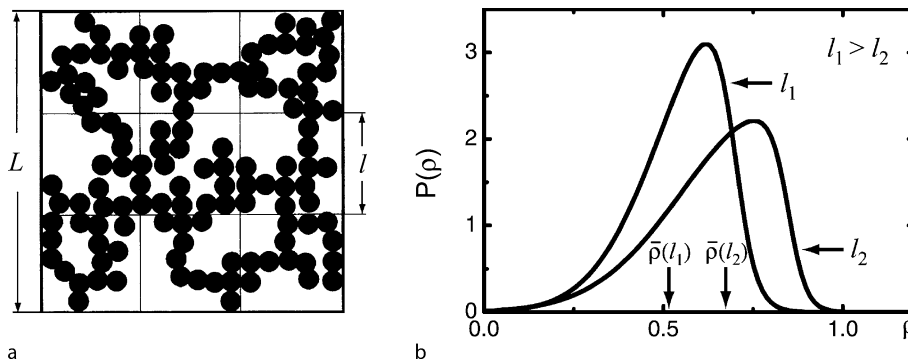
Fractal Structures in Condensed Matter Physics, Figure 3
a Correlated random fractal structure in which particles are randomly distributed, but correlated with each other, and **b** the distribution functions of local densities ρ with finite values, where the average mass densities depend on l
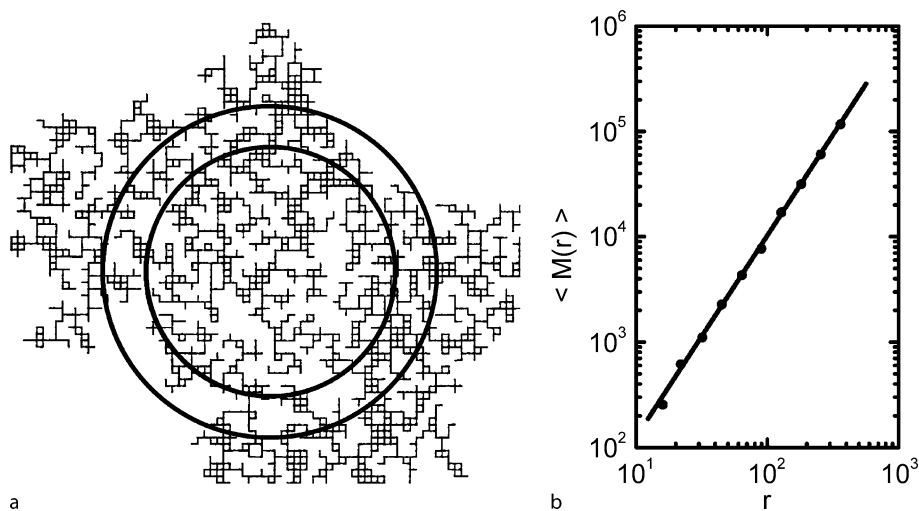
Fractal Structures in Condensed Matter Physics, Figure 4
a 2d site-percolation network and circles with different radii. **b** The power law relation holds between *r* and the number of particles in the sphere of radius *r*, indicating the fractal dimension of the 2d network is $D_f = 1.89\ldots = 91/48$

for random and disordered structures, the so-called *random fractals*, in the meaning of statistical average.

The percolation network made by putting particles or bonds on a lattice with the probability *p* is a typical example of random fractals. The theory of percolation was initiated in 1957 by S.R. Broadbent and J.M. Hammersley [5] in connection with the diffusion of gases through porous media. Since their work, it has been widely accepted that the percolation theory describes a large number of physical and chemical phenomena such as gelation processes, transport in amorphous materials, hopping conduction in doped semiconductors, the quantum Hall effect, and many other applications. In addition, it forms the basis for studies of the flow of liquids or gases through porous media. Percolating networks thus serve as a model which helps us to understand physical properties of complex fractal structures.

For both deterministic and random fractals, it is remarkable that no characteristic length scale exists, and this is a key feature of fractal structures. In other words, fractals are defined to be objects invariant under isotropic scale transformations, i. e., uniform dilatation of the system in every spatial direction. In contrast, there exist systems which are invariant under *anisotropic* transformations. These are called *self-affine fractals*.

Determining Fractal Dimensions

There are several methods to determine fractal dimensions D_f of complex structures encountered in condensed mat-

ter physics. The following methods for obtaining the fractal dimension D_f are known to be quite efficient.

Coverage Method

The idea of coverage in the definition of the capacity dimension (see (5)) can be applied to obtain the fractal dimension D_f of material surfaces. An example is the fractality of rough surfaces or inner surfaces of porous media. The fractal nature is probed by changing the sizes of adsorbed molecules on solid surfaces. Power laws are verified by plotting the total number of adsorbed molecules versus their size *r*. The area of a surface can be estimated with the aid of molecules weakly adsorbed by van der Waals forces. Gas molecules are adsorbed on empty sites until the surface is uniformly covered with a layer one molecule thick. Provided that the radius *r* of one adsorbed molecule and the number of adsorbed molecules $N(r)$ are known, the surface area *S* obtained by molecules is given by

$$S(r) \propto N(r)r^2 . \tag{10}$$

If the surface of the adsorbate is perfectly smooth, we expect the measured area to be independent of the radius *r* of the probe molecules, which indicates the power law

$$N(r) \propto r^{-2} . \tag{11}$$

However, if the surface of the adsorbate is rough or contains pores that are small compared with *r*, less of the surface area *S* is accessible with increasing size *r*. For a fractal

surface with fractal dimension D_f, (11) gives the relation

$$N(r) \propto r^{-D_f} . \tag{12}$$

Box-Counting Method

Consider as an example a set of particles distributed in a space. First, we divide the space into small boxes of size r and count the number of boxes containing more than one particle, which we denote by $N(r)$. From the definition of the capacity dimension (4), the number of particle

$$N(r) \propto r^{-D_f} . \tag{13}$$

For homogeneous objects distributed in a d-dimensional space, the number of boxes of size r becomes, of course

$$N(r) \propto r^{-d} .$$

Correlation Function

The fractal dimension D_f can be obtained via the correlation function, which is the fundamental statistical quantity observed by means of X-ray, light, and neutron scattering experiments. These techniques are available to bulk materials (not surface), and is widely used in condensed matter physics. Let $\rho(r)$ be the number density of atoms at position r. The density-density correlation function $G(r, r')$ is defined by

$$G(r, r') = \langle \rho(r)\rho(r') \rangle , \tag{14}$$

where $\langle \ldots \rangle$ denotes an ensemble average. This gives the correlation of the number-density fluctuation. Provided that the distribution is isotropic, the correlation function becomes a function of only one variable, the radial distance $r = |r - r'|$, which is defined in spherical coordinates. Because of the translational invariance of the system on average, r' can be fixed at the coordinate origin $r' = 0$. We can write the correlation function as

$$G(r) = \langle \rho(r)\rho(0) \rangle . \tag{15}$$

The quantity $\langle \rho(r)\rho(0) \rangle$ is proportional to the probability that a particle exists at a distance r from another particle. This probability is proportional to the particle density $\rho(r)$ within a sphere of radius r. Since $\rho(r) \propto r^{D_f - d}$ for a fractal distribution, the correlation function becomes

$$G(r) \propto r^{D_f - d} , \tag{16}$$

where D_f and d are the fractal and the embedded Euclidean dimensions, respectively. This relation is often used directly to determine D_f for random fractal structures.

The scattering intensity in an actual experiment is proportional to the structure factor $S(q)$, which is the Fourier transform of the correlation function $G(r)$. The structure factor is calculated from (16) as

$$S(q) = \frac{1}{V} \int_V G(r) e^{iq \cdot r} dr \propto q^{-D_f} \tag{17}$$

where V is the volume of the system. Here dr is the d-dimensional volume element. Using this relation, we can determine the fractal dimension D_f from the data obtained by scattering experiments.

When applying these methods to obtain the fractal dimension D_f, we need to take care over the following point. Any fractal structures found in nature must have upper and lower length-limits for their fractality. There usually exists a *crossover* from homogeneous to fractal. Fractal properties should be observed only between these limits.

We describe in the succeeding Sections several examples of fractal structures encountered in condensed matter physics.

Polymer Chains in Solvents

Since the concept of fractal was coined by B.B. Mandelbrot in 1975, scientists reinterpreted random complex structures found in condensed matter physics in terms of fractals. They found that a lot of objects are classified as fractal structures. We show at first from polymer physics an instructive example exhibiting the fractal structure. That is an early work by P.J. Flory in 1949 on the relationship between the mean-square end-to-end distance of a polymer chain $\langle r^2 \rangle$ and the degree of polymerization N. Consider a dilute solution of separate coils in a solvent, where the total length of a flexible polymer chain with a monomer length a is Na. The simplest idealization views the polymer chain in analogy with a *Brownian motion* of a random walker. The walk is made by a succession of N steps from the origin $r = 0$ to the end point r. According to the *central limit theorem* of the probability theory, the probability to find a walker at r after N steps ($N \gg 1$) follows the *diffusion equation* and we have the expression for the probability to find a particle after N steps at r

$$P_N(r) = (2\pi Na^2/3)^{-3/2} \exp(-3r^2/2Na^2) , \tag{18}$$

where the prefactor arises from the normalization of $P_N(r)$. The mean squared distance calculated from $P_N(r)$ becomes

$$\langle r^2 \rangle = \int r^2 P_N(r) d^3r = Na^2 . \tag{19}$$

Then, the mean-average end-to-end distance of a polymer chain yields $R = \langle r^2 \rangle^{1/2} = N^{1/2}a$. Since the number

of polymerization N corresponds to the total mass M of a polymer chain, the use of (19) leads to the relation such as $M(R) \sim R^2$. The mass $M(R)$ can be considered as a measure of a polymer chain, the fractal dimension of this ideal chain as well as the trace of Brown motion becomes $D_f = 2$ for any d-dimensional embedding space.

The entropy of the idealized chain of the length $L = Na$ is obtained from (18) as

$$S(r) = S(0) - \frac{3r^2}{2R^2} , \tag{20}$$

from which the free energy $F_{el} = U - TS$ is obtained as

$$F_{el}(r) = F_{el}(0) + \frac{3k_B T r^2}{2R^2} . \tag{21}$$

Here U is assumed to be independent of distinct configurations of polymer chains. This is an elastic energy of an ideal chain due to entropy where F_{el} decreases as $N \to$ large. P.J. Flory added the repulsive energy term due to monomer-monomer interactions, the so-called *excluded volume effect*. This has an analogy with *self-avoiding random walk*. The contribution to the free energy is obtained by the virial expansion into the power series on the concentration $c_{int} = N/r^d$. According to the mean field theory on the repulsive term $F_{int} \propto c_{int}^2$, we have the total free-energy F such as

$$\frac{F}{k_B T} = \frac{3r^2}{2Na^2} + \frac{v(T)N^2}{r^d} , \tag{22}$$

where $v(T)$ is the excluded volume parameter. We can obtain a minimum of $F(r)$ at $r = R$ by differentiating $F(r)$ with respect to r such that

$$M(R) \propto R^{\frac{d+2}{3}} . \tag{23}$$

Here the number of polymerization N corresponds to the total mass $M(R)$ of a polymer chain. Thus, we have the fractal dimension $D_f = (d + 2)/3$, in particular, $D_f = 5/3 = 1.666\ldots$ for a polymer chain in a solvent.

Aggregates and Flocs

The structures of a wide variety of flocculated colloids in suspension (called aggregates or flocs) can be described in terms of *fractals*. A colloidal suspension is a fluid containing small charged particles that are kept apart by Coulomb repulsion and kept afloat by Brownian motion. A change in the particle-particle interaction can be induced by varying the chemical composition of the solution and in this manner an aggregation process can be initiated. Aggregation processes are classified into two simple types: diffusion-limited aggregation (DLA) and diffusion-limited

cluster-cluster aggregation (DLCA), where a DLA is due to the cluster-particle coalescence and a DLCA to the cluster-cluster flocculation. In most cases, actual aggregates involve a complex interplay between a variety of flocculation processes. The pioneering work was done by M.V. Smoluchowski in 1906, who formulated a kinetic theory for the irreversible aggregation of particles into clusters and further clusters combining with clusters. The inclusion of cluster-cluster aggregation makes this process distinct from the DLA process due to particle-cluster interaction. There are two distinct limiting regimes of the irreversible colloidal aggregation process: the diffusion-limited CCA (DLCA) in dilute solutions and the reaction-limited CCA (RLCA) in dense solutions. The DLCA is due to the fast process determined by the time for the clusters to encounter each other by diffusion, and the RLCA is due to the slow process since the cluster-cluster repulsion has to dominate thermal activation.

Much of our understanding on the mechanism forming aggregates or flocs has been mainly due to computer simulations. The first simulation was carried out by Vold in 1963 [23], who used the ballistic aggregation model and found that the number of particles $N(r)$ within a distance r measured from the first seed particle is given by $N(r) \sim r^{2.3}$. Though this relation surely exhibits the scaling form of (2), the applicability of this model for real systems was doubted in later years. The researches on fractal aggregates has been developed from a simulation model on DLA introduced by T.A. Witten and L.M. Sander in 1981 [26] and on the DLCA model proposed by P. Meakin in 1983 [14] and M. Kolb et al. in 1983 [11], independently. The DLA has been used to describe diverse phenomena forming fractal patterns such as electro-depositions, surface corrosions and dielectric breakdowns. In the simplest version of the DLA model for irreversible colloidal aggregation, a particle is located at an initial site $r = 0$ as a seed for cluster formation. Another particle starts a random walk from a randomly chosen site in the spherical shell of radius r with width $dr (\ll r)$ and center $r = 0$. As a first step, a random walk is continued until the particle contacts the seed. The cluster composed of two particles is then formed. Note that the finite-size of particles is the very reason of dendrite structures of DLA. This procedure is repeated many times, in each of which the radius r of the starting spherical shell should be much larger than the gyration radius of the cluster. If the number of particles contained in the DLA cluster is huge (typically $10^4 \sim 10^8$), the cluster generated by this process is highly branched, and forms fractal structures in the meaning of statistical average. The fractality arises from the fact that the faster growing parts of the cluster shield the other parts, which there-
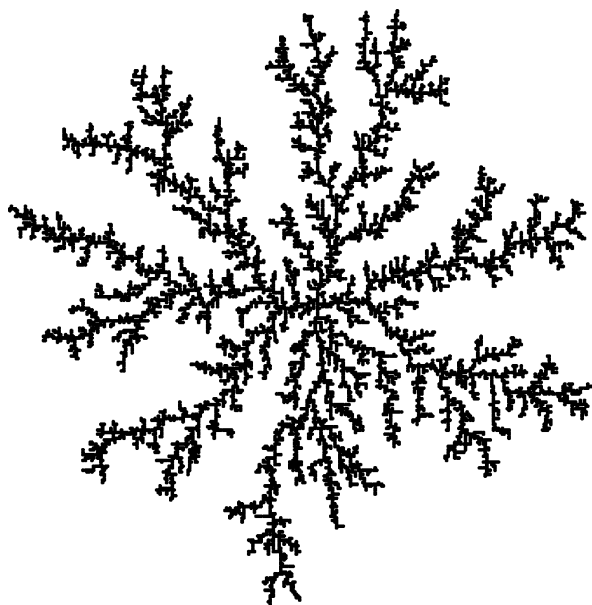
Fractal Structures in Condensed Matter Physics, Figure 5
Simulated results of a 2d diffusion-limited aggregation (DLA). The number of particles contained in this DLA cluster is 10^4

illustrates a simulated result for a 2d DLA cluster obtained by the procedure mentioned above. The number of particles N inside a sphere of radius L ($\ll$ the gyration radius of the cluster) follows the scaling law given by

$$N \propto L^{D_f} , \qquad (24)$$

where the fractal dimension takes a value of $D_f \approx 1.71$ for the 2d DLA cluster and $D_f \approx 2.5$ for the 3d DLA cluster without an underlying lattice. Note that these fractal dimensions are sensitive to the embedding lattice structure. The reason for this open structure is that a wandering molecule will settle preferentially near one of the tips of the fractal, rather than inside a cluster. Thus, different sites have different growth probabilities, which are high near the tips and decrease with increasing depth inside a cluster.

One of the most extensively studied DLA processes is the growth of metallic forms by electrochemical deposition. The scaling properties of electrodeposited metals were pointed out by R.M. Brady and R.C. Ball in 1984 for copper electrodepositions. The confirmation of the fractality for zinc metal leaves was made by M. Matsushita et al. in 1984. In their experiments [13], zinc metal leaves are grown two-dimensionally by electrodeposition. The structures clearly recover the pattern obtained by computer simulations for the DLA model proposed by T.A. Witten and L.M. Sander in 1981. Figure 6 shows a typical zinc

fore become less accessible to incoming particles. An arriving random walker is far more likely to attach to one of the tips of the cluster. Thus, the essence of the fractal-pattern formation arises surely from nonlinear process. Figure 5
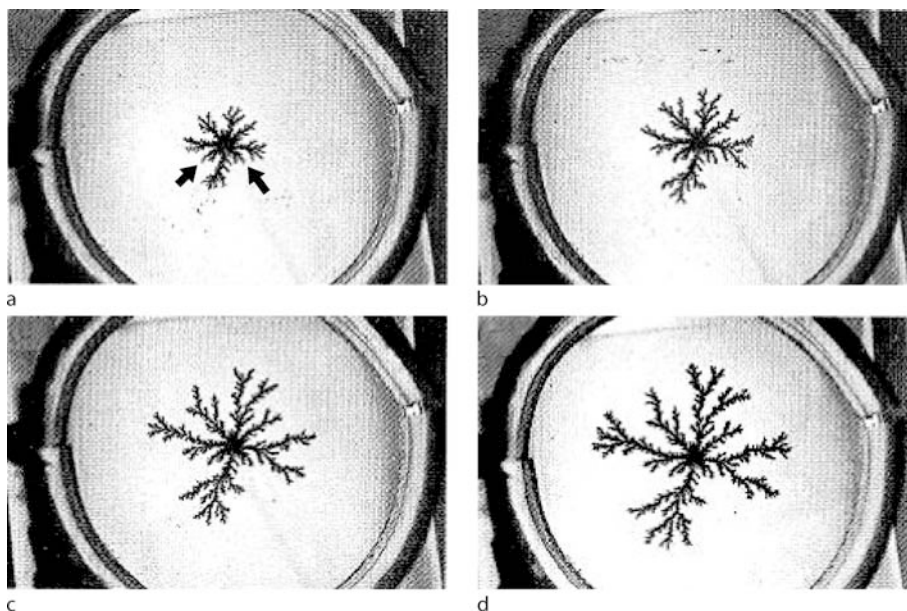
Fractal Structures in Condensed Matter Physics, Figure 6
The fractal structures of zinc metal leaves grown by electrodeposition. Photographs **a–d** were taken 3,5,9, and 15 min after initiating the electrolysis, respectively. After [13]

dendrite that was deposited on the cathode in one of these experiments. The fractal dimensionality $D_f = 1.66 \pm 0.33$ was obtained by computing the density-density correlation function $G(r)$ for patterns grown at applied voltages of less than 8V.

The fractality of uniformly sized gold-colloid aggregates according to the DLCA was experimentally demonstrated by D.A. Weitz in 1984 [25]. They used transmission-electron micrographs to determine the fractal dimension of this systems to be $D_f = 1.75$. They also performed quasi-elastic light-scattering experiments to investigate the dynamic characteristics of DLCA of aqueous gold colloids. They confirmed the scaling behaviors for the dependence of the mean cluster size on both time and initial concentration.

These works were performed *consciously* to examine the fractality of aggregates. There had been earlier works exhibiting the mass-size scaling relationship for actual aggregates. J.M. Beeckmans [2] pointed out in 1963 the power law behaviors by analyzing the data for aerosol and precipitated smokes in the literature (1922–1961). He used in his paper the term "aggregates-within-aggregates", implying the fractality of aggregates. However, the data available at that stage were not adequate and scattered. Therefore, this work did not provide decisive results on the fractal dimensions of aggregates. There were smarter experiments by N. Tambo and Y. Watanabe in 1967 [20], which precisely determined fractal dimensions of flocs formed in an aqueous solution. These were performed without being aware of the concept of fractals. Original works were published in Japanese. The English versions of these works were published in 1979 [21]. We discuss these works below.

Flocs generated in aqueous solutions have been the subject of numerous studies ranging from basic to applied sciences. In particular, the settling process of flocs formed in water incorporating kaolin colloids is relevant to water and wastewater treatment. The papers by N. Tambo and Y. Watanabe pioneered the discussion on the so-called *fractal approach* to floc structures; they performed their own settling experiments to clarifying the size dependences of mass densities for clay-aluminum flocs by using Stokes' law $u_r \propto \Delta\rho(r)r^2$ where $\Delta\rho$ is the difference between the densities of water and flocs taking so-small values $\Delta\rho \sim 0.01$–0.001 g/cm^3. Thus, the settling velocities u_r are very slow of the order of 0.001 m/sec for flocs of sizes $r \sim 0.1$ mm, which enabled them to perform precise measurements. Since flocs are very fragile aggregates, they made the settling experiments with special cautions on convection and turbulence, and by careful and intensive experiments of flocculation conditions. They confirmed

from thousands of pieces of data the scaling relationship between settling velocities u_r and sizes of aggregates such as $u_r \propto r^b$. From the analysis of these data, they found the scaling relation between effective mass densities and sizes of flocs such as $\Delta\rho(r) \propto r^{-c}$, where the exponents c were found to take values from 1.25 to 1.00 depending on the aluminum-ion concentration, showing that the fractal dimensions become $D_f = 1.75$ to 2.00 with increasing aluminum-ion concentration. This is because the repulsive force between charged clay-particles is screened, and van der Waals attractive force dominates between the pair of particles. It is remarkable that these fractal dimensions D_f show excellent agreement with those determined for actual DLCA and RLCA clusters in the 1980s by using various experimental and computer simulation methods. Thus, they had found that the size dependences of mass densities of flocs are controlled by the aluminum-ion concentration dosed/suspended particle concentration, which they named the ALT ratio. These correspond
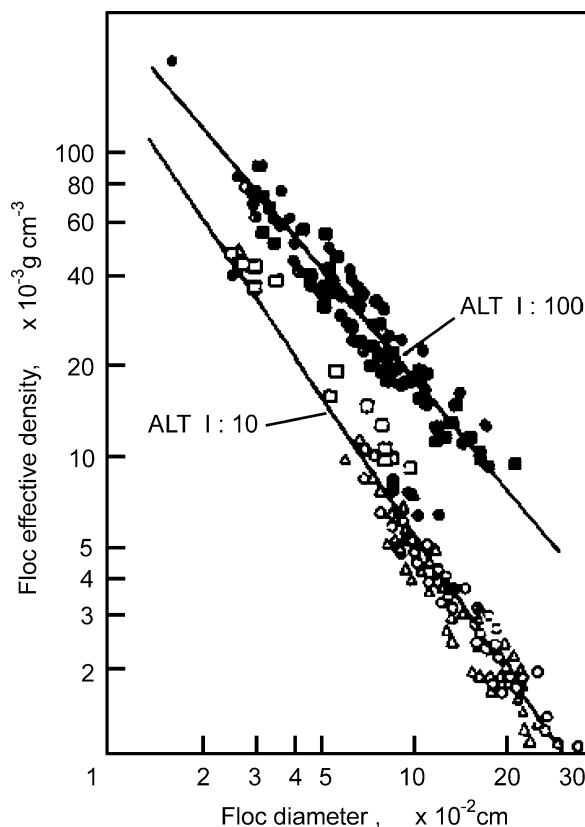
Fractal Structures in Condensed Matter Physics, Figure 7
Observed scaling relations between floc densities and their diameters where aluminum chloride is used as coagulants. After [21]

to the transition from DLCA (established now taking the value of $D_f \approx 1.78$ from computer simulations) process to the RLCA one (established at present from computer simulations as $D_f \approx 2.11$). The ALT ratio has since the publication of the paper been used in practice as a criterion for the coagulation to produce flocs with better settling properties and less sludge volume. We show their experimental data in Fig. 7, which demonstrate clearly that flocs (aggregates) are fractal.

Aerogels

Silica aerogels are extremely light materials with porosities as high as 98% and take fractal structures. The initial step in the preparation is the hydrolysis of an alkoxysilane $Si(OR)_4$, where R is CH_3 or C_2H_5. The hydrolysis produces silicon hydroxide $Si(OH)_4$ groups which polycondense into siloxane bonds –Si–O–Si–, and small particles start to grow in the solution. These particles bind to each other by *diffusion-limited cluster-cluster aggregation* (DLCA) (see Sect. "Aggregates and Flocs") until eventually they produce a disordered network filling the reaction volume. After suitable aging, if the solvent is extracted above the critical point, the open porous structure of the network is preserved and decimeter-size monolithic blocks with a range of densities from 50 to 500 kg/m^3 can be obtained. As a consequence, aerogels exhibit unusual physical properties, making them suitable for a number of practical applications, such as Cerenkov radiation detectors, supports for catalysis, or thermal insulators.

Silica aerogels possess two different length scales. One is the radius r of primary particles. The other length is the correlation length of the gel. At intermediate length scales, lying between these two length scales, the clusters possess a fractal structure and at larger length scales the gel is a homogeneous porous glass. Aerogels have a very low thermal conductivity, solid-like elasticity, and very large internal surfaces.

In elastic neutron scattering experiments, the scattering differential cross-section measures the Fourier components of spatial fluctuations in the mass density. For aerogels, the differential cross-section is the product of three factors, and is expressed by

$$\frac{d\sigma}{d\Omega} = Af^2(q)S(q)C(q) + B \, . \tag{25}$$

Here A is a coefficient proportional to the particle concentration and $f(q)$ is the primary-particle form factor. The structure factor $S(q)$ describes the correlation between particles in a cluster and $C(q)$ accounts for cluster-cluster correlations. The incoherent background is expressed by B. The structure factor $S(q)$ is proportional to the spatial Fourier transform of the density-density correlation function defined by (16), and is given by (17). Since the structure of the aerogel is fractal up to the correlation length ξ of the system and homogeneous for larger scales, the correlation function $G(r)$ is expressed by (25) for $r \ll \xi$ and $G(r) = $ Const. for $r \gg \xi$. Corresponding to this, the structure factor $S(q)$ is given by (17) for $q\xi \gg 1$, while $S(q)$ is independent of q for $q\xi \ll 1$. The wavenumber regime for which $S(q)$ becomes a constant is called the Guinier regime. The value of D_f can be deduced from the slope of the observed intensity versus momentum transfer ($q\xi \gg 1$) in a double logarithmic plot. For very large q, there exists a regime called the *Porod regime* in which the scattering intensity is proportional to q^{-4}.
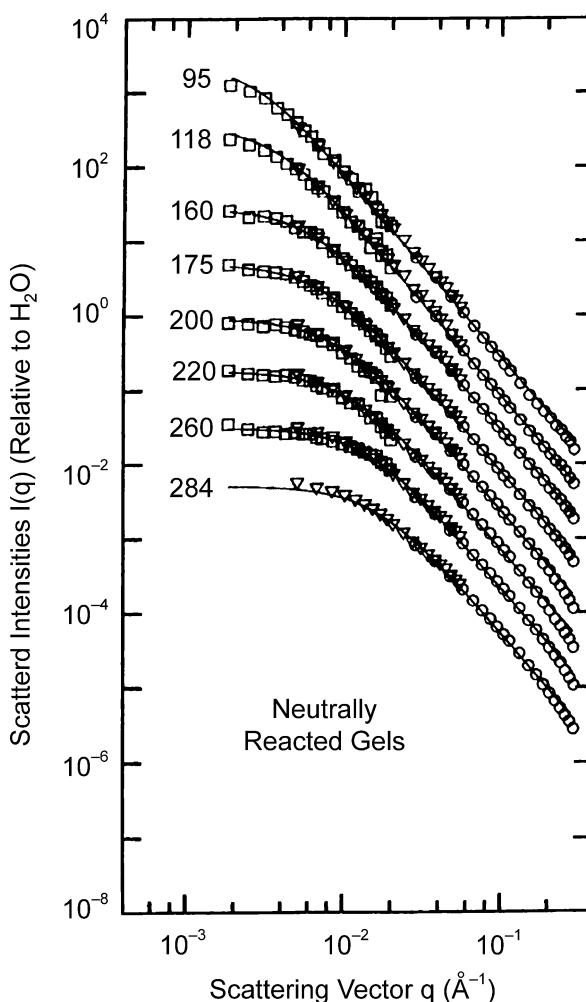
Fractal Structures in Condensed Matter Physics, Figure 8
Scattered intensities for eight neutrally reacted samples. Curves are labeled with ρ in kg/m^3. After [22]

The results in Fig. 8 by R. Vacher et al. [22] are from small-angle neutron scattering experiments on silica aerogels. The various curves are labeled by the macroscopic density ρ of the corresponding sample in Fig. 8. For example, 95 refers to a neutrally reacted sample with $\rho = 95\,\mathrm{kg/m^3}$. Solid lines represent best fits. They are presented even in the particle regime $q > 0.15\,\text{Å}^{-1}$ to emphasize that the fits do not apply in the region, particularly for the denser samples. Remarkably, D_f is independent of sample density to within experimental accuracy: $D_\mathrm{f} = 2.40 \pm 0.03$ for samples 95 to 360. The departure of $S(q)$ from the q^{-D_f} dependence at large q indicates the presence of particles with gyration radii of a few Å.

Dynamical Properties of Fractal Structures

The dynamics of fractal objects is deeply related to the time-scale problems such as diffusion, vibration and transport on fractal support. For the diffusion of a particle on any d-dimensional ordinary Euclidean space, it is well known that the mean-square displacement $\langle r^2(t)\rangle$ is proportional to the time such as $\langle r^2(t)\rangle \propto t$ for any Euclidean dimension d (see also (19)). This is called *normal diffusion*. While, on fractal supports, a particle more slowly diffuses, and the mean-square displacement follows the power law

$$\langle r^2(t)\rangle \propto t^{2/d_\mathrm{w}}, \qquad (26)$$

where d_w is termed the exponent of *anomalous diffusion*. The exponents is expressed as $d_\mathrm{w} = 2 + \theta$ with a positive $\theta > 0$ (see (31)), implying that the diffusion becomes slower compared with the case of normal diffusion. This is because the inequality $2/d_\mathrm{w} < 1$ always holds. This slow diffusions on fractal supports are called *anomalous diffusion*.

The scaling relation between the length-scale and the time-scale can be easily extended to the problem of atomic vibrations of elastic fractal-networks. This is because various types of equations governing dynamics can be mapped onto the diffusion equation. This implies that both equations are governed by the same eigenvalue problem, namely, the replacement of eigenvalues $\omega \to \omega^2$ between the diffusion equation and the equation of atomic vibrations is justified. Thus, the basic properties of vibrations of fractal networks, such as the density of states, the dispersion relation and the localization/delocalization property, can be derived from the same arguments for diffusion on fractal networks. The dispersion relation between the frequency ω and the wavelength $\Lambda(\omega)$ is obtained from (26) by using the reciprocal relation $t \to \omega^{-2}$ (here the diffusion problem is mapped onto the vibrational

one) and $\langle r^2(t)\rangle \to \Lambda(\omega)^{-2}$. Thus we obtain the dispersion relation for vibrational excitations on fractal networks such as

$$\omega \propto \Lambda(\omega)^{d_\mathrm{w}/2}. \qquad (27)$$

If $d_\mathrm{w} = 2$, we have the ordinary dispersion relation $\omega \propto \Lambda(\omega)$ for elastic waves excited on homogeneous systems.

Consider the diffusion of a random walker on a *percolating fractal network*. How does $\langle r^2(t)\rangle$ behave in the case of fractal percolating networks? For this, P.G. de Gennes in 1976 [7] posed the problem called an ant in the labyrinth. Y. Gefen et al. in 1983 [9] gave a fundamental description of this problem in terms of a *scaling argument*. D. Ben-Avraham and S. Havlin in 1982 [3] investigated this problem in terms of Monte Carlo simulations. The work by Y. Gefen [9] triggered further developments in the dynamics of fractal systems, where the *spectral (or fracton) dimension d_s* is a key dimension for describing the dynamics of fractal networks, in addition to the fractal dimension D_f. The fractal dimension D_f characterizes how the geometrical distribution of a static structure depends on its length scale, whereas the spectral dimension d_s plays a central role in characterizing dynamic quantities on fractal networks. These dynamical properties are described in a unified way by introducing a new dynamic exponent called the *spectral or fracton* dimension defined by

$$d_\mathrm{s} = \frac{2D_\mathrm{f}}{d_\mathrm{w}}. \qquad (28)$$

The term *fracton*, coined by S. Alexander and R. Orbach in 1982 [1], denotes vibrational modes peculiar to fractal structures. The characteristics of fracton modes cover a rich variety of physical implications. These modes are strongly localized in space and their localization length is of the order of their wavelengths.

We give below the explicit form of the exponent of anomalous diffusion d_w by illustrating percolation fractal networks. The mean-square displacement $\langle r^2(t)\rangle$ after a sufficiently long time t should follow the anomalous diffusion described by (26). For a finite network with a size ξ, the mean-square distance at sufficiently large time becomes $\langle r^2(t)\rangle \approx \xi^2$, so we have the diffusion coefficient for anomalous diffusion from (26) such as

$$D \propto \xi^{2-d_\mathrm{w}}. \qquad (29)$$

For percolating networks, the diffusion constant D in the vicinity of the critical percolation density p_c behaves

$$D \propto (p - p_\mathrm{c})^{t-\beta} \propto \xi^{-(t-\beta)/\nu}, \qquad (30)$$

where t is called the *conductivity exponent* defined by $\sigma_{dc} \sim (p - p_c)^t$, β the exponent for the percolation order parameter defined by $S(p) \propto (p - p_c)^{\beta}$, and ν the exponent for the correlation length defined by $\xi \propto |p - p_c|^{-\nu}$, respectively. Comparing (29) and (30), we have the relation between exponents such as

$$d_w = 2 + \frac{t - \beta}{\nu} = 2 + \theta . \tag{31}$$

Due to the condition $t > \beta$, and hence $\theta > 0$, implying that the diffusion becomes slow compared with the case of normal diffusion. This slow diffusion is called anomalous diffusion.

Spectral Density of States and Spectral Dimensions

The spectral density of states of atomic vibrations is the most fundamental quantity describing the dynamic properties of homogeneous or fractal systems such as specific heats, heat transport, scattering of waves and others. The simplest derivation of the spectral density of states (abbreviated, SDOS) of a homogeneous elastic system is given below. The density of states at ω is defined as the number of modes per particle, which is expressed by

$$D(\Delta\omega) = \frac{1}{\Delta\omega L^d} , \tag{32}$$

where $\Delta\omega$ is the frequency interval between adjacent eigenfrequencies close to ω and L is the linear size of the system. In the lowest frequency region, $\Delta\omega$ is the lowest eigenfrequency which depends on the size L. The relation between the frequency $\Delta\omega$ and L is obtained from the well-known linear dispersion relationship $\omega = \upsilon k$, where υ is the velocity of phonons (quantized elastic waves) such that

$$\Delta\omega = \frac{2\pi\upsilon}{\lambda} \propto \frac{1}{L} . \tag{33}$$

The substitution of (33) into (32) yields

$$D(\Delta\omega) \propto \Delta\omega^{d-1} . \tag{34}$$

Since this relation holds for any length scale L due to the scale-invariance property of homogeneous systems, we can replace the frequency $\Delta\omega$ by an arbitrary ω. Therefore, we obtain the conventional Debye density of states as

$$D(\omega) \propto \omega^{d-1} . \tag{35}$$

It should be noted that this derivation is based on the scale invariance of the system, suggesting that we can derive the

SDOS for fractal networks in the same line with this treatment. Consider the SDOS of a fractal structure of size L with fractal dimension D_f. The density of states per particle at the lowest frequency $\Delta\omega$ for this system is, as in the case of (32), written as

$$D(\Delta\omega) \propto \frac{1}{L_f^D \Delta\omega} . \tag{36}$$

Assuming that the dispersion relation for $\Delta\omega$ corresponding to (33) is

$$\Delta\omega \propto L^{-z} , \tag{37}$$

we can eliminate L from (36) and obtain

$$D(\Delta\omega) \propto \Delta\omega^{D_f/z - 1} . \tag{38}$$

The exponent z of the dispersion relation (37) is evaluated from the exponent of anomalous diffusion d_w. Considering the mapping correspondence between diffusion and atomic vibrations, we can replace $\langle r^2(t) \rangle$ and t by L^2 and $1/\Delta\omega^2$, respectively. Equation (26) can then be read as

$$L \propto \Delta\omega^{-2/d_w} . \tag{39}$$

The comparison of (28),(37) and (39) leads to

$$z = \frac{d_w}{2} = \frac{D_f}{d_s} . \tag{40}$$

Since the system has a scale-invariant fractal (self-similar) structure $\Delta\omega$, can be replaced by an arbitrary frequency ω. Hence, from (38) and (40) the SDOS for fractal networks is found to be

$$D(\omega) \propto \omega^{d_s - 1} , \tag{41}$$

and the dispersion relation (39) becomes

$$\omega \propto L(\omega)^{-D_f/d_s} . \tag{42}$$

For percolating networks, the spectral dimension is obtained from (40)

$$d_s = \frac{2D_f}{2 + \theta} = \frac{2\upsilon D_f}{2\upsilon + \mu - \beta} . \tag{43}$$

This exponent d_s is called the *fracton dimension* after S. Alexander and R. Orbach [1] or the *spectral dimension* after R. Rammal and G. Toulouse [17], hereafter we use the term *spectral dimension* for d_s. S. Alexander and R. Orbach [1] estimated the values of d_s for percolating networks on d-dimensional Euclidean lattices from the known values of the exponents D_f, υ, μ and β. They

pointed out that, while these exponents depend largely on d, the spectral dimension (fracton) dimension d_s does not.

The spectral dimension d_s can be obtained from the value of the conductivity exponent t or vice versa. In the case of percolating networks, the conductivity exponent t is related to d_s through (43), which means that the conductivity $\sigma_{dc} \sim (p - p_c)^t$ is also characterized by the spectral dimension d_s. In this sense, the spectral dimension d_s is an intrinsic exponent related to the dynamics of fractal systems. We can determine the precise values of d_s from the numerical calculations of the spectral density of states of percolation fractal networks.

The fracton SDOS for 2d, 3d, and 4d bond percolation networks at the percolation threshold $p = p_c$ are given in Fig. 9a and b, which were calculated by K. Yakubo and T. Nakayama in 1989. These were obtained by large-scale computer simulations [27]. At $p = p_c$, the correlation length diverges as $\xi \propto |p - p_c|^{-\nu}$ and the network has a fractal structure at any length scale. Therefore, fracton SDOS should be recovered in the wide frequency range $\omega_L \ll \omega \ll \omega_D$, where ω_D is the Debye cutoff frequency and ω_L is the lower cutoff determined by the system size. The SDOSs and the integrated SDOSs per atom are shown by the filled squares for a 2d bond percolation (abbreviated, BP) network at $p_c = 0.5$. The lowest frequency ω_L is quite small ($\omega \sim 10^{-5}$ for the 2d systems) as seen from the results in Fig. 9 because of the large sizes of the systems.

The spectral dimension d_s is obtained as $d_s = 1.33 \pm 0.11$ from Fig. 9a, whereas data in Fig. 9b give the more precise value $d_s = 1.325 \pm 0.002$. The SDOS and the integrated SDOS for 3d BP networks at $p_c = 0.249$ are given in Fig. 9a and b by the filled triangles (middle). The spectral dimension d_s is obtained as $d_s = 1.31 \pm 0.02$ from Fig. 9a and $d_s = 1.317 \pm 0.003$ from Fig. 9b. The SDOS and the integrated SDOS of 4d BP networks at $p_c = 0.160$.

A typical mode pattern of a fracton on a 2d percolation network is shown in Fig. 10a, where the eigenmode belongs to the angular frequency $\omega = 0.04997$. To bring out the details more clearly, Fig. 10b by K. Yakubo and T. Nakayama [28] shows cross-sections of this fracton mode along the line drawn in Fig. 10a. Filled and open circles represent occupied and vacant sites in the percolation network, respectively. We see that the fracton core (the largest amplitude) possesses very clear boundaries for the edges of the excitation, with an almost step-like character and a long tail in the direction of the weak segments. It should be noted that displacements of atoms in dead ends (weakly connected portions in the percolation network) move in phase, and fall off sharply at their edges.

The spectral dimension can be obtained exactly for deterministic fractals. In the case of the d-dimensional Sierpinski gasket, the spectral dimension is given by [17]

$$d_s = \frac{2 \log(d + 1)}{\log(d + 3)} .$$
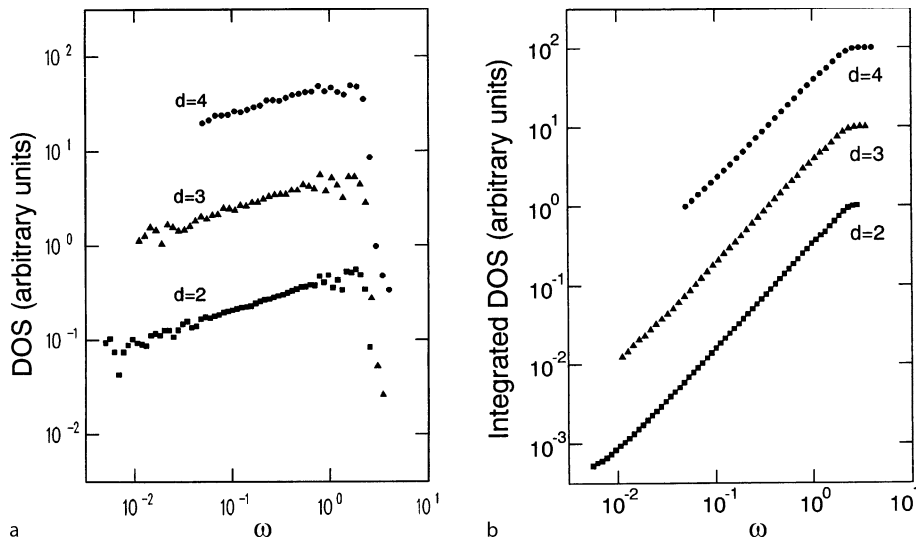
Fractal Structures in Condensed Matter Physics, Figure 9
a Spectral densities of states (SDOS) per atom for 2d, 3d, and 4d BP networks at $p = p_c$. The angular frequency ω is defined with mass units $m = 1$ and force constant $K_{ij} = 1$. The networks are formed on 1100 × 1100 (2d), 100 × 100 × 100 (3d), and 30 × 30 × 30 × 30 (4d) lattices with periodic boundary conditions, respectively. **b** Integrated densities of states for the same
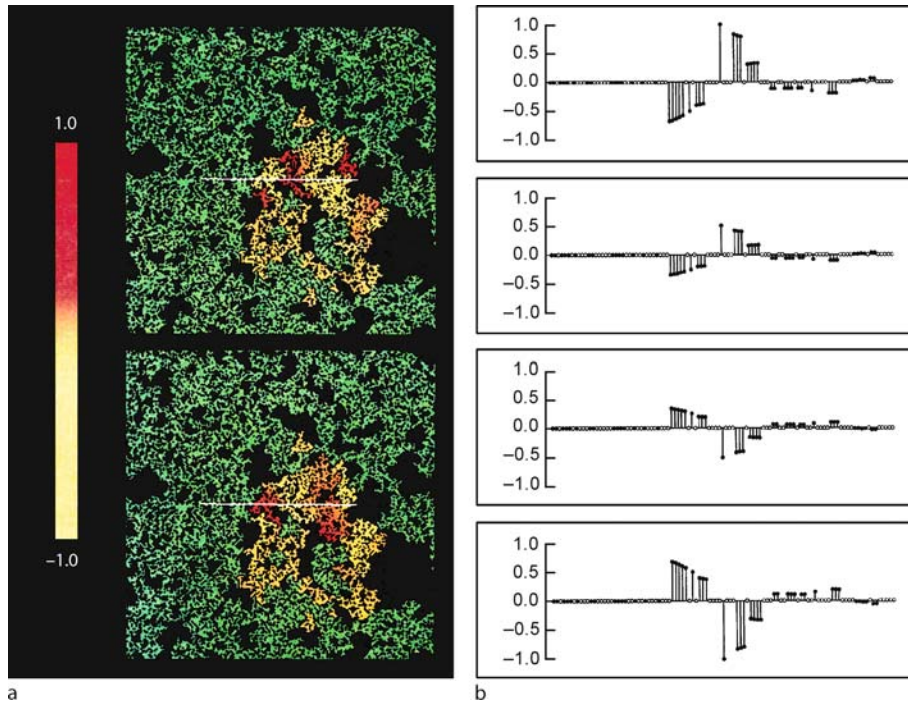
Fractal Structures in Condensed Matter Physics, Figure 10
a Typical fracton mode ($\omega = 0.04997$) on a 2d network. *Bright region* represents the large amplitude portion of the mode. **b** Cross-section of the fracton mode shown in **a** along the *white line*. The four figures are snapshots at different times. After [28]

We see from this that the upper bound for a Sierpinski gasket is $d_s = 2$ as $d \to \infty$. The spectral dimension for the Mandelbrot–Given fractal depicted is also calculated analytically as

$$d_s = \frac{2 \log 8}{\log 22} = 1.345 \ldots .$$

This value is close to those for percolating networks mentioned above, in addition to the fact that the fractal dimension $d_s = \log 8/\log 3$ of the Mandelbrot–Given fractal is close to $D_f = 91/48$ for 2d percolating networks and that the Mandelbrot–Given fractal has a structure with nodes, links, and blobs as in the case of percolating networks.

For real systems, E. Courtens et al. in 1988 [6] observed fracton excitations in aerogels by means of inelastic light scattering.

Future Directions

The significance of fractal researches in sciences is that the very idea of fractals opposes *reductionism*. Modern physics has developed by making efforts to elucidate the physical mechanisms of smaller and smaller structures such as molecules, atoms, and elementary particles. An example in condensed matter physics is the band theory of electrons in solids. Energy spectra of electrons can be obtained by incorporating group theory based on the translational and rotational symmetry of the systems. The use of this mathematical tool greatly simplifies the treatment of systems composed of 10^{22} atoms. If the energy spectrum of a unit cell *molecule* is solved, the whole energy spectrum of the solid can be computed by applying the group theory. In this context, the problem of an ordered solid is reduced to that of a unit cell. Weakly disordered systems can be handled by regarding impurities as a small perturbation to the corresponding ordered systems. However, a different approach is required for elucidating the physical properties of strongly disordered/complex systems with correlations, or of medium-scale objects, for which it is difficult to find an easily identifiable small parameter that would allow a perturbative analysis. For such systems, the concept of fractals plays an important role in building pictures of the realm of nature.

Our established knowledge on fractals is mainly due to experimental observations or computer simulations. The researches are at the phenomenological level stage, not at the intrinsic level, except for a few examples. Concerning future directions of the researches on fractals in condensed

matter physics apart from such a question as "What kinds of fractal structures are involved in condensed matter?", we should consider two directions: one is the very basic aspect such as the problem "Why are there numerous examples showing fractal structures in nature/condensed matter?" However, this type of question is hard. The kinetic growth-mechanisms of fractal systems have a rich variety of applications from the basic to applied sciences and attract much attention as one of the important subjects in non-equilibrium statistical physics and nonlinear physics. Network formation in society is one example where the kinetic growth is relevant. However, many aspects related to the mechanisms of network formations remain puzzling because arguments are at the phenomenological stage. If we compare with researches on Brownian motion as an example, the DLA researches need to advance to the stage of Einstein's intrinsic theory [8], or that of Smoluchowski [18] and H. Nyquist [15]. It is notable that the DLA is a stochastic version of the Hele–Shaw problem, the flow in composite fluids with high and low viscosities: the particles diffuse in the DLA, while the fluid pressure diffuses in Hele–Shaw flow [19]. These are deeply related to each other and involve *many open questions* for basic physics and mathematical physics.

Concerning the opposite direction, one of the important issues in fractal research is to explore practical uses of fractal structures. In fact, the characteristics of fractals are applied to many cases such as the formation of tailor-made nano-scale fractal structures, fractal-shaped antennae with much reduced sizes compared with those of ordinary antennae, and fractal molecules sensitive to frequencies in the infrared region of light.

Deep insights into fractal physics in condensed matter will open the door to new sciences and its application to technologies in the near future.

Bibliography

Primary Literature

1. Alexander S, Orbach R (1982) Density of states: Fractons. J Phys Lett 43:L625–631
2. Beeckmans JM (1963) The density of aggregated solid aerosol particles. Ann Occup Hyg 7:299–305
3. Ben-Avraham D, Havlin S (1982) Diffusion on percolation at criticality. J Phys A 15:L691–697
4. Brady RM, Ball RC (1984) Fractal growth of copper electro-deposits. Nature 309:225–229
5. Broadbent SR, Hammersley JM (1957) Percolation processes I: Crystals and mazes. Proc Cambridge Philos Soc 53:629–641
6. Courtens E, Vacher R, Pelous J, Woignier T (1988) Observation of fractons in silica aerogels. Europhys Lett 6:L691–697
7. de Gennes PG (1976) La percolation: un concept unificateur. Recherche 7:919–927
8. Einstein A (1905) Über die von der molekularkinetischen Theorie der Waerme geforderte Bewegung von in Ruhenden Fluessigkeiten Suspendierten Teilchen. Ann Phys 17:549–560
9. Gefen Y, Aharony A, Alexander S (1983) Anomalous diffusion on percolating clusters. Phys Rev Lett 50:70–73
10. Hausdorff F (1919) Dimension und Aeusseres Mass. Math Ann 79:157–179
11. Kolb M, Botel R, Jullien R (1983) Scaling of kinetically growing clusters. Phys Rev Lett 51:1123–1126
12. Mandelbrot BB, Given JA (1984) Physical properties of a new fractal model of percolation clusters. Phys Rev Lett 52:1853–1856
13. Matsushita M, et al (1984) Fractal structures of zinc metal leaves grown by electro-deposition. Phys Rev Lett 53:286–289
14. Meakin P (1983) Formation of fractal clusters and networks by irreversible diffusion-limited aggregation. Phys Rev Lett 51:1119–1122
15. Nyquist H (1928) Termal agitation nof electric charge in conductors. Phys Rev 32:110–113
16. Perrin J (1909) Movement brownien et realite moleculaire. Ann Chim Phys 19:5–104
17. Rammal R, Toulouze G (1983) Random walks on fractal structures and percolation clusters. J Phys Lett 44:L13–L22
18. Smoluchowski MV (1906) Zur Kinematischen Theorie der Brownschen Molekular Bewegung und der Suspensionen. Ann Phys 21:756–780
19. Saffman PG, Taylor GI (1959) The penetration of a fluid into a porous medium or hele-shaw cell containing a more viscous fluid. Proc Roy Soc Lond Ser A 245:312–329
20. Tambo N, Watanabe Y (1967) Study on the density profiles of aluminium flocs I (in japanese). Suidou Kyoukai Zasshi 397:2–10; ibid (1968) Study on the density profiles of aluminium flocs II (in japanese) 410:14–17
21. Tambo N, Watanabe Y (1979) Physical characteristics of flocs I: The floc density function and aluminium floc. Water Res 13:409–419
22. Vacher R, Woignier T, Pelous J, Courtens E (1988) Structure and self-similarity of silica aerogels. Phys Rev B 37:6500–6503
23. Vold MJ (1963) Computer simulation of floc formation in a colloidal suspension. J Colloid Sci 18:684–695
24. Weitz DA, Oliveria M (1984) Fractal structures formed by kinetic aggregation of aqueous cold colloids. Phys Rev Lett 52:1433–1436
25. Weitz DA, Huang JS, Lin MY, Sung J (1984) Dynamics of diffusion-limited kinetics aggregation. Phys Rev Lett 53:1657–1660
26. Witten TA, Sander LM (1981) Diffusion-limited aggregation, a kinetic critical phenomenon. Phys Rev Lett 47:1400–1403
27. Yakubo K, Nakayama T (1989) Direct observation of localized fractons excited on percolating nets. J Phys Soc Jpn 58:1504–1507
28. Yakubo K, Nakayama T (1989) Fracton dynamics of percolating elastic networks: energy spectrum and localized nature. Phys Rev B 40:517–523

Books and Reviews

Barabasi AL, Stanley HE (1995) Fractal concepts in surface growth. Cambridge University Press, Cambridge
Ben-Avraham D, Havlin S (2000) Diffusion and reactions in fractals and disordered systems. Cambridge University Press, Cambridge

Bunde A, Havlin S (1996) Fractals and disordered systems. Springer, New York

de Gennes PG (1979) Scaling concepts in polymer physics. Cornell University Press, Itahca

Falconer KJ (1989) Fractal geometry: Mathematical foundations and applications. Wiley, New York

Feder J (1988) Fractals. Plenum, New York

Flory PJ (1969) Statistical mechanics of chain molecules. Inter Science, New York

Halsey TC (2000) Diffusion-limited aggregation: A model for pattern formation. Phys Today 11:36–41

Kadanoff LP (1976) Domb C, Green MS (eds) Phase transitions and critical phenomena 5A. Academic Press, New York

Kirkpatrick S (1973) Percolation and conduction. Rev Mod Phys 45:574–588

Mandelbrot BB (1979) Fractals: Form, chance and dimension. Freeman, San Francisco

Mandelbrot BB (1982) The fractal geometry of nature. Freeman, San Francisco

Meakin P (1988) Fractal aggregates. Adv Colloid Interface Sci 28:249–331

Meakin P (1998) Fractals, scaling and growth far from equilibrium. Cambridge University Press, Cambridge

Nakayama T, Yakubo K, Orbach R (1994) Dynamical properties of fractal networks: Scaling, numerical simulations, and physical realizations. Rev Mod Phys 66:381–443

Nakayama T, Yakubo K (2003) Fractal concepts in condensed matter. Springer, Heidelberg

Sahimi M (1994) Applications of percolation theory. Taylor and Francis, London

Schroeder M (1991) Fractals, chaos, power laws. W.H. Freeman, New York

Stauffer D, Aharony A (1992) Introduction to percolation theory, 2nd edn. Taylor and Francis, London

Vicsek T (1992) Fractal growth phenomena, 2nd edn. World Scientific, Singapore

Fractals and Wavelets: What Can We Learn on Transcription and Replication from Wavelet-Based Multifractal Analysis of DNA Sequences?

Alain Arneodo[1], Benjamin Audit[1],
Edward-Benedict Brodie of Brodie[1],
Samuel Nicolay[2], Marie Touchon[3,5],
Yves d'Aubenton-Carafa[4], Maxime Huvet[4],
Claude Thermes[4]

[1] Laboratoire Joliot–Curie and Laboratoire de Physique, ENS-Lyon CNRS, Lyon Cedex, France

[2] Institut de Mathématique, Université de Liège, Liège, Belgium

[3] Génétique des Génomes Bactériens, Institut Pasteur, CNRS, Paris, France

[4] Centre de Génétique Moléculaire, CNRS, Gif-sur-Yvette, France

[5] Atelier de Bioinformatique, Université Pierre et Marie Curie, Paris, France

Article Outline

Glossary

Fractal Fractals are complex mathematical objects that are invariant with respect to dilations (**self-similarity**) and therefore do not possess a characteristic length scale. Fractal objects display scale-invariance properties that can either fluctuate from point to point (**multifractal**) or be homogeneous (**monofractal**). Mathematically, these properties should hold over all scales. However, in the real world, there are necessarily lower and upper bounds over which self-similarity applies.

Wavelet transform The continuous wavelet transform (WT) is a mathematical technique introduced in the early 1980s to perform time-frequency analysis. The WT has been early recognized as a mathematical microscope that is well adapted to characterize the scale-invariance properties of fractal objects and to reveal the hierarchy that governs the spatial distribution of the singularities of multifractal measures and functions. More specifically, the WT is a space-scale analysis which consists in expanding signals in terms of wavelets that are constructed from a single function, the analyzing wavelet, by means of translations and dilations.

Wavelet transform modulus maxima method The WTMM method provides a unified statistical (thermodynamic) description of multifractal distributions including measures and functions. This method relies on the computation of partition functions from the wavelet transform skeleton defined by the wavelet transform modulus maxima (WTMM). This skeleton provides an adaptive space-scale partition of the fractal distribution under study, from which one can extract the $D(h)$ singularity spectrum as the equivalent of a thermodynamic potential (entropy). With some appropriate choice of the analyzing wavelet, one can show that the WTMM method provides a natural generalization of the classical box-counting and structure function techniques.

Compositional strand asymmetry The DNA double helix is made of two strands that are maintained together by hydrogen bonds involved in the base-pairing between Adenine (resp. Guanine) on one strand and Thymine (resp. Cytosine) on the other strand. Under no-strand bias conditions, i. e. when mutation rates are identical on the two strands, in other words when the two strands are strictly equivalent, one expects equimolarities of adenine and thymine and of guanine and cytosine on each DNA strand, a property named Chargaff's second parity rule. Compositional strand asymmetry refers to deviations from this rule which can be assessed by measuring departure from intrastrand equimolarities. Note two major biological processes, **transcription** and **replication**, both requiring the opening of the double helix, actually break the symmetry between the two DNA strands and can thus be at the origin of compositional strand asymmetries.

Eukaryote Organisms whose cells contain a nucleus, the structure containing the genetic material arranged into chromosomes. Eukaryotes constitute one of the three domains of life, the two others, called prokaryotes (without nucleus), being the eubacteria and the archaebacteria.

Transcription Transcription is the process whereby the DNA sequence of a gene is enzymatically copied into a complementary messenger RNA. In a following step, **translation** takes place where each messenger RNA serves as a template to the biosynthesis of a specific protein.

Replication DNA replication is the process of making an identical copy of a double-stranded DNA molecule. DNA replication is an essential cellular function responsible for the accurate transmission of genetic information though successive cell generations. This process starts with the binding of initiating proteins to a DNA locus called **origin of replication**. The recruitment of additional factors initiates the bi-directional progression of two replication forks along the chromosome. In eukaryotic cells, this binding event happens at a multitude of replication origins along each chromosome from which replication propagates until two converging forks collide at a **terminus of replication**.

Chromatin Chromatin is the compound of DNA and proteins that forms the chromosomes in living cells. In eukaryotic cells, chromatin is located in the nucleus.

Histones Histones are a major family of proteins found in eukaryotic chromatin. The wrapping of DNA around a core of 8 histones forms a **nucleosome**, the first step of eukaryotic DNA compaction.

Definition of the Subject

The continuous wavelet transform (WT) is a mathematical technique introduced in signal analysis in the early 1980s [1,2]. Since then, it has been the subject of considerable theoretical developments and practical applications in a wide variety of fields. The WT has been early recognized as a mathematical microscope that is well adapted to reveal the hierarchy that governs the spatial distribution of singularities of multifractal measures [3,4,5]. What makes the WT of fundamental use in the present study is that its singularity scanning ability equally applies to singular functions than to singular measures [3,4,5,6,7, 8,9,10,11]. This has led Alain Arneodo and his collaborators [12,13,14,15,16] to elaborate a unified thermodynamic description of multifractal distributions including measures and functions, the so-called Wavelet Transform Modulus Maxima (WTMM) method. By using wavelets instead of boxes, one can take advantage of the freedom in the choice of these "generalized oscillating boxes" to get rid of possible (smooth) polynomial behavior that might either mask singularities or perturb the estimation of their strength h (Hölder exponent), remedying in this way for one of the main failures of the classical multifractal methods (e. g. the box-counting algorithms in the case of measures and the structure function method in the case of functions [12,13,15,16]). The other fundamental advantage of using wavelets is that the skeleton defined by the WTMM [10,11], provides an adaptive space-scale partitioning from which one can extract the $D(h)$ singularity spectrum via the Legendre transform of the scaling exponents $\tau(q)$ (q real, positive as well as negative) of some partition functions defined from the WT skeleton. We refer the reader to Bacry et al. [13], Jaffard [17,18] for rigorous

mathematical results and to Hentschel [19] for the theoretical treatment of random multifractal functions.

Applications of the WTMM method to 1D signals have already provided insights into a wide variety of problems [20], e. g., the validation of the log-normal cascade phenomenology of fully developed turbulence [21,22,23, 24] and of high-resolution temporal rainfall [25,26], the characterization and the understanding of long-range correlations in DNA sequences [27,28,29,30], the demonstration of the existence of causal cascade of information from large to small scales in financial time series [31,32], the use of the multifractal formalism to discriminate between healthy and sick heartbeat dynamics [33,34], the discovery of a Fibonacci structural ordering in 1D cuts of diffusion limited aggregates (DLA) [35,36,37,38]. The canonical WTMM method has been further generalized from 1D to 2D with the specific goal to achieve multifractal analysis of rough surfaces with fractal dimensions D_F anywhere between 2 and 3 [39,40,41]. The 2D WTMM method has been successfully applied to characterize the intermittent nature of satellite images of the cloud structure [42,43], to perform a morphological analysis of the anisotropic structure of atomic hydrogen (H_I) density in Galactic spiral arms [44] and to assist in the diagnosis in digitized mammograms [45]. We refer the reader to Arneodo et al. [46] for a review of the 2D WTMM methodology, from the theoretical concepts to experimental applications. In a recent work, Kestener and Arneodo [47] have further extended the WTMM method to 3D analysis. After some convincing test applications to synthetic 3D monofractal Brownian fields and to 3D multifractal realizations of singular cascade measures as well as their random function counterpart obtained by fractional integration, the 3D WTMM method has been applied to dissipation and enstrophy 3D numerical data issued from direct numerical simulations (DNS) of isotropic turbulence. The results so-obtained have revealed that the multifractal spatial structure of both dissipation and enstrophy fields are likely to be well described by a multiplicative cascade process clearly nonconservative. This contrasts with the conclusions of previous box-counting analysis [48] that failed to estimate correctly the corresponding multifractal spectra because of their intrinsic inability to master non-conservative singular cascade measures [47].

For many years, the multifractal description has been mainly devoted to scalar measures and functions. However, in physics as well as in other fundamental and applied sciences, fractals appear not only as deterministic or random scalar fields but also as vector-valued deterministic or random fields. Very recently, Kestener and Arneodo [49,50] have combined singular value decomposi-

tion techniques and WT analysis to generalize the multifractal formalism to vector-valued random fields. The so-called Tensorial Wavelet Transform Modulus Maxima (TWTMM) method has been applied to turbulent velocity and vorticity fields generated in $(256)^3$ DNS of the incompressible Navier–Stokes equations. This study reveals the existence of an intimate relationship $D_{\mathbf{v}}(h + 1) = D_{\boldsymbol{\omega}}(h)$ between the singularity spectra of these two vector fields that are found significantly more intermittent that previously estimated from longitudinal and transverse velocity increment statistics. Furthermore, thanks to the singular value decomposition, the TWTMM method looks very promising for future simultaneous multifractal and structural (vorticity sheets, vorticity filaments) analysis of turbulent flows [49,50].

Introduction

The possible relevance of scale invariance and fractal concepts to the structural complexity of genomic sequences has been the subject of considerable increasing interest [20,51,52]. During the past fifteen years or so, there has been intense discussion about the existence, the nature and the origin of the long-range correlations (LRC) observed in DNA sequences. Different techniques including mutual information functions [53,54], auto-correlation functions [55,56], power-spectra [54,57,58], "DNA walk" representation [52,59], Zipf analysis [60,61] and entropies [62,63], were used for the statistical analysis of DNA sequences. For years there has been some permanent debate on rather struggling questions like the fact that the reported LRC might be just an artifact of the compositional heterogeneity of the genome organization [20,27,52, 55,56,64,65,66,67]. Another controversial issue is whether or not LRC properties are different for protein-coding (exonic) and non-coding (intronic, intergenic) sequences [20, 27,52,54,55,56,57,58,59,61,68]. Actually, there were many objective reasons for this somehow controversial situation. Most of the pioneering investigations of LRC in DNA sequences were performed using different techniques that all consisted in measuring power-law behavior of some characteristic quantity, e. g., the fractal dimension of the DNA walk, the scaling exponent of the correlation function or the power-law exponent of the power spectrum. Therefore, in practice, they all faced the same difficulties, namely finite-size effects due to the finiteness of the sequence [69,70,71] and statistical convergence issue that required some precautions when averaging over many sequences [52,65]. But beyond these practical problems, there was also a more fundamental restriction since the measurement of a unique exponent characterizing the

global scaling properties of a sequence failed to resolve multifractality [27], and thus provided very poor information upon the nature of the underlying LRC (if they were any). Actually, it can be shown that for a homogeneous (monofractal) DNA sequence, the scaling exponents estimated with the techniques previously mentioned, can all be expressed as a function of the so-called Hurst or roughness exponent H of the corresponding DNA walk landscape [20,27,52]. $H = 1/2$ corresponds to classical Brownian, i. e. uncorrelated random walk. For any other value of H, the steps (increments) are either positively correlated ($H > 1/2$: Persistent random walk) or anti-correlated ($H < 1/2$: Anti-persistent random walk).

One of the main obstacles to LRC analysis in DNA sequences is the genuine mosaic structure of these sequences which are well known to be formed of "patches" of different underlying composition [72,73,74]. When using the "DNA walk" representation, these patches appear as trends in the DNA walk landscapes that are likely to break scale-invariance [20,52,59,64,65,66,67,75,76]. Most of the techniques, e. g. the variance method, used for characterizing the presence of LRC are not well adapted to study non-stationary sequences. There have been some phenomenological attempts to differentiate local patchiness from LRC using ad hoc methods such as the so-called "min-max method" [59] and the "detrended fluctuation analysis" [77]. In previous works [27,28], the WT has been emphasized as a well suited technique to overcome this difficulty. By considering analyzing wavelets that make the WT microscope blind to low-frequency trends, any bias in the DNA walk can be removed and the existence of power-law correlations with specific scale invariance properties can be revealed accurately. In [78], from a systematic WT analysis of human exons, CDSs and introns, LRC were found in non-coding sequences as well as in coding regions somehow hidden in their inner codon structure. These results made rather questionable the model based on genome plasticity proposed at that time to account for the reported absence of LRC in coding sequences [27,28,52,54,59,68]. More recently, some structural interpretation of these LRC has emerged from a comparative multifractal analysis of DNA sequences using structural coding tables based on nucleosome positioning data [29,30]. The application of the WTMM method has revealed that the corresponding DNA chain bending profiles are monofractal (homogeneous) and that there exists two LRC regimes. In the 10–200 bp range, LRC are observed for eukaryotic sequences as quantified by a Hurst exponent value $H \simeq 0.6$ (but not for eubacterial sequences for which $H = 0.5$) as the signature of the nucleosomal structure. These LRC were shown to favor the autonomous

formation of small (a few hundred bps) 2D DNA loops and in turn the propensity of eukaryotic DNA to interact with histones to form nucleosomes [79,80]. In addition, these LRC might induce some local hyperdiffusion of these loops which would be a very attractive interpretation of the nucleosomal repositioning dynamics. Over larger distances ($\gtrsim 200$ bp), stronger LRC with $H \simeq 0.8$ seem to exist in any sequence [29,30]. These LRC are actually observed in the *S. cerevisiae* nucleosome positioning data [81] suggesting that they are involved in the nucleosome organization in the so-called 30 nm chromatin fiber [82]. The fact that this second regime of LRC is also present in eubacterial sequences shows that it is likely to be a possible key to the understanding of the structure and dynamics of both eukaryotic and prokaryotic chromatin fibers. In regards to their potential role in regulating the hierarchical structure and dynamics of chromatin, the recent report [83] of sequence-induced LRC effects on the conformations of naked DNA molecules deposited onto mica surface under 2D thermodynamic equilibrium observed by Atomic Force Microscopy (AFM) is a definite experimental breakthrough.

Our purpose here is to take advantage of the availability of fully sequenced genomes to generalize the application of the WTMM method to genome-wide multifractal sequence analysis when using codings that have a clear functional meaning. According to the second parity rule [84,85], under no strand-bias conditions, each genomic DNA strand should present equimolarities of adenines A and thymines T and of guanines G and cytosines C [86,87]. Deviations from intrastrand equimolarities have been extensively studied during the past decade and the observed skews have been attributed to asymmetries intrinsic to the replication and transcription processes that both require the opening of the double helix. Actually, during these processes mutational events can affect the two strands differently and an asymmetry can result if one strand undergoes different mutations, or is repaired differently than the other strand. The existence of transcription and/or replication associated strand asymmetries has been mainly established for prokaryote, organelle and virus genomes [88,89,90,91,92,93,94]. For a long time the existence of compositional biases in eukaryotic genomes has been unclear and it is only recently that (i) the statistical analysis of eukaryotic gene introns have revealed the presence of transcription-coupled strand asymmetries [95,96,97] and (ii) the genome wide multiscale analysis of mammalian genomes has clearly shown some departure from intrastrand equimolarities in intergenic regions and further confirmed the existence of replication-associated strand asymmetries [98,99,100]. In this

manuscript, we will review recent results obtained when using the WT microscope to explore the scale invariance properties of the TA and GC skew profiles in the 22 human autosomes [98,99,100]. These results will enlighten the richness of information that can be extracted from these functional codings of DNA sequences including the prediction of 1012 putative human replication origins. In particular, this study will reveal a remarkable human gene organization driven by the coordination of transcription and replication [101].

A Wavelet-Based Multifractal Formalism

The Continuous Wavelet Transform

The WT is a space-scale analysis which consists in expanding signals in terms of *wavelets* which are constructed from a single function, the *analyzing wavelet* ψ, by means of translations and dilations. The WT of a real-valued function f is defined as [1,2]:

$$T_\psi[f](x_0, a) = \frac{1}{a} \int_{-\infty}^{+\infty} f(x)\, \psi\left(\frac{x - x_0}{a}\right) dx,$$ (1)

where x_0 is the space parameter and $a\ (> 0)$ the scale parameter. The analyzing wavelet ψ is generally chosen to be well localized in both space and frequency. Usually ψ is required to be of zero mean for the WT to be invertible. But for the particular purpose of singularity tracking that is of interest here, we will further require ψ to be orthogonal to low-order polynomials in Fig. 1 [7,8,9,10,11,12,13, 14,15,16]:

$$\int_{-\infty}^{+\infty} x^m \psi(x) dx, \quad 0 \le m < n_\psi.$$ (2)

As originally pointed out by Mallat and collaborators [10,11], for the specific purpose of analyzing the regularity of a function, one can get rid of the redundancy of the WT by concentrating on the WT skeleton defined by its modulus maxima only. These maxima are defined, at each scale a, as the local maxima of $|T_\psi[f](x, a)|$ considered as a function of x. As illustrated in Figs. 2e, 2f, these WTMM are disposed on connected curves in the space-scale (or time-scale) half-plane, called *maxima lines*. Let us define $\mathcal{L}(a_0)$ as the set of all the maxima lines that exist at the scale a_0 and which contain maxima at any scale $a \le a_0$. An important feature of these maxima lines, when analyzing singular functions, is that there is at least one maxima line pointing towards each singularity [10,11,16].

Scanning Singularities with the Wavelet Transform Modulus Maxima

The strength of the singularity of a function f at point x_0 is given by the *Hölder* exponent, i. e., the largest exponent such that there exists a polynomial $P_n(x - x_0)$ of order $n < h(x_0)$ and a constant $C > 0$, so that for any point x in a neighborhood of x_0, one has [7,8,9,10,11,13,16]:

$$|f(x) - P_n(x - x_0)| \le C\,|x - x_0|^h.$$ (3)

If f is n times continuously differentiable at the point x_0, then one can use for the polynomial $P_n(x - x_0)$, the order-n Taylor series of f at x_0 and thus prove that $h(x_0) > n$. Thus $h(x_0)$ measures how irregular the function f is at the point x_0. The higher the exponent $h(x_0)$, the more regular the function f.

The main interest in using the WT for analyzing the regularity of a function lies in its ability to be blind to polynomial behavior by an appropriate choice of the analyzing wavelet ψ. Indeed, let us assume that according to Eq. (3), f has, at the point x_0, a local scaling (Hölder) exponent $h(x_0)$; then, assuming that the singularity is not oscillating [11,102,103], one can easily prove that the local behavior of f is mirrored by the WT which locally behaves like [7,8,9,10,11,12,13,14,15,16,17,18]:

$$T_\psi[f](x_0, a) \sim a^{h(x_0)}, \quad a \to 0^+,$$ (4)

provided $n_\psi > h(x_0)$, where n_ψ is the number of vanishing moments of ψ (Eq. (2)). Therefore one can extract the exponent $h(x_0)$ as the slope of a log-log plot of the WT amplitude versus the scale a. On the contrary, if one chooses $n_\psi < h(x_0)$, the WT still behaves as a power-law but with a scaling exponent which is n_ψ:

$$T_\psi[f](x_0, a) \sim a^{n_\psi}, \quad a \to 0^+.$$ (5)

Thus, around a given point x_0, the faster the WT decreases when the scale goes to zero, the more regular f is around that point. In particular, if $f \in C^\infty$ at x_0 ($h(x_0) = +\infty$), then the WT scaling exponent is given by n_ψ, i. e. a value which is dependent on the shape of the analyzing wavelet. According to this observation, one can hope to detect the points where f is smooth by just checking the scaling behavior of the WT when increasing the order n_ψ of the analyzing wavelet [12,13,14,15,16].

Remark 1 A very important point (at least for practical purpose) raised by Mallat and Hwang [10] is that the local scaling exponent $h(x_0)$ can be equally estimated by looking at the value of the WT modulus along a maxima line converging towards the point x_0. Indeed one can prove that both Eqs. (4) and (5) still hold when following a maxima line from large down to small scales [10,11].

The Wavelet Transform Modulus Maxima Method

As originally defined by Parisi and Frisch [104], the multifractal formalism of multi-affine functions amounts to compute the so-called *singularity spectrum* $D(h)$ defined as the Hausdorff dimension of the set where the Hölder exponent is equal to h [12,13,16]:

$$D(h) = \dim_H \{x, \ h(x) = h\} \ , \qquad (6)$$

where h can take, a priori, positive as well as negative real values (e. g., the Dirac distribution $\delta(x)$ corresponds to the Hölder exponent $h(0) = -1$) [17].

A natural way of performing a multifractal analysis of fractal functions consists in generalizing the "classical" multifractal formalism [105,106,107,108,109] using wavelets instead of boxes. By taking advantage of the freedom in the choice of the "generalized oscillating boxes" that are the wavelets, one can hope to get rid of possible smooth behavior that could mask singularities or perturb the estimation of their strength h. But the major difficulty with respect to box-counting techniques [48,106,110,111, 112] for singular measures, consists in defining a covering of the support of the singular part of the function with our set of wavelets of different sizes. As emphasized in [12,13, 14,15,16], the branching structure of the WT skeletons of fractal functions in the (x, a) half-plane enlightens the hierarchical organization of their singularities (Figs. 2e, 2f). The WT skeleton can thus be used as a guide to position, at a considered scale a, the oscillating boxes in order to obtain a partition of the singularities of f. The wavelet transform modulus maxima (WTMM) method amounts to compute the following partition function in terms of WTMM coefficients [12,13,14,15,16]:

$$Z(q, a) = \sum_{l \in \mathcal{L}(a)} \left(\sup_{\substack{(x, a') \in l \\ a' \leq a}} \left| T_\psi[f](x, a') \right| \right)^q \ , \qquad (7)$$

where $q \in \mathbb{R}$ and the sup can be regarded as a way to define a scale adaptative "Hausdorff-like" partition. Now from the deep analogy that links the multifractal formalism to thermodynamics [12,113], one can define the exponent $\tau(q)$ from the power-law behavior of the partition function:

$$Z(q, a) \sim a^{\tau(q)} \ , \quad a \to 0^+ \ , \qquad (8)$$

where q and $\tau(q)$ play respectively the role of the inverse temperature and the free energy. The main result of this wavelet-based multifractal formalism is that in place of the energy and the entropy (i. e. the variables conjugated to q and τ), one has h, the Hölder exponent, and $D(h)$, the singularity spectrum. This means that the singularity spectrum of f can be determined from the Legendre transform

of the partition function scaling exponent $\tau(q)$ [13,17,18]:

$$D(h) = \min_q (qh - \tau(q)) \ . \qquad (9)$$

From the properties of the Legendre transform, it is easy to see that *homogeneous* monofractal functions that involve singularities of unique Hölder exponent $h = \partial\tau/\partial q$, are characterized by a $\tau(q)$ spectrum which is a *linear* function of q. On the contrary, a *nonlinear* $\tau(q)$ curve is the signature of nonhomogeneous functions that exhibit *multifractal* properties, in the sense that the Hölder exponent $h(x)$ is a fluctuating quantity that depends upon the spatial position x.

Defining Our Battery of Analyzing Wavelets

There are almost as many analyzing wavelets as applications of the continuous WT [3,4,5,12,13,14,15,16]. In the present work, we will mainly used the class of analyzing wavelets defined by the successive derivatives of the Gaussian function:

$$g^{(N)}(x) = \frac{\mathrm{d}^N}{\mathrm{d}x^N} e^{-x^2/2} \ , \qquad (10)$$

for which $n_\psi = N$ and more specifically $g^{(1)}$ and $g^{(2)}$ that are illustrated in Figs. 1a, 1b.

Remark 2 The WT of a signal f with $g^{(N)}$ (Eq. (10)) takes the following simple expression:

$$T_{g^{(N)}}[f](x, a) = \frac{1}{a} \int_{-\infty}^{+\infty} f(y) g^{(N)} \left(\frac{y - x}{a} \right) \mathrm{d}y \ ,$$

$$= a^N \frac{\mathrm{d}^N}{\mathrm{d}x^N} T_{g^{(0)}}[f](x, a) \ . \qquad (11)$$

Equation (11) shows that the WT computed with $g^{(N)}$ at scale a is nothing but the Nth derivative of the signal $f(x)$ smoothed by a dilated version $g^{(0)}(x/a)$ of the Gaussian function. This property is at the heart of various applications of the WT microscope as a very efficient multi-scale singularity tracking technique [20].

With the specific goal of disentangling the contributions to the nucleotide composition strand asymmetry coming respectively from transcription and replication processes, we will use in Sect. "A Wavelet-Based Methodology to Disentangle Transcription- and Replication-Associated Strand Asymmetries Reveals a Remarkable Gene Organization in the Human Genome", an adapted analyzing wavelet of the following form (Fig. 1c) [101,114]:

$$\phi_R(x) = -\left(x - \frac{1}{2}\right) \ , \quad \text{for } x \in \left[-\frac{1}{2}, \frac{1}{2}\right]$$

$$= 0 \quad \text{elsewhere} \ . \qquad (12)$$

F
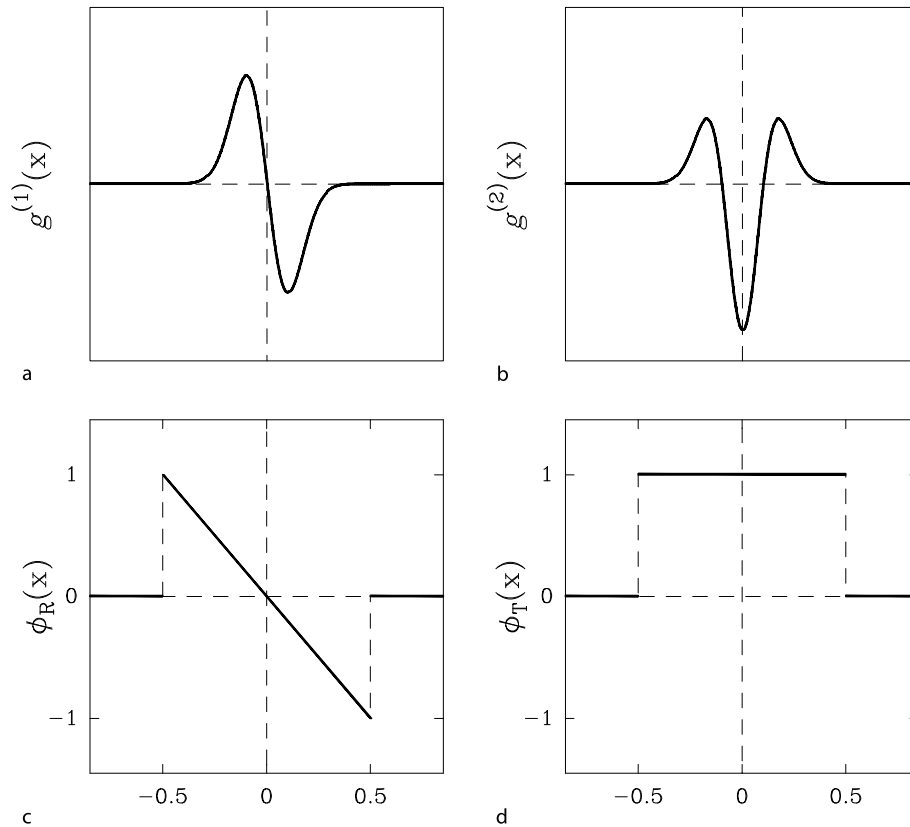
Fractals and Wavelets: What Can We Learn on Transcription and Replication …?, Figure 1
Set of analyzing wavelets $\psi(x)$ that can be used in Eq. (1). **a** $g^{(1)}$ and **b** $g^{(2)}$ as defined in Eq. (10). **c** ϕ_R as defined in Eq. (12), that will be used in Sect. "**A Wavelet-Based Methodology to Disentangle Transcription- and Replication-Associated Strand Asymmetries Reveals a Remarkable Gene Organization in the Human Genome**" to detect replication domains. **d** Box function ϕ_T that will be used in Sect. "**A Wavelet-Based Methodology to Disentangle Transcription- and Replication-Associated Strand Asymmetries Reveals a Remarkable Gene Organization in the Human Genome**" to model step-like skew profiles induced by transcription

By performing multi-scale pattern recognition in the (space, scale) half-plane with this analyzing wavelet, we will be able to define replication domains bordered by putative replication origins in the human genome and more generally in mammalian genomes [101,114].

Test Applications of the WTMM Method on Monofractal and Multifractal Synthetic Random Signals

This section is devoted to test applications of the WTMM method to random functions generated either by *additive* models like fractional Brownian motions [115] or by *multiplicative* models like random $\mathcal{W}$-cascades on wavelet dyadic trees [21,22,116,117]. For each model, we first wavelet transform 1000 realizations of length $L = 65\,536$ with the first order ($n_\psi = 1$) analyzing wavelet $g^{(1)}$. From the WT skeletons defined by the WTMM, we compute the mean partition function (Eq. (7)) from which we extract the annealed $\tau(q)$ (Eq. (8)) and, in turn, $D(h)$ (Eq. (9)) multifractal spectra. We systematically test the robustness of our estimates with respect to some change of the shape of the analyzing wavelet, in particular when increasing the number n_ψ of zero moments, going from $g^{(1)}$ to $g^{(2)}$ (Eq. (10)).

Fractional Brownian Signals   Since its introduction by Mandelbrot and van Ness [115], the fractional Brownian motion (fBm) B_H has become a very popular model in signal and image processing [16,20,39]. In 1D, fBm has proved useful for modeling various physical phenomena with long-range dependence, e. g., "$1/f$" noises. The fBm exhibits a power spectral density $S(k) \sim 1/k^\beta$, where the spectral exponent $\beta = 2H + 1$ is related to the Hurst exponent H. fBm has been extensively used as test stochas-
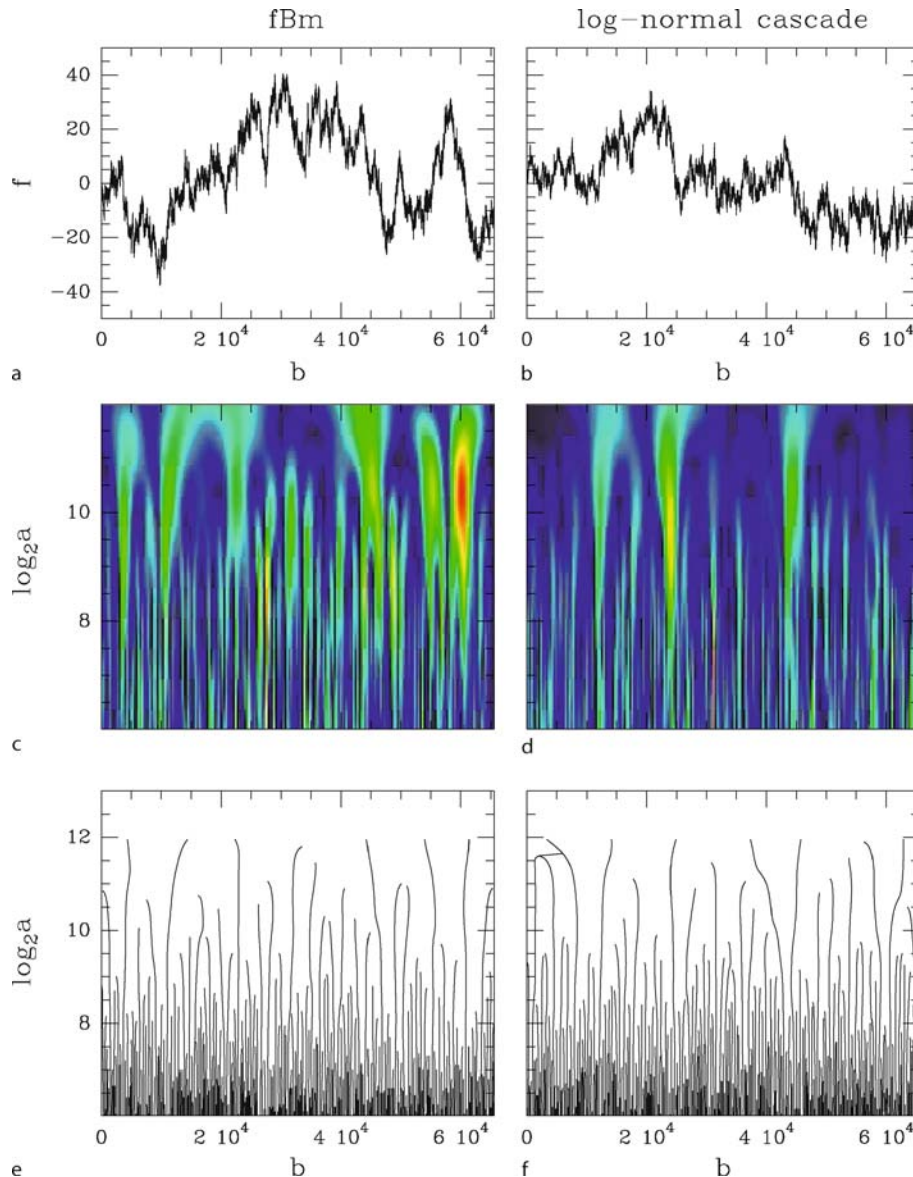
Fractals and Wavelets: What Can We Learn on Transcription and Replication ...?, Figure 2
WT of monofractal and multifractal stochastic signals. *Fractional Brownian motion*: **a** a realization of $B_{1/3}$ ($L = 65\,536$); **c** WT of $B_{1/3}$ as coded, independently at each scale a, using 256 colors from *black* ($|T_\psi| = 0$) to *red* ($\max_b |T_\psi|$); **e** WT skeleton defined by the set of all the maxima lines. *Log-normal random $\mathcal{W}$-cascades*: **b** a realization of the log-normal $\mathcal{W}$-cascade model ($L = 65\,536$) with the following parameter values $m = -0.355\ln 2$ and $\sigma^2 = 0.02\ln 2$ (see [116]); **d** WT of the realization in **b** represented with the same color coding as in **c**; **f** WT skeleton. The analyzing wavelet is $g^{(1)}$ (see Fig. 1a)

tic signals for Hurst exponent measurements. In Figs. 2, 3 and 4, we report the results of a statistical analysis of fBm's using the WTMM method [12,13,14,15,16]. We mainly concentrate on $B_{1/3}$ since it has a $k^{-5/3}$ power-spectrum similar to the spectrum of the multifractal stochastic signal we will study next. Actually, our goal is to demon-

strate that, where the power spectrum analysis fails, the WTMM method succeeds in discriminating unambiguously between these two fractal signals. The numerical signals were generated by filtering uniformly generated pseudo-random noise in Fourier space in order to have the required $k^{-5/3}$ spectral density. A $B_{1/3}$ fractional Brownian
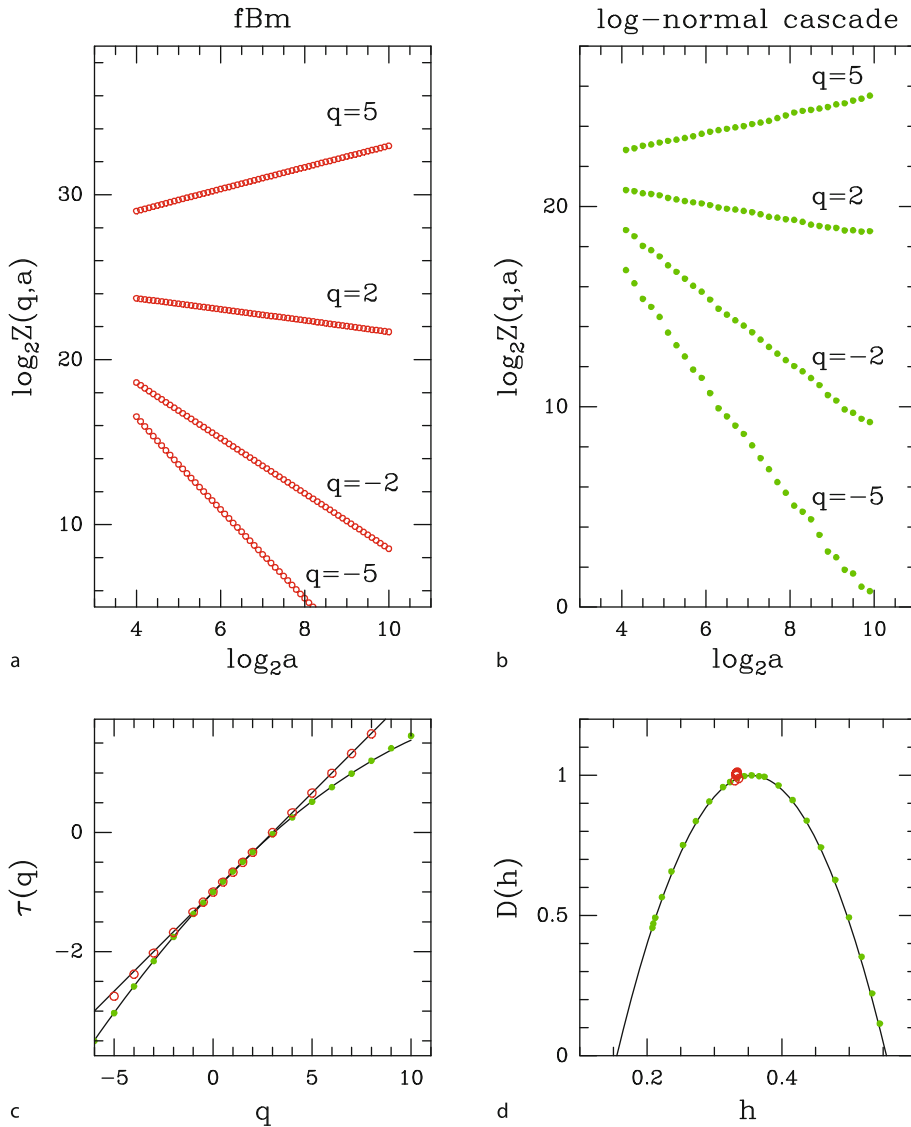
Fractals and Wavelets: What Can We Learn on Transcription and Replication …?, Figure 3
Determination of the $\tau(q)$ and $D(h)$ multifractal spectra of fBm $B_{1/3}$ (*red circles*) and log-normal random $\mathcal{W}$-cascades (*green dots*) using the WTMM method. **a** $\log_2 Z(q, a)$ vs. $\log_2 a$: $B_{1/3}$. **b** $\log_2 Z(q, a)$ vs. $\log_2 a$: Log-normal $\mathcal{W}$-cascades with the same parameters as in Fig. 2b. **c** $\tau(q)$ vs. q; the *solid lines* correspond respectively to the theoretical spectra (13) and (16). **d** $D(h)$ vs. h; the *solid lines* correspond respectively to the theoretical predictions (14) and (17). The analyzing wavelet is $g^{(1)}$. The reported results correspond to annealed averaging over 1000 realizations of $L = 65\,536$

trail is shown in Fig. 2a. Figure 2c illustrates the WT coded, independently at each scale a, using 256 colors. The analyzing wavelet is $g^{(1)}$ ($n_\psi = 1$). Figure 3a displays some plots of $\log_2 Z(q, a)$ versus $\log_2(a)$ for different values of q, where the partition function $Z(q, a)$ has been computed on the WTMM skeleton shown in Fig. 2e, according to the definition (Eq. (7)). Using a linear regression fit, we then obtain the slopes $\tau(q)$ of these graphs. As shown in Fig. 3c,

when plotted versus q, the data for the exponents $\tau(q)$ consistently fall on a straight line that is remarkably fitted by the theoretical prediction:

$$\tau(q) = qH - 1 , \tag{13}$$

with $H = 1/3$. From the Legendre transform of this linear $\tau(q)$ (Eq. (9)), one gets a $D(h)$ singularity spectrum that
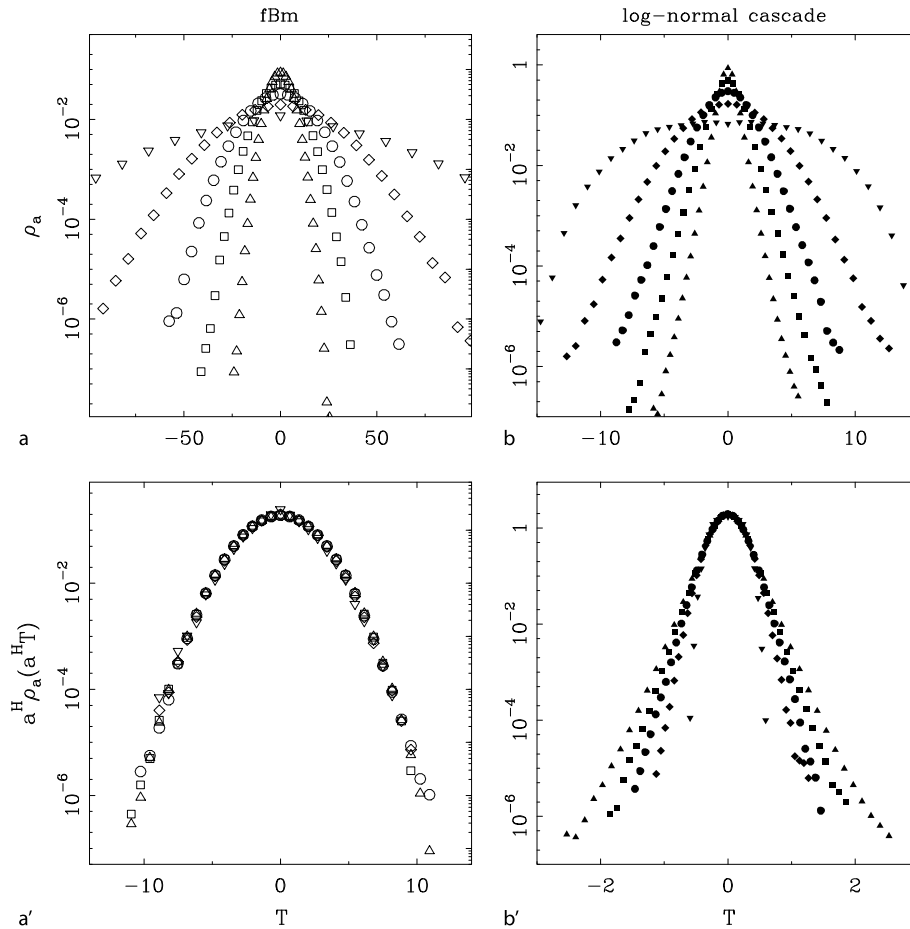
Fractals and Wavelets: What Can We Learn on Transcription and Replication …?, Figure 4
Probability distribution functions of wavelet coefficient values of fBm $B_{1/3}$ (*open symbols*) and log-normal random $\mathcal{W}$-cascades (*filled symbols*) with the same parameters as in Fig. 2b. **a** ρ_a vs. $T_{g^{(1)}}$ for the set of scales $a = 10$ ($\triangle$), 50 ($\square$), 100 ($\bigcirc$), 1000 ($\diamond$), 9000 ($\triangledown$); **a′** $a^H \rho_a(a^H T_{g^{(1)}})$ vs. $T_{g^{(1)}}$ with $H = 1/3$; The symbols have the same meaning as in **a**. **b** ρ_a vs. $T_{g^{(1)}}$ for the set of scales $a = 10$ ($\blacktriangle$), 50 ($\blacksquare$), 100 ($\bullet$), 1000 ($\blacklozenge$), 9000 ($\blacktriangledown$); (**b′**) $a^H \rho_a(a^H T_{g^{(1)}})$ vs. $T_{g^{(1)}}$ with $H = -m/\ln 2 = 0.355$. The analyzing wavelet is $g^{(1)}$ (Fig. 1a)

reduces to a single point:

$$
\begin{aligned}
D(h) &= 1 && \text{if } h = H, \\
&= -\infty && \text{if } h \neq H.
\end{aligned} \tag{14}
$$

Thus, as expected theoretically [16,115], one finds that the fBm $B_{1/3}$ is a nowhere differentiable homogeneous fractal signal with a unique Hölder exponent $h = H = 1/3$. Note that similar good estimates are obtained when using analyzing wavelets of different order (e. g. $g^{(2)}$), and this whatever the value of the index H of the fBm [12,13,14,15,16].

Within the perspective of confirming the monofractality of fBm's, we have studied the probability density function (pdf) of wavelet coefficient values $\rho_a(T_{g^{(1)}}(., a))$, as computed at a fixed scale a in the fractal scaling range. According to the monofractal scaling properties, one expects

these pdfs to satisfy the self-similarity relationship [20, 27,28]:

$$
a^H \rho_a(a^H T) = \rho(T), \tag{15}
$$

where $\rho(T)$ is a "universal" pdf (actually the pdf obtained at scale $a = 1$) that does not depend on the scale parameter a. As shown in Figs. 4a, 4a′ for $B_{1/3}$, when plotting $a^H \rho_a(a^H T)$ vs. T, all the ρ_a curves corresponding to different scales (Fig. 4a) remarkably collapse on a unique curve when using a unique exponent $H = 1/3$ (Fig. 4a′). Furthermore the so-obtained universal curve cannot be distinguished from a parabola in semi-log representation as the signature of the monofractal Gaussian statistics of fBm fluctuations [16,20,27].

Random $\mathcal{W}$-Cascades Multiplicative cascade models have enjoyed increasing interest in recent years as the paradigm of multifractal objects [16,19,48,105,107, 108,118]. The notion of cascade actually refers to a self-similar process whose properties are defined multiplicatively from coarse to fine scales. In that respect, it occupies a central place in the statistical theory of turbulence [48, 104]. Originally, the concept of self-similar cascades was introduced to model multifractal measures (e. g. dissipation or enstrophy) [48]. It has been recently generalized to the construction of scale-invariant signals (e. g. longitudinal velocity, pressure, temperature) using orthogonal wavelet basis [116,119]. Instead of redistributing the measure over sub-intervals with multiplicative weights, one allocates the wavelet coefficients in a multiplicative way on the dyadic grid. This method has been implemented to generate multifractal functions (with weights W) from a given deterministic or probabilistic multiplicative process. Along the line of the modeling of fully developed turbulent signals by log-infinitely divisible multiplicative processes [120,121], we will mainly concentrate here on the log-normal $\mathcal{W}$-cascades in order to calibrate the WTMM method. If m and σ^2 are respectively the mean and the variance of $\ln W$ (where W is a multiplicative random variable with log-normal probability distribution), then, as shown in [116], a straightforward computation leads to the following $\tau(q)$ spectrum:

$$
\begin{aligned}
\tau(q) &= -\log_2\langle W^q\rangle - 1\,, \quad \forall q \in \mathbb{R} \\
&= -\frac{\sigma^2}{2\ln 2}q^2 - \frac{m}{\ln 2}q - 1\,,
\end{aligned}
\tag{16}
$$

where $\langle\ldots\rangle$ means ensemble average. The corresponding $D(h)$ singularity spectrum is obtained by Legendre transforming $\tau(q)$ (Eq. (9)):

$$
D(h) = -\frac{(h + m/\ln 2)^2}{2\sigma^2/\ln 2} + 1\,.
\tag{17}
$$

According to the convergence criteria established in [116], m and σ^2 have to satisfy the conditions: $m < 0$ and $|m|/\sigma > \sqrt{2\ln 2}$. Moreover, by solving $D(h) = 0$, one gets the following bounds for the support of the $D(h)$ singularity spectrum: $h_{\min} = -m/\ln 2 - (\sqrt{2}\sigma)/\sqrt{\ln 2}$ and $h_{\max} = -m/\ln 2 + (\sqrt{2}\sigma)/\sqrt{\ln 2}$.

In Fig. 2b is illustrated a realization of a log-normal $\mathcal{W}$-cascade for the parameter values $m = -0.355\ln 2$ and $\sigma^2 = 0.02\ln 2$. The corresponding WT and WT skeleton as computed with $g^{(1)}$ are shown in Figs. 2d and 2f respectively. The results of the application of the WTMM method are reported in Fig. 3. As shown in Fig. 3b, when plotted versus the scale parameter a in a logarithmic rep-

resentation, the annealed average of the partition functions $Z(q, a)$ displays a well defined scaling behavior over a range of scales of about 5 octaves. Note that scaling of quite good quality is found for a rather wide range of q values: $-5 \leq q \leq 10$. When processing to a linear regression fit of the data over the first four octaves, one gets the $\tau(q)$ spectrum shown in Fig. 3c. This spectrum is clearly a nonlinear function of q, the hallmark of multifractal scaling. Moreover, the numerical data are in remarkable agreement with the theoretical quadratic prediction (Eq. (16)). Similar quantitative agreement is observed on the $D(h)$ singularity spectrum in Fig. 3d which displays a single humped parabola shape that characterizes intermittent fluctuations corresponding to Hölder exponents values ranging from $h_{\min} = 0.155$ to $h_{\max} = 0.555$. Unfortunately, to capture the strongest and the weakest singularities, one needs to compute the $\tau(q)$ spectrum for very large values of $|q|$. This requires the processing of many more realizations of the considered log-normal random $\mathcal{W}$-cascade. The multifractal nature of log-normal $\mathcal{W}$-cascade realizations is confirmed in Figs. 4b, 4b′ where the self-similarity relationship (Eq. (15)) is shown not to apply. Actually there does not exist a H value allowing to superimpose onto a single curve the WT pdfs computed at different scales.

The test applications reported in this section demonstrate the ability of the WTMM method to resolve multifractal scaling of 1D signals, a hopeless task for classical power spectrum analysis. They were used on purpose to calibrate and to test the reliability of our methodology, and of the corresponding numerical tools, with respect to finite-size effects and statistical convergence.

Bifractality of Human DNA Strand-Asymmetry Profiles Results from Transcription

During genome evolution, mutations do not occur at random as illustrated by the diversity of the nucleotide substitution rate values [122,123,124,125]. This non-randomness is considered as a by-product of the various DNA mutation and repair processes that can affect each of the two DNA strands differently. Asymmetries of substitution rates coupled to transcription have been mainly observed in prokaryotes [88,89,91], with only preliminary results in eukaryotes. In the human genome, excess of T was observed in a set of gene introns [126] and some large-scale asymmetry was observed in human sequences but they were attributed to replication [127]. Only recently, a comparative analysis of mammalian sequences demonstrated a transcription-coupled excess of G+T over A+C in the coding strand [95,96,97]. In contrast to the substitution

biases observed in bacteria presenting an excess of C→T transitions, these asymmetries are characterized by an excess of purine (A→G) transitions relatively to pyrimidine (T→C) transitions. These might be a by-product of the transcription-coupled repair mechanism acting on uncorrected substitution errors during replication [128]. In this section, we report the results of a genome-wide multifractal analysis of strand-asymmetry DNA walk profiles in the human genome [129]. This study is based on the computation of the TA and GC skews in non-overlapping 1 kbp windows:

$$S_{\mathrm{TA}} = \frac{n_{\mathrm{T}} - n_{\mathrm{A}}}{n_{\mathrm{T}} + n_{\mathrm{A}}}, \quad S_{\mathrm{GC}} = \frac{n_{\mathrm{G}} - n_{\mathrm{C}}}{n_{\mathrm{G}} + n_{\mathrm{C}}}, \quad (18)$$

where n_{A}, n_{C}, n_{G} and n_{T} are respectively the numbers of A, C, G and T in the windows. Because of the observed correlation between the TA and GC skews, we also considered the total skew

$$S = S_{\mathrm{TA}} + S_{\mathrm{GC}}. \quad (19)$$

From the skews $S_{\mathrm{TA}}(n)$, $S_{\mathrm{GC}}(n)$ and $S(n)$, obtained along the sequences, where n is the position (in kbp units) from the origin, we also computed the cumulative skew profiles (or skew walk profiles):

$$\Sigma_{\mathrm{TA}}(n) = \sum_{j=1}^{n} S_{\mathrm{TA}}(j), \quad \Sigma_{\mathrm{GC}}(n) = \sum_{j=1}^{n} S_{\mathrm{GC}}(j), \quad (20)$$

and

$$\Sigma(n) = \sum_{j=1}^{n} S(j). \quad (21)$$

Our goal is to show that the skew DNA walks of the 22 human autosomes display an unexpected (with respect to previous monofractal diagnosis [27,28,29,30]) bifractal scaling behavior in the range 10 to 40 kbp as the signature of the presence of transcription-induced jumps in the LRC noisy S profiles. Sequences and gene annotation data ("ref-Gene") were retrieved from the UCSC Genome Browser (May 2004). We used RepeatMasker to exclude repetitive elements that might have been inserted recently and would not reflect long-term evolutionary patterns.

Revealing the Bifractality of Human Skew DNA Walks with the WTMM Method

As an illustration of our wavelet-based methodology, we show in Fig. 5 the S skew profile of a fragment of human chromosome 6 (Fig. 5a), the corresponding skew DNA

walk (Fig. 5b) and its space-scale wavelet decomposition using the Mexican hat analyzing wavelet $g^{(2)}$ (Fig. 1b). When computing $Z(q, a)$ (Eq. (7)) from the WT skeletons of the skew DNA walks Σ of the 22 human autosomes, we get convincing power-law behavior for $-1.5 \leq q \leq 3$ (data not shown). In Fig. 6a are reported the $\tau(q)$ exponents obtained using a linear regression fit of $\ln Z(q, a)$ vs. $\ln a$ over the range of scales $10\,\mathrm{kbp} \leq a \leq 40\,\mathrm{kbp}$. All the data points remarkably fall on two straight lines $\tau_1(q) = 0.78q - 1$ and $\tau_2(q) = q - 1$ which strongly suggests the presence of two types of singularities $h_1 = 0.78$ and $h_2 = 1$, respectively on two sets S_1 and S_2 with the same Haussdorf dimension $D = -\tau_1(0) = -\tau_2(0) = 1$, as
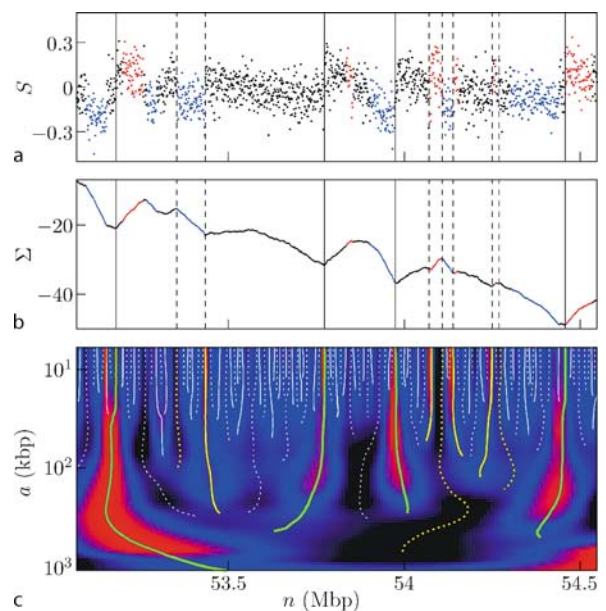
Fractals and Wavelets: What Can We Learn on Transcription and Replication ...?, Figure 5
a Skew profile $S(n)$ (Eq. (19)) of a repeat-masked fragment of human chromosome 6; *red* (resp. *blue*) 1 kbp window points correspond to (+) genes (resp. (−) genes) lying on the Watson (resp. Crick) strand; *black points* to intergenic regions. **b** Cumulated skew profile $\Sigma(n)$ (Eq. (21)). **c** WT of Σ; $T_{g^{(2)}}(n, a)$ is coded from *black* (min) to *red* (max); the WT skeleton defined by the maxima lines is shown in *solid* (resp. *dashed*) lines corresponding to *positive* (resp. *negative*) WT values. For illustration *yellow solid* (resp. *dashed*) *maxima lines* are shown to point to the positions of 2 *upward* (resp. 2 *downward*) *jumps* in S (*vertical dashed lines* in **a** and **b**) that coincide with gene transcription starts (resp. ends). In *green* are shown maxima lines that persist above $a \geq 200$ kbp and that point to sharp upward jumps in S (*vertical solid lines* in **a** and **b**) that are likely to be the locations of putative replication origins (see Sect. "From the Detection of Relication Origins Using the Wavelet Transform Microscope to the Modeling of Replication in Mammalian Genomes") [98,100]; note that 3 out of those 4 jumps are co-located with transcription start sites [129]
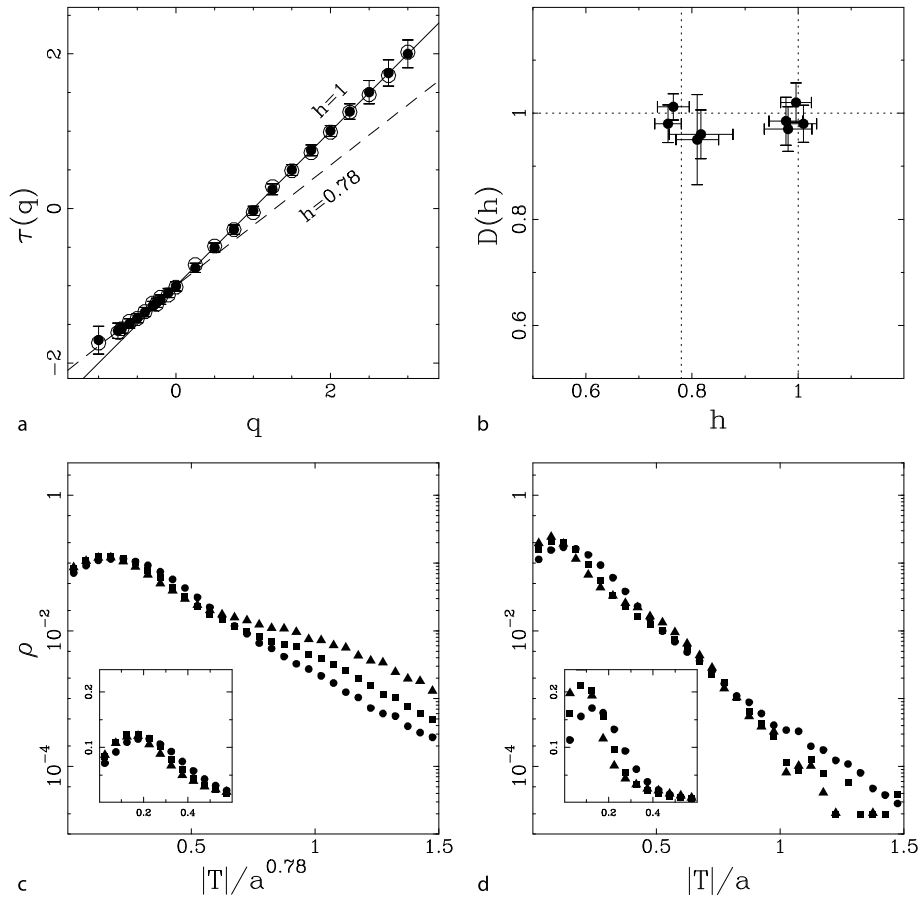
Fractals and Wavelets: What Can We Learn on Transcription and Replication …?, Figure 6
Multifractal analysis of $\Sigma(n)$ of the 22 human (*filled symbols*) and 19 mouse (*open circle*) autosomes using the WTMM method with $g^{(2)}$ over the range 10 kbp $\leq a \leq$ 40 kbp [129]. a $\tau(q)$ vs. q. b $D(h)$ vs. h. c WTMM pdf: ρ is plotted versus $|T|/a^H$ where $H = h_1 = 0.78$, in semi-log representation; the *inset* is an enlargement of the pdf central part in linear representation. d Same as in c but with $H = h_2 = 1$. In c and d, the symbols correspond to scales $a = 10$ (●), 20 (■) and 40 kbp (▲)

confirmed when computing the $D(h)$ singularity spectrum in Fig. 6b. This observation means that $Z(q, a)$ can be split in two parts [12,16]:

$$Z(q, a) = C_1(q)a^{qh_1-1} + C_2(q)a^{qh_2-1} , \qquad (22)$$

where $C_1(q)$ and $C_2(q)$ are prefactors that depend on q. Since $h_1 < h_2$, in the limit $a \mapsto 0^+$, the partition function is expected to behave like $Z(q, a) \sim C_1(q)a^{qh_1-1}$ for $q > 0$ and like $Z(q, a) \sim C_2(q)a^{qh_2-1}$ for $q < 0$, with a so-called phase transition [12,16] at the critical value $q_c = 0$. Surprisingly, it is the contribution of the weakest singularities $h_2 = 1$ that controls the scaling behavior of $Z(q, a)$ for $q > 0$ while the strongest ones $h_1 = 0.78$ actually dominate for $q < 0$ (Fig. 6a). This inverted behavior originates from finite (1 kbp) resolution which prevents the observation of the predicted scaling behavior in the limit $a \mapsto 0^+$.

The prefactors $C_1(q)$ and $C_2(q)$ in Eq. (22) are sensitive to (i) the number of maxima lines in the WT skeleton along which the WTMM behave as a^{h_1} or a^{h_2} and (ii) the relative amplitude of these WTMM. Over the range of scales used to estimate $\tau(q)$, the WTMM along the maxima lines pointing (at small scale) to $h_2 = 1$ singularities are significantly larger than those along the maxima lines associated to $h_1 = 0.78$ (see Figs. 6c, 6d). This implies that the larger $q > 0$, the stronger the inequality $C_2(q) \gg C_1(q)$ and the more pronounced the relative contribution of the second term in the r.h.s. of Eq. (22). On the opposite for $q < 0$, $C_1(q) \gg C_2(q)$ which explains that the strongest singularities $h_1 = 0.78$ now control the scaling behavior of $Z(q, a)$ over the explored range of scales.

In Figs. 6c, 6d are shown the WTMM pdfs computed at scales $a = 10$, 20 and 40 kbp after rescaling by a^{h_1} and

a^{h_2} respectively. We note that there does not exist a value of H such that all the pdfs collapse on a single curve as expected from Eq. (15) for monofractal DNA walks. Consistently with the $\tau(q)$ data in Fig. 6a and with the inverted scaling behavior discussed above, when using the two exponents $h_1 = 0.78$ and $h_2 = 1$, one succeeds in superimposing respectively the central (bump) part (Fig. 6c) and the tail (Fig. 6d) of the rescaled WTMM pdfs. This corroborates the bifractal nature of the skew DNA walks that display two competing scale-invariant components of Hölder exponents: (i) $h_1 = 0.78$ corresponds to LRC homogeneous fluctuations previously observed over the range $200\,\mathrm{bp} \lesssim a \lesssim 20\,\mathrm{kbp}$ in DNA walks generated with structural codings [29,30] and (ii) $h_2 = 1$ is associated to convex $\vee$ and concave $\wedge$ shapes in the DNA walks Σ indicating the presence of discontinuities in the derivative of Σ, i. e., of jumps in S (Figs. 5a, 5b). At a given scale a, according to Eq. (11), a large value of the WTMM in Fig. 5c corresponds to a strong derivative of the smoothed S profile and the maxima line to which it belongs is likely to point to a jump location in S. This is particularly the case for the colored maxima lines in Fig. 5c: Upward (resp. downward) jumps (Fig. 5a) are so-identified by the maxima lines corresponding to positive (resp. negative) values of the WT.

Transcription-Induced Step-like Skew Profiles in the Human Genome

In order to identify the origin of the jumps observed in the skew profiles, we have performed a systematic investigation of the skews observed along 14 854 intron containing genes [96,97]. In Fig. 7 are reported the mean values of S_{TA} and S_{GC} skews for all genes as a function of the distance to the 5′- or 3′- end. At the 5′ gene extremities (Fig. 7a), a sharp transition of both skews is observed from about zero values in the intergenic regions to finite positive values in transcribed regions ranging between 4 and 6% for $\bar{S}_{\mathrm{TA}}$ and between 3 and 5% for $\bar{S}_{\mathrm{GC}}$. At the gene 3′- extremities (Fig. 7b), the TA and GC skews also exhibit transitions from significantly large values in transcribed regions to very small values in untranscribed regions. However, in comparison to the steep transitions observed at 5′- ends, the 3′- end profiles present a slightly smoother transition pattern extending over $\sim 5\,\mathrm{kbp}$ and including regions downstream of the 3′- end likely reflecting the fact that transcription continues to some extent downstream of the polyadenylation site. In pluricellular organisms, mutations responsible for the observed biases are expected to have mostly occurred in germ-line cells. It could happen that gene 3′- ends annotated in the databank differ from the poly-A sites effectively used in the germ-line cells. Such differences would then lead to some broadening of the skew profiles.

From Skew Multifractal Analysis to Gene Detection

In Fig. 8 are reported the results of a statistical analysis of the jump amplitudes in human S profiles [129]. For maxima lines that extend above $a^* = 10\,\mathrm{kbp}$ in the WT skeleton (see Fig. 5c), the histograms obtained for upward and downward variations are quite similar, especially
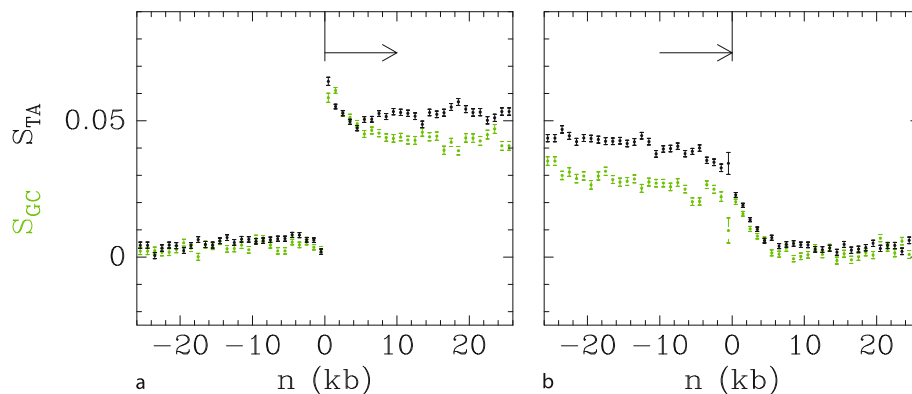
Fractals and Wavelets: What Can We Learn on Transcription and Replication …?, Figure 7
TA (●) and GC (*green* ●) skew profiles in the regions surrounding 5′ and 3′ gene extremities [96]. S_{TA} and S_{GC} were calculated in 1 kbp windows starting from each gene extremities in both directions. In abscissa is reported the distance (*n*) of each 1 kbp window to the indicated gene extremity; zero values of abscissa correspond to 5′- (**a**) or 3′- (**b**) gene extremities. In ordinate is reported the mean value of the skews over our set of 14 854 intron-containing genes for all 1 kbp windows at the corresponding abscissa. *Error bars represent the standard error of the means*
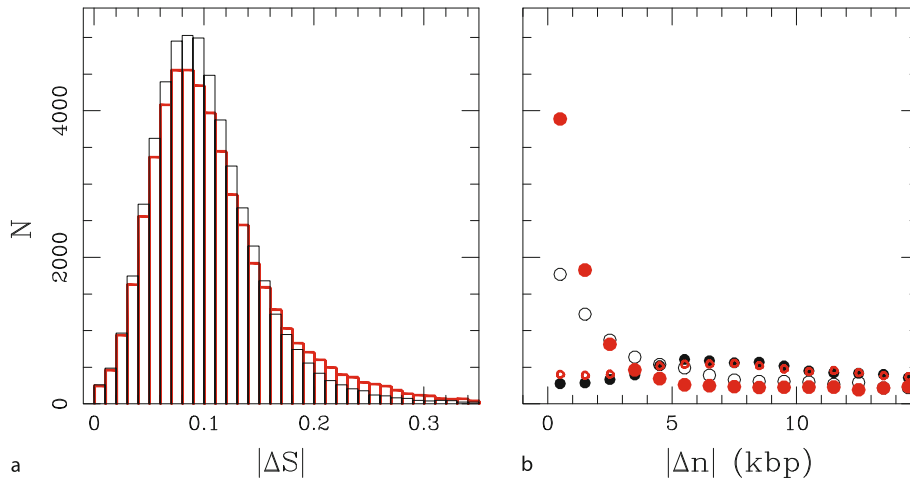
Fractals and Wavelets: What Can We Learn on Transcription and Replication …?, Figure 8
Statistical analysis of skew variations at the singularity positions determined at scale 1 kbp from the maxima lines that exist at scales
$a \geq 10$ kbp **in the WT skeletons of the 22 human autosomes** [129]. **For each singularity, we computed the variation amplitudes**
$\Delta S = \bar{S}(3') - \bar{S}(5')$ **over two adjacent 5 kbp windows, respectively in the 3' and 5' directions and the distances** Δn **to the closest**
TSS (resp. TES). a Histograms $N(|\Delta S|)$ **for upward** ($\Delta S > 0$, **red**) **and downward** ($\Delta S < 0$, **black**) **skew variations. b Histograms of the**
distances Δn **of upward** (**red**) **or downward** (**black**) **jumps with** $|\Delta S| \geq 0.1$ **to the closest TSS** (●, **red** ●) **and TES** (○, **red** ○)

their tails that are likely to correspond to jumps in the *S* profiles (Fig. 8a). When computing the distance between upward or downward jumps ($|\Delta S| \geq 0.1$) to the closest transcription start (TSS) or end (TES) sites (Fig. 8b), we reveal that the number of upward jumps in close proximity ($|\Delta n| \lesssim 3$ kpb) to TSS over-exceeds the number of such jumps close to TES. Similarly, downward jumps are preferentially located at TES. These observations are consistent with the step-like shape of skew profiles induced by transcription: $S > 0$ (resp. $S < 0$) is constant along a (+) (resp. (−)) genes and $S = 0$ in the intergenic regions (Fig. 7) [96]. Since a step-like pattern is edged by one upward and one downward jump, the set of human genes that are significantly biased is expected to contribute to an even number of $\Delta S > 0$ and $\Delta S < 0$ jumps when exploring the range of scales $10 \lesssim a \lesssim 40$ kbp, typical of human gene size. Note that in Fig. 8a, the number of sharp upward jumps actually slightly exceeds the number of sharp downward jumps, consistently with the experimental observation that whereas TSS are well defined, TES may extend over 5 kbp resulting in smoother downward skew transitions (Fig. 7b). This TES particularity also explains the excess of upward jumps found close to TSS as compared to the number of downward jumps close to TES (Fig. 8b).

In Fig. 9a, we report the analysis of the distance of TSS to the closest upward jump [129]. For a given upward jump amplitude, the number of TSS with a jump

within $|\Delta n|$ increases faster than expected (as compared to the number found for randomized jump positions) up to $|\Delta n| \simeq 2$ kbp. This indicates that the probability to find an upward jump within a gene promoter region is significantly larger than elsewhere. For example, out of 20 023 TSS, 36% (7228) are delineated within 2 kbp by a jump with $\Delta S > 0.1$. This provides a very reasonable estimate for the number of genes expressed in germline cells as compared to the 31.9% recently experimentally found to be bound to Pol II in human embryonic stem cells [130].

Combining the previous results presented in Figs. 8b and 9a, we report in Fig. 9b an estimate of the efficiency/coverage relationship by plotting the proportion of upward jumps ($\Delta S > \Delta S^*$) lying in TSS proximity as a function of the number of so-delineated TSS [129]. For a given proximity threshold $|\Delta n|$, increasing ΔS^* results in a decrease of the number of delineated TSS, characteristic of the right tail of the gene bias pdf. Concomitant to this decrease, we observe an increase of the efficiency up to a maximal value corresponding to some optimal value for ΔS^*. For $|\Delta n| < 2$ kbp, we reach a maximal efficiency of 60% for $\Delta S^* = 0.225$; 1403 out of 2342 upward jumps delineate a TSS. Given the fact that the actual number of human genes is estimated to be significantly larger ($\sim 30\,000$) than the number provided by refGene, a large part of the the 40% (939) of upward jumps that have not been associated to a refGene could be explained by this
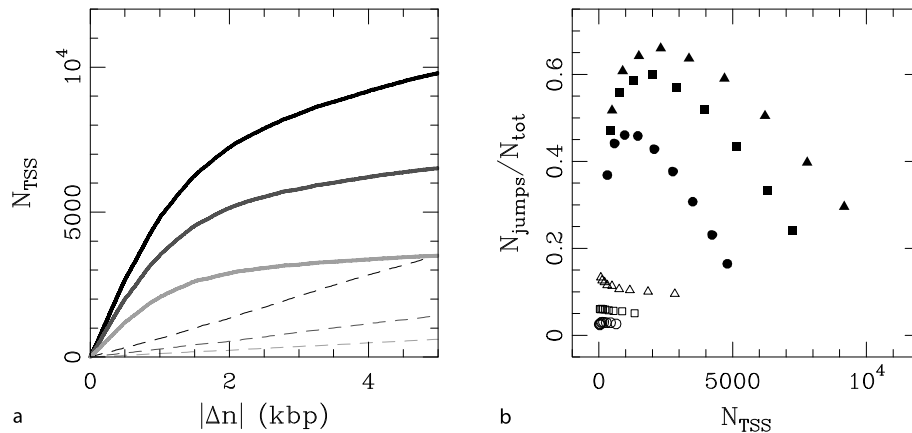
Fractals and Wavelets: What Can We Learn on Transcription and Replication ...?, Figure 9
a Number of TSS with an upward jump within $|\Delta n|$ (*abscissa*) for jump amplitudes $\Delta S > 0.1$ (*black*), 0.15 (*dark gray*) and 0.2 (*light gray*). *Solid lines* correspond to true jump positions while *dashed lines* to the same analysis when jump positions were randomly drawn along each chromosome [129]. **b** Among the $N_{tot}(\Delta S^*)$ upward jumps of amplitude larger than some threshold ΔS^*, we plot the proportion of those that are found within 1 kbp (●), 2 kbp (■) or 4 kbp (▲) of the closest TSS vs. the number N_{TSS} of the so-delineated TSS. Curves were obtained by varying ΔS^* from 0.1 to 0.3 (from *right* to *left*). *Open symbols* correspond to similar analyzes performed on random upward jump and TSS positions

limited coverage. In other words, jumps with sufficiently high amplitude are very good candidates for the location of highly-biased gene promoters. Let us point that out of the above 1403 (resp. 2342) upward jumps, 496 (resp. 624) jumps are still observed at scale $a^* = 200$ kbp. We will see in the next section that these jumps are likely to also correspond to replication origins underlying the fact that large upward jumps actually result from the cooperative contributions of both transcription- and replication- associated biases [98,99,100,101]. The observation that 80% (496/624) of the predicted replication origins are co-located with TSS enlightens the existence of a remarkable gene organization at replication origins [101].

To summarize, we have demonstrated the bifractal character of skew DNA walks in the human genome. When using the WT microscope to explore (repeat-masked) scales ranging from 10 to 40 kbp, we have identified two competing homogeneous scale-invariant components characterized by Hölder exponents $h_1 = 0.78$ and $h_2 = 1$ that respectively correspond to LRC colored noise and sharp jumps in the original DNA compositional asymmetry profiles. Remarkably, the so-identified upward (resp. downward) jumps are mainly found at the TSS (resp. TES) of human genes with high transcription bias and thus very likely highly expressed. As illustrated in Fig. 6a, similar bifractal properties are also observed when investigating the 19 mouse autosomes. This suggests that the results reported in this section are general features of mammalian genomes [129].

From the Detection of Relication Origins Using the Wavelet Transform Microscope to the Modeling of Replication in Mammalian Genomes

DNA replication is an essential genomic function responsible for the accurate transmission of genetic information through successive cell generations. According to the so-called "replicon" paradigm derived from prokaryotes [131], this process starts with the binding of some "initiator" protein to a specific "replicator" DNA sequence called *origin of replication*. The recruitment of additional factors initiate the bi-directional progression of two divergent replication forks along the chromosome. One strand is replicated continuously (leading strand), while the other strand is replicated in discrete steps towards the origin (lagging strand). In eukaryotic cells, this event is initiated at a number of replication origins and propagates until two converging forks collide at a *terminus of replication* [132]. The initiation of different replication origins is coupled to the cell cycle but there is a definite flexibility in the usage of the replication origins at different developmental stages [133,134,135,136, 137]. Also, it can be strongly influenced by the distance and timing of activation of neighboring replication origins, by the transcriptional activity and by the local chromatin structure [133,134,135,137]. Actually, sequence requirements for a replication origin vary significantly between different eukaryotic organisms. In the unicellular eukaryote *Saccharomyces cerevisiae*, the replication ori-

gins spread over 100–150 bp and present some highly conserved motifs [132]. However, among eukaryotes, *S. cerevisiae* seems to be the exception that remains faithful to the replicon model. In the fission yeast *Schizosaccharomyces pombe*, there is no clear consensus sequence and the replication origins spread over at least 800 to 1000 bp [132]. In multicellular organisms, the nature of initiation sites of DNA replication is even more complex. Metazoan replication origins are rather poorly defined and initiation may occur at multiple sites distributed over a thousand of base pairs [138]. The initiation of replication at random and closely spaced sites was repeatedly observed in *Drosophila* and *Xenopus* early embryo cells, presumably to allow for extremely rapid *S* phase, suggesting that any DNA sequence can function as a replicator [136,139,140]. A developmental change occurs around midblastula transition that coincides with some remodeling of the chromatin structure, transcription ability and selection of preferential initiation sites [136,140]. Thus, although it is clear that some sites consistently act as replication origins in most eukaryotic cells, the mechanisms that select these sites and the sequences that determine their location remain elusive in many cell types [141,142]. As recently proposed by many authors [143,144,145], the need to fulfill specific requirements that result from cell diversification may have led multicellular eukaryotes to develop various epigenetic controls over the replication origin selection rather than to conserve specific replication sequence. This might explain that only very few replication origins have been identified so far in multicellular eukaryotes, namely around 20 in metazoa and only about 10 in human [146]. Along the line of this epigenetic interpretation, one might wonder what can be learned about eukaryotic DNA replication from DNA sequence analysis.

Replication Induced Factory-Roof Skew Profiles in Mammalian Genomes

The existence of replication associated strand asymmetries has been mainly established in bacterial genomes [87,90, 92,93,94]. S_{GC} and S_{TA} skews abruptly switch sign (over few kbp) from negative to positive values at the replication origin and in the opposite direction from positive to negative values at the replication terminus. This step-like profile is characteristic of the replicon model [131] (see Fig. 13, left panel). In eukaryotes, the existence of compositional biases is unclear and most attempts to detect the replication origins from strand compositional asymmetry have been inconclusive. Several studies have failed to show compositional biases related to replication, and analysis of nucleotide substitutions in the region of the *β-globin* repli-

cation origin in primates does not support the existence of mutational bias between the leading and the lagging strands [92,147,148]. Other studies have led to rather opposite results. For instance, strand asymmetries associated with replication have been observed in the subtelomeric regions of *Saccharomyces cerevisiae* chromosomes, supporting the existence of replication-coupled asymmetric mutational pressure in this organism [149].

As shown in Fig. 10a for the TOP1 replication origin [146], most of the known replication origins in the human genome correspond to rather sharp (over several kbp) transitions from negative to positive *S* (S_{TA} as well as S_{GC}) skew values that clearly emerge from the noisy background. But when examining the behavior of the skews at larger distances from the origin, one does not observe a step-like pattern with upward and downward jumps at the origin and termination positions, respectively, as expected for the bacterial replicon model (Fig. 13, left panel). Surprisingly, on both sides of the upward jump, the noisy *S* profile decreases steadily in the 5′ to 3′ direction without clear evidence of pronounced downward jumps. As shown in Figs. 10b–10d, sharp upward jumps of amplitude $\Delta S \gtrsim 15\%$, similar to the ones observed for the known replication origins (Fig. 10a), seem to exist also at many other locations along the human chromosomes. But the most striking feature is the fact that in between two neighboring major upward jumps, not only the noisy *S* profile does not present any comparable downward sharp transition, but it displays a remarkable decreasing linear behavior. At chromosome scale, we thus get jagged *S* profiles that have the aspect of "factory roofs" [98,100,146]. Note that the jagged *S* profiles shown in Figs. 10a–10d look somehow disordered because of the extreme variability in the distance between two successive upward jumps, from spacing ~ 50–100 kbp (~ 100–200 kbp for the native sequences) mainly in GC rich regions (Fig. 10d), up to 1–2 Mbp (~ 2–3 Mbp for native sequences) (Fig. 10c) in agreement with recent experimental studies [150] that have shown that mammalian replicons are heterogeneous in size with an average size ~ 500 kbp, the largest ones being as large as a few Mbp. But what is important to notice is that some of these segments between two successive skew upward jumps are entirely intergenic (Figs. 10a, 10c), clearly illustrating the particular profile of a strand bias resulting solely from replication [98,100, 146]. In most other cases, we observe the superimposition of this replication profile and of the step-like profiles of (+) and (−) genes (Fig. 7), appearing as upward and downward blocks standing out from the replication pattern (Fig. 10c). Importantly, as illustrated in Figs. 10e, 10f, the factory-roof pattern is not specific to human se-
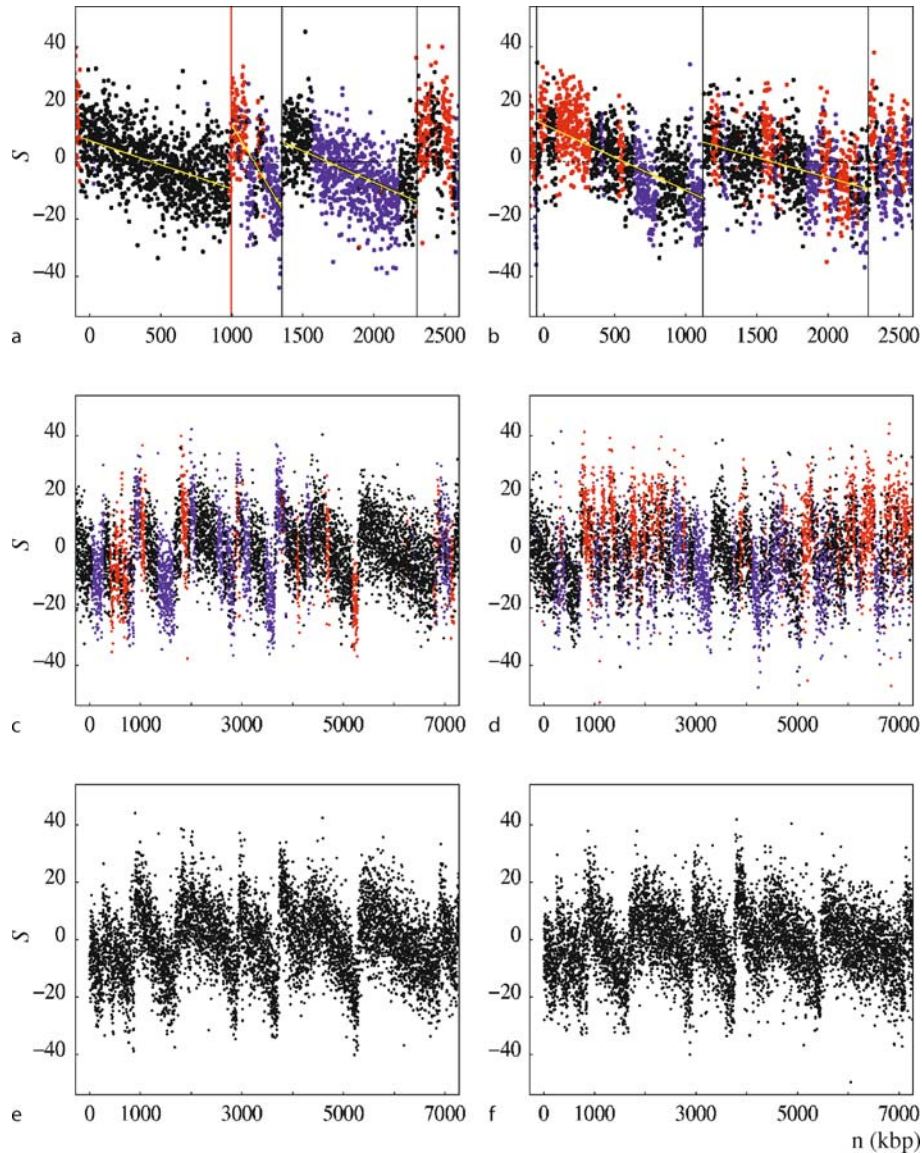
Fractals and Wavelets: What Can We Learn on Transcription and Replication …?, Figure 10
S profiles along mammalian genome fragments [100,146]. **a** Fragment of human chromosome 20 including the TOP1 origin (*red vertical line*). **b** and **c** Human chromosome 4 and chromosome 9 fragments, respectively, with low GC content (36%). **d** Human chromosome 22 fragment with larger GC content (48%). In **a** and **b**, *vertical lines* correspond to selected putative origins (see Subsect. "Detecting Replication Origins from the Skew WT Skeleton"); *yellow lines* are linear fits of the *S* values between successive putative origins. *Black intergenic regions*; *red*, (+) genes; *blue*, (−) genes. Note the fully intergenic regions upstream of TOP1 in **a** and from positions 5290–6850 kbp in **c**. **e** Fragment of mouse chromosome 4 homologous to the human fragment shown in **c**. **f** Fragment of dog chromosome 5 syntenic to the human fragment shown in **c**. In **e** and **f**, genes are not represented

quences but is also observed in numerous regions of the mouse and dog genomes [100]. Hence, the presence of strand asymmetry in regions that have strongly diverged during evolution further supports the existence of compostional bias associated with replication in mammalian germ-line cells [98,100,146].

Detecting Replication Origins from the Skew WT Skeleton

We have shown in Fig. 10a that experimentally determined human replication origins coincide with large-amplitude upward transitions in noisy skew profiles. The

corresponding ΔS ranges between 14% and 38%, owing to possible different replication initiation efficiencies and/or different contributions of transcriptional biases (Sect. "Bifractality of Human DNA Strand-Asymmetry Profiles Results from Transcription"). Along the line of the jump detection methodology described in Sect. "Bifractality of Human DNA Strand-Asymmetry Profiles Results from Transcription", we have checked that upward jumps observed in the skew S at these known replication origins correspond to maxima lines in the WT skeleton that extend to rather large scales $a > a^* = 200$ kbp. This observation has led us to select the maxima lines that exist above $a^* = 200$ kbp, i. e. a scale which is smaller than the typical replicon size and larger than the typical gene size [98, 100]. In this way, we not only reduce the effect of the noise but we also reduce the contribution of the upward (5′ extremity) and backward (3′ extremity) jumps associated to the step-like skew pattern induced by transcription only (Sect. "Bifractality of Human DNA Strand-Asymmetry Profiles Results from Transcription"), to the benefit of maintaining a good sensitivity to replication induced jumps. The detected jump locations are estimated as the positions at scale 20 kbp of the so-selected maxima lines. According to Eq. (11), upward (resp. downward) jumps are identified by the maxima lines corresponding to positive (resp. negative) values of the WT as illustrated in Fig. 5c by the green solid (resp. dashed) maxima lines. When applying this methodology to the total skew S along the repeat-masked DNA sequences of the 22 human autosomal chromosomes, 2415 upward jumps

are detected and, as expected, a similar number (namely 2686) of downward jumps. In Fig. 11a are reported the histograms of the amplitude $|\Delta S|$ of the so-identified upward ($\Delta S > 0$) and downward ($\Delta S < 0$) jumps respectively. These histograms no longer superimpose as previously observed at smaller scales in Fig. 8a, the former being significantly shifted to larger $|\Delta S|$ values. When plotting $N(|\Delta S| > \Delta S^*)$ versus ΔS^* in Fig. 11b, we can see that the number of large amplitude upward jumps overexceeds the number of large amplitude downward jumps. These results confirm that most of the sharp upward transitions in the S profiles in Fig. 10 have no sharp downward transition counterpart [98,100]. This excess likely results from the fact that, contrasting with the prokaryote replicon model (Fig. 13, left panel) where downward jumps result from precisely positioned replication terminations, in mammals termination appears not to occur at specific positions but to be randomly distributed. Accordingly the small number of downward jumps with large $|\Delta S|$ is likely to result from transcription (Fig. 5) and not from replication. These jumps are probably due to highly biased genes that also generate a small number of large-amplitude upward jumps, giving rise to false-positive candidate replication origins. In that respect, the number of large downward jumps can be taken as an estimation of the number of false positives. In a first step, we have retained as acceptable a proportion of 33% of false positives. As shown in Fig. 11b, this value results from the selection of upward and downward jumps of amplitude $|\Delta S| \geq 12.5\%$, corresponding to a ratio of upward over downward jumps
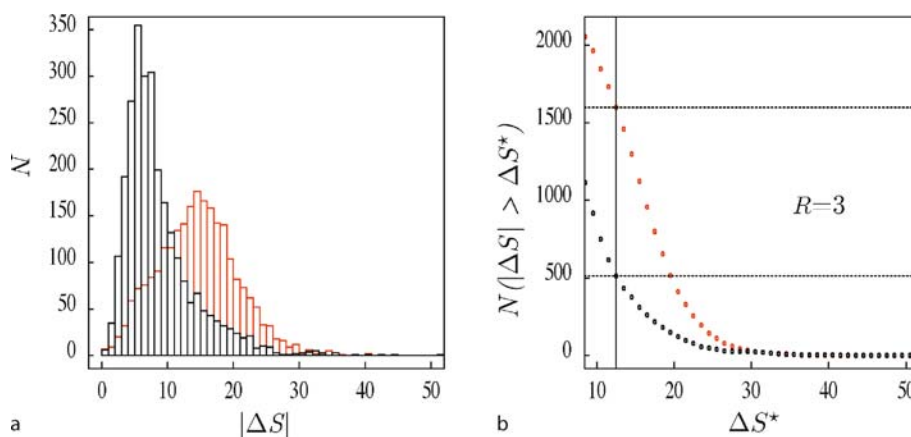
Statistical analysis of the sharp jumps detected in the S profiles of the 22 human autosomal chromosomes by the WT microscope at scale $a^* = 200$ kbp for repeat-masked sequences [98,100]. $|\Delta S| = |\bar{S}(3') - \bar{S}(5')|$, where the averages were computed over the two adjacent 20 kbp windows, respectively, in the 3′ and 5′ direction from the detected jump location. **a** Histograms $N(|\Delta S|)$ of $|\Delta S|$ values. **b** $N(|\Delta S| > \Delta S^*)$ vs. ΔS^*. In **a** and **b**, the *black* (resp. *red*) *line* corresponds to downward $\Delta S < 0$ (resp. upward $\Delta S > 0$) jumps. $R = 3$ corresponds to the ratio of upward over downward jumps presenting an amplitude $|\Delta S| \geq 12.5\%$ (see text)
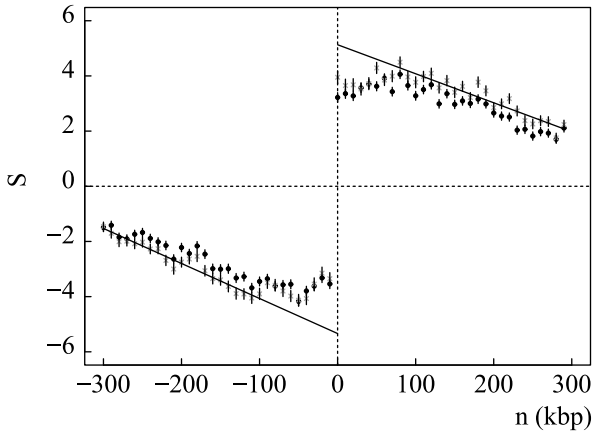
Fractals and Wavelets: What Can We Learn on Transcription and Replication …?, Figure 12

Mean skew profile of intergenic regions around putative replication origins [100]. The skew S was calculated in 1 kbp windows (Watson strand) around the position (±300 kbp without repeats) of the 1012 detected upward jumps; 5′ and 3′ transcript extremities were extended by 0.5 and 2 kbp, respectively (●), or by 10 kbp at both ends (✱). The abscissa represents the distance (in kbp) to the corresponding origin; the ordinate represents the skews calculated for the windows situated in intergenic regions (mean values for all discontinuities and for 10 consecutive 1 kbp window positions). The skews are given in percent (*vertical bars*, SEM). The *lines* correspond to linear fits of the values of the skew (✱) for $n < -100$ kbp and $n > 100$ kbp

$R = 3$. Let us notice that the value of this ratio is highly variable along the chromosome [146] and significantly larger than 1 for $G+C \lesssim 42\%$.

In a final step, we have decided [98,100,146] to retain as putative replication origins upward jumps with $|\Delta S| \geq 12.5\%$ detected in regions with $G+C \leq 42\%$. This selection leads to a set of 1012 candidates among which our estimate of the proportion of true replication origins is 79% ($R = 4.76$). In Fig. 12 is shown the mean skew profile calculated in intergenic windows on both sides of the 1012 putative replication origins [100]. This mean skew profile presents a rather sharp transition from negative to positive values when crossing the origin position. To avoid any bias in the skew values that could result from incompletely annotated gene extremities (e. g. 5′ and 3′ UTRs), we have removed 10-kbp sequences at both ends of all annotated transcripts. As shown in Fig. 12, the removal of these intergenic sequences does not significantly modifies the mean skew profile, indicating that the observed values do not result from transcription. On both sides of the jump, we observe a linear decrease of the bias with some flattening of the profile close to the transition point. Note that, due to (i) the potential presence of signals implicated in replication initiation and (ii) the possible existence of dispersed origins [151], one might question the meaningfulness of this flattening that leads to a significant underestimate of the jump amplitude. Furthermore, according to our detection methodology, the numerical uncertainty on
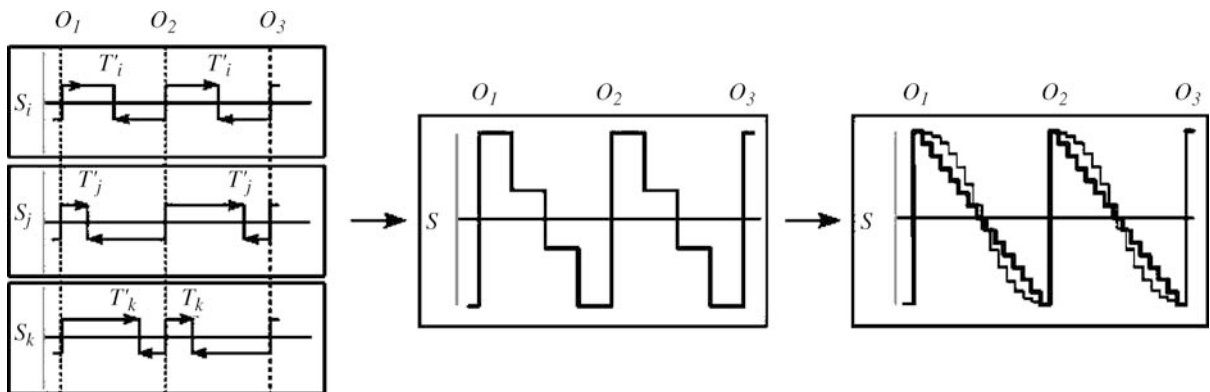
Fractals and Wavelets: What Can We Learn on Transcription and Replication …?, Figure 13

Model of replication termination [98,100]. Schematic representation of the skew profiles associated with three replication origins O_1, O_2, and O_3; we suppose that these replication origins are adjacent, bidirectional origins with similar replication efficiency. The abscissa represents the sequence position; the ordinate represents the S value (arbitrary units). *Upward* (or *downward*) steps correspond to origin (or termination) positions. For convenience, the termination sites are symmetric relative to O_2. (*Left*) Three different termination positions T_i, T_j, and T_k, leading to elementary skew profiles S_i, S_j, and S_k as predicted by the replicon model [146]. (*Center*) Superposition of these three profiles. (*Right*) Superposition of a large number of elementary profiles leading to the final factory-roof pattern. In the simple model, termination occurs with equal probability on both sides of the origins, leading to the linear profile (*thick line*). In the alternative model, replication termination is more likely to occur at lower rates close to the origins, leading to a flattening of the profile (*gray line*)

the putative origin position estimate may also contribute to this flattening. As illustrated in Fig. 12, when extrapolating the linear behavior observed at distances > 100 kbp from the jump, one gets a skew of 5.3%, i. e. a value consistent with the skew measured in intergenic regions around the six experimentally known replication origins namely $7.0 \pm 0.5\%$. Overall, the detection of sharp upward jumps in the skew profiles with characteristics similar to those of experimentally determined replication origins and with no downward counterpart further supports the existence, in human chromosomes, of replication-associated strand asymmetries, leading to the identification of numerous putative replication origins active in germ-line cells.

A Model of Replication in Mammalian Genomes

Following the observation of jagged skew profiles similar to factory roofs in Subsect. "Replication Induced Factory-Roof Skew Profiles in Mammalian Genomes", and the quantitative confirmation of the existence of such (piecewise linear) profiles in the neighborhood of 1012 putative origins in Fig. 12, we have proposed, in Touchon et al. [100] and Brodie of Brodie et al. [98], a rather crude model for replication in the human genome that relies on the hypothesis that the replication origins are quite well positioned while the terminations are randomly distributed. Although some replication terminations have been found at specific sites in *S. cerevisiae* and to some extent in *Schizosaccharomyces pombe* [152], they occur randomly between active origins in *Xenopus* egg extracts [153, 154]. Our results indicate that this property can be extended to replication in human germ-line cells. As illustrated in Fig. 13, replication termination is likely to rely on the existence of numerous potential termination sites distributed along the sequence. For each termination site (used in a small proportion of cell cycles), strand asymmetries associated with replication will generate a step-like skew profile with a downward jump at the position of termination and upward jumps at the positions of the adjacent origins (as in bacteria). Various termination positions will thus correspond to classical replicon-like skew profiles (Fig. 13, left panel). Addition of these profiles will generate the intermediate profile (Fig. 13, central panel). In a simple picture, we can reasonably suppose that termination occurs with constant probability at any position on the sequence. This behavior can, for example, result from the binding of some termination factor at any position between successive origins, leading to a homogeneous distribution of termination sites during successive cell cycles. The final skew profile is then a linear segment decreasing between successive origins (Fig. 13, right panel).

Let us point out that firing of replication origins during time interval of the *S* phase [155] might result in some flattening of the skew profile at the origins as sketched in Fig. 13 (right panel, gray curve). In the present state, our results [98,100,146] support the hypothesis of random replication termination in human, and more generally in mammalian cells (Fig. 10), but further analyzes will be necessary to determine what scenario is precisely at work.

A Wavelet-Based Methodology to Disentangle Transcription- and Replication-Associated Strand Asymmetries Reveals a Remarkable Gene Organization in the Human Genome

During the duplication of eukaryotic genomes that occurs during the *S* phase of the cell cycle, the different replication origins are not all activated simultaneously [132,135, 138,150,155,156]. Recent technical developments in genomic clone microarrays have led to a novel way of detecting the temporal order of DNA replication [155,156]. The arrays are used to estimate *replication timing ratios* i. e. ratios between the average amount of DNA in the *S* phase at a locus along the genome and the usual amount of DNA present in the *G*1 phase for that locus. These ratios should vary between 2 (throughout the *S* phase, the amount of DNA for the earliest replicating regions is twice the amount during *G*1 phase) and 1 (the latest replicating regions are not duplicated until the end of *S* phase). This approach has been successfully used to generate genome-wide maps of replication timing for *S. cerevisiae* [157], *Drosophila melanogaster* [137] and human [158]. Very recently, two new analyzes of human chromosome 6 [156] and 22 [155] have improved replication timing resolution from 1 Mbp down to ∼ 100 kbp using arrays of overlapping tile path clones. In this section, we report on a very promising first step towards the experimental confirmation of the thousand putative replication origins described in Sect. "From the Detection of Relication Origins Using the Wavelet Transform Microscope to the Modeling of Replication in Mammalian Genomes". The strategy will consist in mapping them on the recent high-resolution timing data [156] and in checking that these regions replicate earlier than their surrounding [114]. But to provide a convincing experimental test, we need as a prerequisite to extract the contribution of the compositional skew specific to replication.

Disentangling Transcription- and Replication-Associated Strand Asymmetries

The first step to detect putative replication domains consists in developing a multi-scale pattern recognition
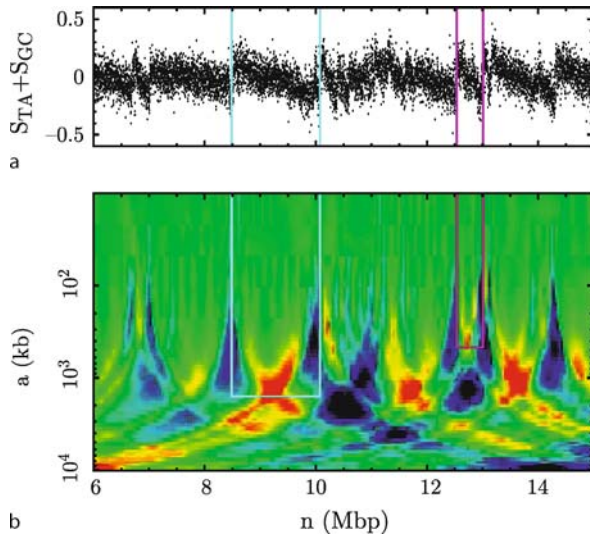
Fractals and Wavelets: What Can We Learn on Transcription and Replication ...?, Figure 14

Wavelet-based analysis of genomic sequences. a Skew profile *S* of a 9 Mbp repeat-masked fragment of human chromosome 21. b WT of *S* using φ_R (Fig. 1c); $T_{\phi_R}[S](n,a)$ is color-coded from *dark-blue* (min; negative values) to *red* (max; positive values) through *green* (null values). *Light-blue* and *purple lines* illustrate the detection of two replication domains of significantly different sizes. Note that in b, *blue cone-shape areas* signing upward jumps point at small scale (*top*) towards the putative replication origins and that the vertical positions of the WT maxima (*red areas*) corresponding to the two indicated replication domains match the distance between the putative replication origins (1.6 Mbp and 470 kbp respectively)

methodology based on the WT of the strand compositional asymmetry *S* using as analyzing wavelet $\phi_R(x)$ (Eq. (12)) that is adapted to perform an objective segmentation of factory-roof skew profiles (Fig. 1c). As illustrated in Fig. 14, the space-scale location of significant maxima values in the 2D WT decomposition (red areas in Fig. 14b) indicates the middle position (spatial location) of candidate replication domains whose size is given by the scale location. In order to avoid false positives, we then check that there does exist a well-defined upward jump at each domain extremity. These jumps appear in Fig. 14b as blue cone-shape areas pointing at small scale to the jumps positions where are located the putative replication origins. Note that because the analyzing wavelet is of zero mean (Eq. (2)), the WT decomposition is insensitive to (global) asymmetry offset.

But as discussed in Sect. "Bifractality of Human DNA Strand-Asymmetry Profiles Results from Transcription", the overall observed skew *S* also contains some contribution induced by transcription that generates step-like

blocks corresponding to (+) and (−) genes [96,97,129]. Hence, when superimposing the replication serrated and transcription step-like skew profiles, we get the following theoretical skew profile in a replication domain [114]:

$$
\begin{aligned}
S(x') &= S_R(x') + S_T(x') \\
&= -2\delta \times \left(x' - \frac{1}{2} \right) + \sum_{\text{gene}} c_g \chi_g(x') ,
\end{aligned}
\tag{23}
$$

where position x' within the domain has been rescaled between 0 and 1, $\delta > 0$ is the replication bias, χ_g is the characteristic function for the gth gene (1 when x' points within the gene and 0 elsewhere) and c_g is its transcriptional bias calculated on the Watson strand (likely to be positive for (+) genes and negative for (−) genes). The objective is thus to detect human replication domains by delineating, in the noisy *S* profile obtained at 1 kbp resolution (Fig. 15a), all chromosomal loci where *S* is well fitted by the theoretical skew profile Eq. (23).

In order to enforce strong compatibility with the mammalian replicon model (Subsect. "A Model of Replication in Mammalian Genomes"), we will only retain the domains the most likely to be bordered by putative replication origins, namely those that are delimited by upward jumps corresponding to a transition from a negative *S* value $< -3\%$ to a positive *S* value $> +3\%$. Also, for each domain so-identified, we will use a least-square fitting procedure to estimate the replication bias δ, and each of the gene transcription bias c_g. The resulting χ^2 value will then be used to select the candidate domains where the noisy *S* profile is well described by Eq. (23). As illustrated in Fig. 15 for a fragment of human chromosome 6 that contains 4 adjacent replication domains (Fig. 15a), this method provides a very efficient way of disentangling the step-like transcription skew component (Fig. 15b) from the serrated component induced by replication (Fig. 15c). Applying this procedure to the 22 human autosomes, we delineated 678 replication domains of mean length $\langle L \rangle = 1.2 \pm 0.6$ Mbp, spanning 28.3% of the genome and predicted 1060 replication origins.

DNA Replication Timing Data Corroborate *in silico* Human Replication Origin Predictions

Chromosome 22 being rather atypical in gene and GC contents, we mainly report here on the correlation analysis [114] between nucleotide compositional skew and timing data for chromosome 6 which is more representative of the whole human genome. Note that timing data for clones completely included in another clone have been removed after checking for timing ratio value consistency leaving 1648 data points. The timing ratio value at each
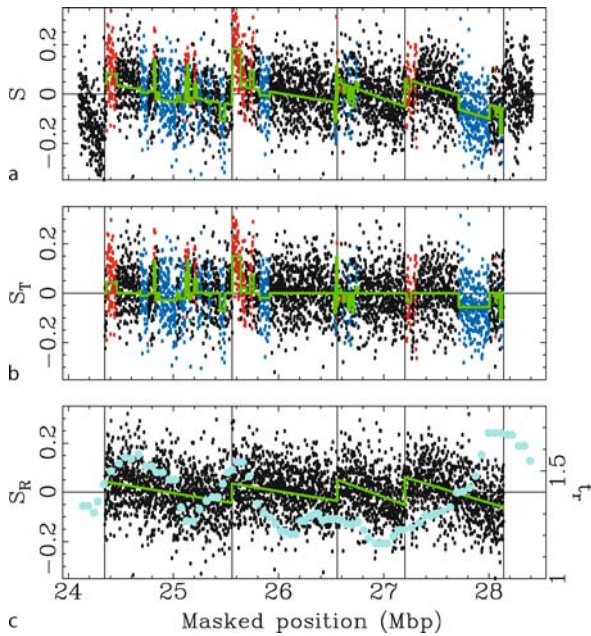
Fractals and Wavelets: What Can We Learn on Transcription and Replication …?, Figure 15

a Skew profile S of a 4.3 Mbp repeat-masked fragment of human chromosome 6 [114]; each point corresponds to a 1 kbp window: *Red*, (+) genes; *blue*, (−) genes; *black*, intergenic regions (the color was defined by majority rule); the estimated skew profile (Eq. (23)) is shown in *green*; *vertical lines* correspond to the locations of 5 putative replication origins that delimit 4 adjacent domains identified by the wavelet-based methodology. **b** Transcription-associated skew S_T obtained by subtracting the estimated replication-associated profile (*green lines* in **c**) from the original S profile in **a**; the estimated transcription step-like profile (second term on the *rhs* of Eq. (23)) is shown in *green*. **c** Replication-associated skew S_R obtained by subtracting the estimated transcription step-like profile (*green lines* in **b**) from the original S profile in **a**; the estimated replication serrated profile (first term in the *rhs* of Eq. (23)) is shown in *green*; the *light-blue dots* correspond to high-resolution t_r data

point has been chosen as the median over the 4 closest data points to remove noisy fluctuations resulting from clone heterogeneity (clone length 100 ± 51 kbp and distance between successive clone mid-points 104 ± 89 kbp), so that the spatial resolution is rather inhomogeneous ~ 300 kbp. Note that using asynchronous cells also results in some smoothing of the data, possibly masking local maxima.

Our wavelet-based methodology has identified 54 replication domains in human chromosome 6 [114]; these domains are bordered by 83 putative replication origins among which 25 are common to two adjacent domains. Four of these contiguous domains are shown in Fig. 15. In Fig. 15c, on top of the replication skew profile S_R, are

reported for comparison the high-resolution timing ratio t_r data from [156]. The histogram of t_r values obtained at the 83 putative origin locations displays a maximum at $t_r \simeq \langle t_r \rangle \simeq 1.5$ (data not shown) and confirms what is observed in Fig. 15c, namely that a majority of the predicted origins are rather early replicating with $t_r \gtrsim 1.4$. This contrasts with the rather low t_r ($\simeq 1.2$) values observed in domain central regions (Fig. 15c). But there is an even more striking feature in the replication timing profile in Fig. 15c: 4 among the 5 predicted origins correspond, relatively to the experimental resolution, to local maxima of the t_r profile. As shown in Fig. 16a, the average t_r profile around the 83 putative replication origins decreases regularly on both sides of the origins over a few (4–6) hundreds kbp confirming statistically that domain borders replicate earlier than their left and right surroundings which is consistent with these regions being true replication origins mostly active early in S phase. In fact, when averaging over the top 20 origins with a well-defined local maximum in the t_r profile, $\langle t_r \rangle$ displays a faster decrease on both sides of the origin and a higher maximum value ~ 1.55 corresponding to the earliest replicating origins. On the opposite, when averaging t_r profiles over the top 10 late replicating origins, we get, as expected, a rather flat mean profile ($t_r \sim 1.2$) (Fig. 16a). Interestingly, these origins are located in rather wide regions of very low GC content ($\lesssim 34\%$, not shown) correlating with chromosomal G banding patterns predominantly composed of GC-poor isochores [159,160]. This illustrates how the statistical contribution of rather flat profiles observed around late replicating origins may significantly affect the overall mean t_r profile. Individual inspection of the 38 replication domains with $L \geq 1$ Mbp shows that, in those domains that are bordered by early replicating origins ($t_r \gtrsim 1.4 - 1.5$), the replication timing ratio t_r and the absolute value of the replication skew $|S_R|$ turn out to be strongly correlated. This is quantified in Fig. 16b by the histogram of the Pearson's correlation coefficient values that is clearly shifted towards positive values with a maximum at ~ 0.4. Altogether the results of this comparative analysis provide the first experimental verification of *in silico* replication origins predictions: The detected putative replication domains are bordered by replication origins mostly active in the early S phase, whereas the central regions replicate more likely in late S phase.

Gene Organization
in the Detected Replication Domains

Most of the 1060 putative replication origins that border the detected replication domains are intergenic (77%)
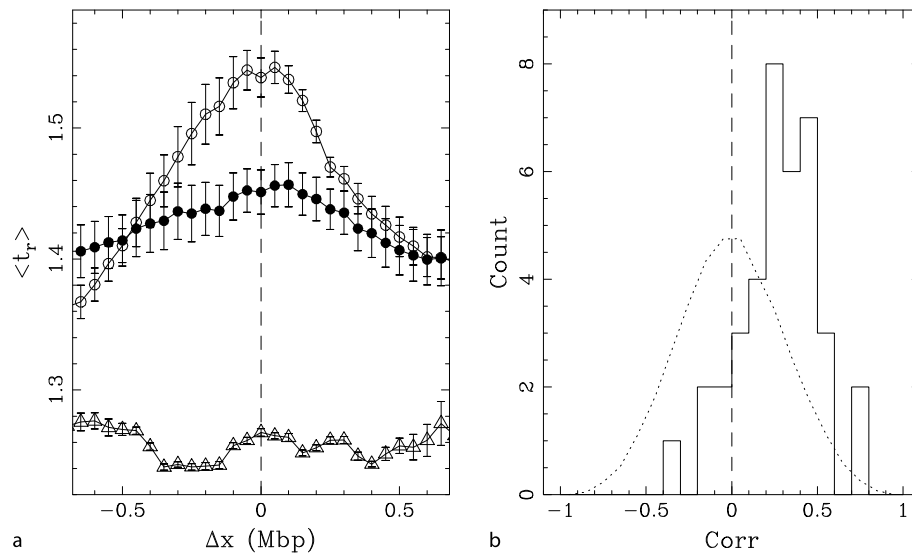
Fractals and Wavelets: What Can We Learn on Transcription and Replication …?, Figure 16
a Average replication timing ratio (±SEM) determined around the 83 putative replication origins (●), 20 origins with well-defined local maxima (○) and 10 late replicating origins (△). Δx is the native distance to the origins in Mbp units [114]. **b** Histogram of Pearson's correlation coefficient values between t_r and the absolute value of S_R over the 38 predicted domains of length $L \geq 1$ Mbp. The *dotted line* corresponds to the expected histogram computed with the correlation coefficients between t_r and $|S|$ profiles over independent windows randomly positioned along chromosome 6 and with the same length distribution as the 38 detected domains

and are located near to a gene promoter more often than would be expected by chance (data not shown) [101]. The replication domains contain approximately equal numbers of genes oriented in each direction (1511 (+) genes and 1507 (−) genes). Gene distributions in the 5′ halves of domains contain more (+) genes than (−) genes, regardless of the total number of genes located in the half-domains (Fig. 17b). Symmetrically, the 3′ halves contain more (−) genes than (+) genes (Fig. 17b). 32.7% of half-domains contain one gene, and 50.9% contain more than one gene. For convenience, (+) genes in the 5′ halves and (−) genes in the 3′ halves are defined as R+ genes, (Fig. 17a): their transcription is, in most cases, oriented in the same direction as the putative replication fork progression (genes transcribed in the opposite direction are defined as R− genes). The 678 replication domains contain significantly more R+ genes (2041) than R− genes (977). Within 50 kbp of putative replication origins, the mean density of R+ genes is 8.2 times greater than that of R− genes. This asymmetry weakens progressively with the distance from the putative origins, up to ∼ 250 kbp (Fig. 17b). A similar asymmetric pattern is observed when the domains containing duplicated genes are eliminated from the analysis, whereas control domains obtained after randomization of domain positions present similar R+ and R− gene density distributions (Supplementary in [101]).

The mean length of the R+ genes near the putative origins is significantly greater (∼ 160 kbp) than that of the R− genes (∼ 50 kbp), however both tend towards similar values (∼ 70 kbp) at the center of the domain (Fig. 17c). Within 50 kbp of the putative origins, the ratio between the numbers of base pairs transcribed in the R+ and R− directions is 23.7; this ratio falls to ∼ 1 at the domain centers (Fig. 17d). In Fig. 17e are reported the results of the analysis of the breadth of expression, N_t (number of tissues in which a gene is expressed) of genes located within the detected domains [101]. As measured by EST data (similar results are obtained by SAGE or microarray data [101]), N_t is found to decrease significantly from the extremities to the center in a symmetrical manner in the 5′ and 3′ half-domains (Fig. 17e). Thus, genes located near the putative replications origins tend to be widely expressed whereas those located far from them are mostly tissue-specific.

To summarize, the results reported in this section provide the first demonstration of quantitative relationships in the human genome between gene expression, orientation and distance from putative replication origins [101]. A possible key to the understanding of this complex architecture is the coordination between replication and transcription [101]. The putative replication origins would mostly be active early in the *S* phase in most tissues. Their activity could result from particular genomic context in-
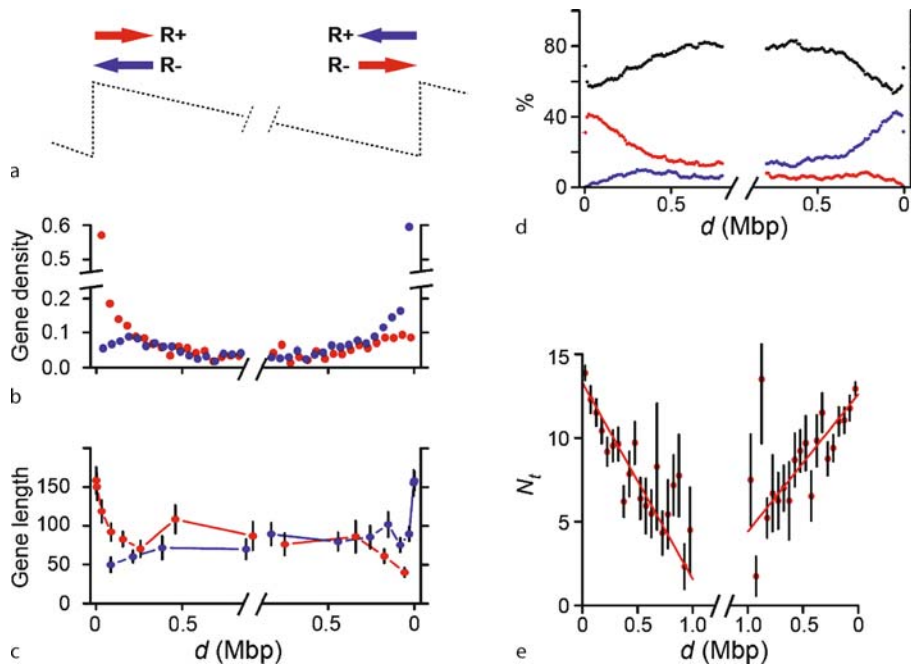
Fractals and Wavelets: What Can We Learn on Transcription and Replication …?, Figure 17
Analysis of the genes located in the identified replication domains [101]. a *Arrows* indicate the R+ orientation, i. e. the same orienta-
tion as the most frequent direction of putative replication fork progression; R— orientation (opposed direction); *red*, (+) genes; *blue*,
(—) genes. **b** Gene density. The density is defined as the number of 5′ ends (for (+) genes) or of 3′ ends (for (—) genes) in 50-kbp
adjacent windows, divided by the number of corresponding domains. In abscissa, the distance, *d*, in Mbp, to the closest domain
extremity. **c** Mean gene length. Genes are ranked by their distance, *d*, from the closest domain extremity, grouped by sets of 150
genes, and the mean length (kbp) is computed for each set. **d** Relative number of base pairs transcribed in the + direction (*red*), — di-
rection (*blue*) and non-transcribed (*black*) determined in 10-kbp adjacent sequence windows. **e** Mean expression breadth using EST
data [101]

volving transcription factor binding sites and/or from the
transcription of their neighboring housekeeping genes.
This activity could also be associated with an open chro-
matin structure, permissive to early replication and gene
expression in most tissues [161,162,163,164]. This open
conformation could extend along the first gene, possi-
bly promoting the expression of further genes. This ef-
fect would progressively weaken with the distance from
the putative replication origin, leading to the observed
decrease in expression breadth. This model is consistent
with a number of data showing that in metazoans, ORC
and RNA polymerase II colocalize at transcriptional pro-
moter regions [165], and that replication origins are deter-
mined by epigenetic information such as transcription fac-
tor binding sites and/or transcription [166,167,168,169]. It
is also consistent with studies in *Drosophila* and humans
that report correlation between early replication timing
and increased probability of expression [137,155,156,165,
170]. Furthermore, near the putative origins bordering
the replication domains, transcription is preferentially ori-
ented in the same direction as replication fork progres-

sion. This co-orientation is likely to reduce head-on colli-
sions between the replication and transcription machiner-
ies, which may induce deleterious recombination events
either directly or via stalling of the replication fork [171,
172]. In bacteria, co-orientation of transcription and repli-
cation has been observed for essential genes, and has been
associated with a reduction in head-on collisions between
DNA and RNA polymerases [173]. It is noteworthy that
in human replication domains such co-orientation usu-
ally occurs in widely-expressed genes located near puta-
tive replication origins. Near domain centers, head-on col-
lisions may occur in 50% of replication cycles, regardless
of the transcription orientation, since there is no prefer-
ential orientation of the replication fork progression in
these regions. However, in most cell types, there should be
few head-on collisions due to the low density and expres-
sion breadth of the corresponding genes. Selective pres-
sure to reduce head-on collisions may thus have con-
tributed to the simultaneous and coordinated organization
of gene orientation and expression breadth along the de-
tected replication domains [101].

Future Directions

From a statistical multifractal analysis of nucleotide strand asymmetries in mammalian genomes, we have revealed the existence of jumps in the noisy skew profiles resulting from asymmetries intrinsic to the transcription and replication processes [98,100]. This discovery has led us to extend our 1D WTMM methodology to an adapted multi-scale pattern recognition strategy in order to detect putative replication domains bordered by replication origins [101,114]. The results reported in this manuscript show that directly from the DNA sequence, we have been able to reveal the existence in the human genome (and very likely in all mammalian genomes), of regions bordered by early replicating origins in which gene position, orientation and expression breadth present a high level of organization, possibly mediated by the chromatin structure.

These results open new perspectives in DNA sequence analysis, chromatin modeling as well as in experiment. From a bioinformatic and modeling point of view, we plan to study the lexical and structural characteristics of our set of putative origins. In particular we will search for conserved sequence motifs in these replication initiation zones. Using a sequence-dependent model of DNA-histones interactions, we will develop physical studies of nucleosome formation and diffusion along the DNA fiber around the putative replication origins. These bioinformatic and physical studies, performed for the first time on a large number of replication origins, should shed light on the processes at work during the recognition of the replication initiation zone by the replication machinery. From an experimental point of view, our study raises new opportunities for future experiments. The first one concerns the experimental validation of the predicted replication origins (e. g. by molecular combing of DNA molecules [174]), which will allow us to determine precisely the existence of replication origins in given genome regions. Large scale study of all candidate origins is in current progress in the laboratory of O. Hyrien (École Normale Supérieure, Paris). The second experimental project consists in using Atomic Force Microscopy (AFM) [175] and Surface Plasmon Resonance Microscopy (SPRM) [176] to visualize and study the structural and mechanical properties of the DNA double helix, the nucleosomal string and the 30 nm chromatin fiber around the predicted replication origins. This work is in current progress in the experimental group of F. Argoul at the Laboratoire Joliot–Curie (ENS, Lyon) [83]. Finally the third experimental perspective concerns in situ studies of replication origins. Using fluorescence techniques (FISH chromosome painting [177]), we plan to study the distributions and dynamics of origins in the cell nucleus, as well as chromosome domains potentially associated with territories and their possible relation to nuclear matrix attachment sites. This study is likely to provide evidence of chromatin rosette patterns as suggested in [146]. This study is under progress in the molecular biology experimental group of F. Mongelard at the Laboratoire Joliot–Curie.

Acknowledgments

Bibliography

Primary Literature

1. Goupillaud P, Grossmann A, Morlet J (1984) Cycle-octave and related transforms in seismic signal analysis. Geoexploration 23:85–102
2. Grossmann A, Morlet J (1984) Decomposition of Hardy functions into square integrable wavelets of constant shape. SIAM J Math Anal 15:723–736
3. Arneodo A, Argoul F, Bacry E, Elezgaray J, Freysz E, Grasseau G, Muzy J-F, Pouligny B (1992) Wavelet transform of fractals. In: Meyer Y (ed) Wavelets and applications. Springer, Berlin, pp 286–352
4. Arneodo A, Argoul F, Elezgaray J, Grasseau G (1989) Wavelet transform analysis of fractals: Application to nonequilibrium phase transitions. In: Turchetti G (ed) Nonlinear dynamics. World Scientific, Singapore, pp 130–180
5. Arneodo A, Grasseau G, Holschneider M (1988) Wavelet transform of multifractals. Phys Rev Lett 61:2281–2284
6. Holschneider M (1988) On the wavelet transform of fractal objects. J Stat Phys 50:963–993
7. Holschneider M, Tchamitchian P (1990) Régularité locale de la fonction non-différentiable de Riemann. In: Lemarié PG (ed) Les ondelettes en 1989. Springer, Berlin, pp 102–124
8. Jaffard S (1989) Hölder exponents at given points and wavelet coefficients. C R Acad Sci Paris Sér. I 308:79–81
9. Jaffard S (1991) Pointwise smoothness, two-microlocalization and wavelet coefficients. Publ Mat 35:155–168
10. Mallat S, Hwang W (1992) Singularity detection and processing with wavelets. IEEE Trans Info Theory 38:617–643
11. Mallat S, Zhong S (1992) Characterization of signals from multiscale edges. IEEE Trans Patt Recog Mach Intell 14:710–732
12. Arneodo A, Bacry E, Muzy J-F (1995) The thermodynamics of fractals revisited with wavelets. Physica A 213:232–275
13. Bacry E, Muzy J-F, Arneodo A (1993) Singularity spectrum of fractal signals from wavelet analysis: Exact results. J Stat Phys 70:635–674

14. Muzy J-F, Bacry E, Arneodo A (1991) Wavelets and multifractal formalism for singular signals: Application to turbulence data. Phys Rev Lett 67:3515–3518

15. Muzy J-F, Bacry E, Arneodo A (1993) Multifractal formalism for fractal signals: The structure-function approach versus the wavelet-transform modulus-maxima method. Phys Rev E 47:875–884

16. Muzy J-F, Bacry E, Arneodo A (1994) The multifractal formalism revisited with wavelets. Int J Bifurc Chaos 4:245–302

17. Jaffard S (1997) Multifractal formalism for functions part I: Results valid for all functions. SIAM J Math Anal 28:944–970

18. Jaffard S (1997) Multifractal formalism for functions part II: Self-similar functions. SIAM J Math Anal 28:971–998

19. Hentschel HGE (1994) Stochastic multifractality and universal scaling distributions. Phys Rev E 50:243–261

20. Arneodo A, Audit B, Decoster N, Muzy J-F, Vaillant C (2002) Wavelet based multifractal formalism: Application to DNA sequences, satellite images of the cloud structure and stock market data. In: Bunde A, Kropp J, Schellnhuber HJ (eds) The science of disasters: Climate disruptions, heart attacks, and market crashes. Springer, Berlin, pp 26–102

21. Arneodo A, Manneville S, Muzy J-F (1998) Towards log-normal statistics in high Reynolds number turbulence. Eur Phys J B 1:129–140

22. Arneodo A, Manneville S, Muzy J-F, Roux SG (1999) Revealing a lognormal cascading process in turbulent velocity statistics with wavelet analysis. Phil Trans R Soc Lond A 357:2415–2438

23. Delour J, Muzy J-F, Arneodo A (2001) Intermittency of 1D velocity spatial profiles in turbulence: A magnitude cumulant analysis. Eur Phys J B 23:243–248

24. Roux S, Muzy J-F, Arneodo A (1999) Detecting vorticity filaments using wavelet analysis: About the statistical contribution of vorticity filaments to intermittency in swirling turbulent flows. Eur Phys J B 8:301–322

25. Venugopal V, Roux SG, Foufoula-Georgiou E, Arneodo A (2006) Revisiting multifractality of high-resolution temporal rainfall using a wavelet-based formalism. Water Resour Res 42:W06D14

26. Venugopal V, Roux SG, Foufoula-Georgiou E, Arneodo A (2006) Scaling behavior of high resolution temporal rainfall: New insights from a wavelet-based cumulant analysis. Phys Lett A 348:335–345

27. Arneodo A, d'Aubenton-Carafa Y, Bacry E, Graves PV, Muzy J-F, Thermes C (1996) Wavelet based fractal analysis of DNA sequences. Physica D 96:291–320

28. Arneodo A, Bacry E, Graves PV, Muzy J-F (1995) Characterizing long-range correlations in DNA sequences from wavelet analysis. Phys Rev Lett 74:3293–3296

29. Audit B, Thermes C, Vaillant C, d'Aubenton Carafa Y, Muzy J-F, Arneodo A (2001) Long-range correlations in genomic DNA: A signature of the nucleosomal structure. Phys Rev Lett 86:2471–2474

30. Audit B, Vaillant C, Arneodo A, d'Aubenton-Carafa Y, Thermes C (2002) Long-range correlations between DNA bending sites: Relation to the structure and dynamics of nucleosomes. J Mol Biol 316:903–918

31. Arneodo A, Muzy J-F, Sornette D (1998) "Direct" causal cascade in the stock market. Eur Phys J B 2:277–282

32. Muzy J-F, Sornette D, Delour J, Arneodo A (2001) Multifractal returns and hierarchical portfolio theory. Quant Finance 1:131–148

33. Ivanov PC, Amaral LA, Goldberger AL, Havlin S, Rosenblum MG, Struzik ZR, Stanley HE (1999) Multifractality in human heartbeat dynamics. Nature 399:461–465

34. Ivanov PC, Rosenblum MG, Peng CK, Mietus J, Havlin S, Stanley HE, Goldberger AL (1996) Scaling behavior of heartbeat intervals obtained by wavelet-based time-series analysis. Nature 383:323–327

35. Arneodo A, Argoul F, Bacry E, Muzy J-F, Tabard M (1992) Golden mean arithmetic in the fractal branching of diffusion-limited aggregates. Phys Rev Lett 68:3456–3459

36. Arneodo A, Argoul F, Muzy J-F, Tabard M (1992) Structural 5-fold symmetry in the fractal morphology of diffusion-limited aggregates. Physica A 188:217–242

37. Arneodo A, Argoul F, Muzy J-F, Tabard M (1992) Uncovering Fibonacci sequences in the fractal morphology of diffusion-limited aggregates. Phys Lett A 171:31–36

38. Kuhn A, Argoul F, Muzy J-F, Arneodo A (1994) Structural-analysis of electroless deposits in the diffusion-limited regime. Phys Rev Lett 73:2998–3001

39. Arneodo A, Decoster N, Roux SG (2000) A wavelet-based method for multifractal image analysis, I. Methodology and test applications on isotropic and anisotropic random rough surfaces. Eur Phys J B 15:567–600

40. Arrault J, Arneodo A, Davis A, Marshak A (1997) Wavelet based multifractal analysis of rough surfaces: Application to cloud models and satellite data. Phys Rev Lett 79:75–78

41. Decoster N, Roux SG, Arneodo A (2000) A wavelet-based method for multifractal image analysis, II. Applications to synthetic multifractal rough surfaces. Eur Phys J B 15:739–764

42. Arneodo A, Decoster N, Roux SG (1999) Intermittency, lognormal statistics, and multifractal cascade process in high-resolution satellite images of cloud structure. Phys Rev Lett 83:1255–1258

43. Roux SG, Arneodo A, Decoster N (2000) A wavelet-based method for multifractal image analysis, III. Applications to high-resolution satellite images of cloud structure. Eur Phys J B 15:765–786

44. Khalil A, Joncas G, Nekka F, Kestener P, Arneodo A (2006) Morphological analysis of H_I features, II. Wavelet-based multifractal formalism. Astrophys J Suppl Ser 165:512–550

45. Kestener P, Lina J-M, Saint-Jean P, Arneodo A (2001) Wavelet-based multifractal formalism to assist in diagnosis in digitized mammograms. Image Anal Stereol 20:169–174

46. Arneodo A, Decoster N, Kestener P, Roux SG (2003) A wavelet-based method for multifractal image analysis: From theoretical concepts to experimental applications. Adv Imaging Electr Phys 126:1–92

47. Kestener P, Arneodo A (2003) Three-dimensional wavelet-based multifractal method: The need for revisiting the multifractal description of turbulence dissipation data. Phys Rev Lett 91:194501

48. Meneveau C, Sreenivasan KR (1991) The multifractal nature of turbulent energy-dissipation. J Fluid Mech 224:429–484

49. Kestener P, Arneodo A (2004) Generalizing the wavelet-based multifractal formalism to random vector fields: Application to three-dimensional turbulence velocity and vorticity data. Phys Rev Lett 93:044501

50. Kestener P, Arneodo A (2007) A multifractal formalism for vector-valued random fields based on wavelet analysis: Application to turbulent velocity and vorticity 3D nu-

merical data. Stoch Environ Res Risk Assess. doi:10.1007/s00477-007-0121-6

51. Li WT, Marr TG, Kaneko K (1994) Understanding long-range correlations in DNA-sequences. Physica D 75:392–416
52. Stanley HE, Buldyrev SV, Goldberger AL, Havlin S, Ossadnik SM, Peng C-K, Simons M (1993) Fractal landscapes in biological systems. Fractals 1:283–301
53. Li W (1990) Mutual information functions versus correlation-functions. J Stat Phys 60:823–837
54. Li W (1992) Generating non trivial long-range correlations and $1/f$ spectra by replication and mutation. Int J Bifurc Chaos 2:137–154
55. Azbel' MY (1995) Universality in a DNA statistical structure. Phys Rev Lett 75:168–171
56. Herzel H, Große I (1995) Measuring correlations in symbol sequences. Physica A 216:518–542
57. Voss RF (1992) Evolution of long-range fractal correlations and $1/f$ noise in DNA base sequences. Phys Rev Lett 68:3805–3808
58. Voss RF (1994) Long-range fractal correlations in DNA introns and exons. Fractals 2:1–6
59. Peng C-K, Buldyrev SV, Goldberger AL, Havlin S, Sciortino F, Simons M, Stanley HE (1992) Long-range correlations in nucleotide sequences. Nature 356:168–170
60. Havlin S, Buldyrev SV, Goldberger AL, Mantegna RN, Peng C-K, Simons M, Stanley HE (1995) Statistical and linguistic features of DNA sequences. Fractals 3:269–284
61. Mantegna RN, Buldyrev SV, Goldberger AL, Havlin S, Peng C-K, Simons M, Stanley HE (1995) Systematic analysis of coding and noncoding DNA sequences using methods of statistical linguistics. Phys Rev E 52:2939–2950
62. Herzel H, Ebeling W, Schmitt A (1994) Entropies of biosequences: The role of repeats. Phys Rev E 50:5061–5071
63. Li W (1997) The measure of compositional heterogeneity in DNA sequences is related to measures of complexity. Complexity 3:33–37
64. Borštnik B, Pumpernik D, Lukman D (1993) Analysis of apparent $1/f^{\alpha}$ spectrum in DNA sequences. Europhys Lett 23:389–394
65. Chatzidimitriou-Dreismann CA, Larhammar D (1993) Long-range correlations in DNA. Nature 361:212–213
66. Nee S (1992) Uncorrelated DNA walks. Nature 357:450
67. Viswanathan GM, Buldyrev SV, Havlin S, Stanley HE (1998) Long-range correlation measures for quantifying patchiness: Deviations from uniform power-law scaling in genomic DNA. Physica A 249:581–586
68. Buldyrev SV, Goldberger AL, Havlin S, Mantegna RN, Matsa ME, Peng C-K, Simons M, Stanley HE (1995) Long-range correlation properties of coding and noncoding DNA sequences: GenBank analysis. Phys Rev E 51:5084–5091
69. Berthelsen CL, Glazier JA, Raghavachari S (1994) Effective multifractal spectrum of a random walk. Phys Rev E 49:1860–1864
70. Li W (1997) The study of correlation structures of DNA sequences: A critical review. Comput Chem 21:257–271
71. Peng C-K, Buldyrev SV, Goldberger AL, Havlin S, Simons M, Stanley HE (1993) Finite-size effects on long-range correlations: Implications for analyzing DNA sequences. Phys Rev E 47:3730–3733
72. Bernardi G (2000) Isochores and the evolutionary genomics of vertebrates. Gene 241:3–17
73. Gardiner K (1996) Base composition and gene distribution: Critical patterns in mammalian genome organization. Trends Genet 12:519–524
74. Li W, Stolovitzky G, Bernaola-Galván P, Oliver JL (1998) Compositional heterogeneity within, and uniformity between, DNA sequences of yeast chromosomes. Genome Res 8:916–928
75. Karlin S, Brendel V (1993) Patchiness and correlations in DNA sequences. Science 259:677–680
76. Larhammar D, Chatzidimitriou-Dreismann CA (1993) Biological origins of long-range correlations and compositional variations in DNA. Nucleic Acids Res 21:5167–5170
77. Peng C-K, Buldyrev SV, Havlin S, Simons M, Stanley HE, Goldberger AL (1994) Mosaic organization of DNA nucleotides. Phys Rev E 49:1685–1689
78. Arneodo A, d'Aubenton-Carafa Y, Audit B, Bacry E, Muzy J-F, Thermes C (1998) Nucleotide composition effects on the long-range correlations in human genes. Eur Phys J B 1:259–263
79. Vaillant C, Audit B, Arneodo A (2005) Thermodynamics of DNA loops with long-range correlated structural disorder. Phys Rev Lett 95:068101
80. Vaillant C, Audit B, Thermes C, Arneodo A (2006) Formation and positioning of nucleosomes: effect of sequence-dependent long-range correlated structural disorder. Eur Phys J E 19:263–277
81. Yuan G-C, Liu Y-J, Dion MF, Slack MD, Wu LF, Altschuler SJ, Rando OJ (2005) Genome-scale identification of nucleosome positions in S. cerevisiae. Science 309:626–630
82. Vaillant C, Audit B, Arneodo A (2007) Experiments confirm the influence of genome long-range correlations on nucleosome positioning. Phys Rev Lett 99:218103
83. Moukhtar J, Fontaine E, Faivre-Moskalenko C, Arneodo A (2007) Probing persistence in DNA curvature properties with atomic force microscopy. Phys Rev Lett 98:178101
84. Chargaff E (1951) Structure and function of nucleic acids as cell constituents. Fed Proc 10:654–659
85. Rudner R, Karkas JD, Chargaff E (1968) Separation of B. subtilis DNA into complementary strands, 3. Direct analysis. Proc Natl Acad Sci USA 60:921–922
86. Fickett JW, Torney DC, Wolf DR (1992) Base compositional structure of genomes. Genomics 13:1056–1064
87. Lobry JR (1995) Properties of a general model of DNA evolution under no-strand-bias conditions. J Mol Evol 40:326–330
88. Beletskii A, Grigoriev A, Joyce S, Bhagwat AS (2000) Mutations induced by bacteriophage T7 RNA polymerase and their effects on the composition of the T7 genome. J Mol Biol 300:1057–1065
89. Francino MP, Ochman H (2001) Deamination as the basis of strand-asymmetric evolution in transcribed Escherichia coli sequences. Mol Biol Evol 18:1147–1150
90. Frank AC, Lobry JR (1999) Asymmetric substitution patterns: A review of possible underlying mutational or selective mechanisms. Gene 238:65–77
91. Freeman JM, Plasterer TN, Smith TF, Mohr SC (1998) Patterns of genome organization in bacteria. Science 279:1827
92. Mrázek J, Karlin S (1998) Strand compositional asymmetry in bacterial and large viral genomes. Proc Natl Acad Sci USA 95:3720–3725
93. Rocha EP, Danchin A, Viari A (1999) Universal replication biases in bacteria. Mol Microbiol 32:11–16

94. Tillier ER, Collins RA (2000) The contributions of replication orientation, gene direction, and signal sequences to base-composition asymmetries in bacterial genomes. J Mol Evol 50:249–257

95. Green P, Ewing B, Miller W, Thomas PJ, Green ED (2003) Transcription-associated mutational asymmetry in mammalian evolution. Nat Genet 33:514–517

96. Touchon M, Nicolay S, Arneodo A, d'Aubenton-Carafa Y, Thermes C (2003) Transcription-coupled TA and GC strand asymmetries in the human genome. FEBS Lett 555:579–582

97. Touchon M, Arneodo A, d'Aubenton-Carafa Y, Thermes C (2004) Transcription-coupled and splicing-coupled strand asymmetries in eukaryotic genomes. Nucleic Acids Res 32:4969–4978

98. Brodie of Brodie E-B, Nicolay S, Touchon M, Audit B, d'Aubenton-Carafa Y, Thermes C, Arneodo A (2005) From DNA sequence analysis to modeling replication in the human genome. Phys Rev Lett 94:248103

99. Nicolay S, Argoul F, Touchon M, d'Aubenton-Carafa Y, Thermes C, Arneodo A (2004) Low frequency rhythms in human DNA sequences: A key to the organization of gene location and orientation? Phys Rev Lett 93:108101

100. Touchon M, Nicolay S, Audit B, Brodie of Brodie E-B, d'Aubenton-Carafa Y, Arneodo A, Thermes C (2005) Replication-associated strand asymmetries in mammalian genomes: Toward detection of replication origins. Proc Natl Acad Sci USA 102:9836–9841

101. Huvet M, Nicolay S, Touchon M, Audit B, d'Aubenton-Carafa Y, Arneodo A, Thermes C (2007) Human gene organization driven by the coordination of replication and transcription. Genome Res 17:1278–1285

102. Arneodo A, Bacry E, Jaffard S, Muzy J-F (1997) Oscillating singularities on Cantor sets: A grand-canonical multifractal formalism. J Stat Phys 87:179–209

103. Arneodo A, Bacry E, Jaffard S, Muzy J-F (1998) Singularity spectrum of multifractal functions involving oscillating singularities. J Fourier Anal Appl 4:159–174

104. Parisi G, Frisch U (1985) Fully developed turbulence and intermittency. In: Ghil M, Benzi R, Parisi G (eds) Turbulence and predictability in geophysical fluid dynamics and climate dynamics. Proc of Int School. North-Holland, Amsterdam, pp 84–88

105. Collet P, Lebowitz J, Porzio A (1987) The dimension spectrum of some dynamical systems. J Stat Phys 47:609–644

106. Grassberger P, Badii R, Politi A (1988) Scaling laws for invariant measures on hyperbolic and non hyperbolic attractors. J Stat Phys 51:135–178

107. Halsey TC, Jensen MH, Kadanoff LP, Procaccia I, Shraiman BI (1986) Fractal measures and their singularities: The characterization of strange sets. Phys Rev A 33:1141–1151

108. Paladin G, Vulpiani A (1987) Anomalous scaling laws in multifractal objects. Phys Rep 156:147–225

109. Rand D (1989) The singularity spectrum for hyperbolic Cantor sets and attractors. Ergod Th Dyn Syst 9:527–541

110. Argoul F, Arneodo A, Elezgaray J, Grasseau G (1990) Wavelet analysis of the self-similarity of diffusion-limited aggregates and electrodeposition clusters. Phys Rev A 41:5537–5560

111. Farmer JD, Ott E, Yorke JA (1983) The dimension of chaotic attractors. Physica D 7:153–180

112. Grassberger P, Procaccia I (1983) Measuring the strangeness of strange attractors. Physica D 9:189–208

113. Bohr T, Tèl T (1988) The thermodynamics of fractals. In: Hao BL (ed) Direction in chaos, vol 2. World Scientific, Singapore, pp 194–237

114. Audit B, Nicolay S, Huvet M, Touchon M, d'Aubenton-Carafa Y, Thermes C, Arneodo A (2007) DNA replication timing data corroborate in silico human replication origin predictions. Phys Rev Lett 99:248102

115. Mandelbrot BB, van Ness JW (1968) Fractional Brownian motions, fractal noises and applications. SIAM Rev 10:422–437

116. Arneodo A, Bacry E, Muzy JF (1998) Random cascades on wavelet dyadic trees. J Math Phys 39:4142–4164

117. Benzi R, Biferale L, Crisanti A, Paladin G, Vergassola M, Vulpiani A (1993) A random process for the construction of multiaffine fields. Physica D 65:352–358

118. Mandelbrot BB (1974) Intermittent turbulence in self-similar cascades: Divergence of high moments and dimension of the carrier. J Fluid Mech 62:331–358

119. Arneodo A, Bacry E, Manneville S, Muzy JF (1998) Analysis of random cascades using space-scale correlation functions. Phys Rev Lett 80:708–711

120. Castaing B, Dubrulle B (1995) Fully-developed turbulence – A unifying point-of-view. J Phys II France 5:895–899

121. Novikov EA (1994) Infinitely divisible distributions in turbulence. Phys Rev E 50:3303–3305

122. Gojobori T, Li WH, Graur D (1982) Patterns of nucleotide substitution in pseudogenes and functional genes. J Mol Evol 18:360–369

123. Li WH, Wu CI, Luo CC (1984) Nonrandomness of point mutation as reflected in nucleotide substitutions in pseudogenes and its evolutionary implications. J Mol Evol 21:58–71

124. Petrov DA, Hartl DL (1999) Patterns of nucleotide substitution in Drosophila and mammalian genomes. Proc Natl Acad Sci USA 96:1475–1479

125. Zhang Z, Gerstein M (2003) Patterns of nucleotide substitution, insertion and deletion in the human genome inferred from pseudogenes. Nucleic Acids Res 31:5338–5348

126. Duret L (2002) Evolution of synonymous codon usage in metazoans. Curr Opin Genet Dev 12:640–649

127. Shioiri C, Takahata N (2001) Skew of mononucleotide frequencies, relative abundance of dinucleotides, and DNA strand asymmetry. J Mol Evol 53:364–376

128. Svejstrup JQ (2002) Mechanisms of transcription-coupled DNA repair. Nat Rev Mol Cell Biol 3:21–29

129. Nicolay S, Brodie of Brodie E-B, Touchon M, Audit B, d'Aubenton-Carafa Y, Thermes C, Arneodo A (2007) Bifractality of human DNA strand-asymmetry profiles results from transcription. Phys Rev E 75:032902

130. Lee TI, Jenner RG, Boyer LA, Guenther MG, Levine SS, Kumar RM, Chevalier B, Johnstone SE, Cole MF, ichi Isono K, Koseki H, Fuchikami T, Abe K, Murray HL, Zucker JP, Yuan B, Bell GW, Herbolsheimer E, Hannett NM, Sun K, Odom DT, Otte AP, Volkert TL, Bartel DP, Melton DA, Gifford DK, Jaenisch R, Young RA (2006) Control of developmental regulators by polycomb in human embryonic stem cells. Cell 125:301–313

131. Jacob F, Brenner S, Cuzin F (1963) On the regulation of DNA replication in bacteria. Cold Spring Harb Symp Quant Biol 28:329–342

132. Bell SP, Dutta A (2002) DNA replication in eukaryotic cells. Annu Rev Biochem 71:333–374

133. Anglana M, Apiou F, Bensimon A, Debatisse M (2003) Dynamics of DNA replication in mammalian somatic cells: Nu-

cleotide pool modulates origin choice and interorigin spacing. Cell 114:385–394

134. Fisher D, Méchali M (2003) Vertebrate HoxB gene expression requires DNA replication. EMBO J 22:3737–3748

135. Gerbi SA, Bielinsky AK (2002) DNA replication and chromatin. Curr Opin Genet Dev 12:243–248

136. Hyrien O, Méchali M (1993) Chromosomal replication initiates and terminates at random sequences but at regular intervals in the ribosomal DNA of Xenopus early embryos. EMBO J 12:4511–4520

137. Schübeler D, Scalzo D, Kooperberg C, van Steensel B, Delrow J, Groudine M (2002) Genome-wide DNA replication profile for Drosophila melanogaster: A link between transcription and replication timing. Nat Genet 32:438–442

138. Gilbert DM (2001) Making sense of eukaryotic DNA replication origins. Science 294:96–100

139. Coverley D, Laskey RA (1994) Regulation of eukaryotic DNA replication. Annu Rev Biochem 63:745–776

140. Sasaki T, Sawado T, Yamaguchi M, Shinomiya T (1999) Specification of regions of DNA replication initiation during embryogenesis in the 65-kilobase DNApolalpha-dE2F locus of Drosophila melanogaster. Mol Cell Biol 19:547–555

141. Bogan JA, Natale DA, Depamphilis ML (2000) Initiation of eukaryotic DNA replication: Conservative or liberal? J Cell Physiol 184:139–150

142. Gilbert DM (2004) In search of the holy replicator. Nat Rev Mol Cell Biol 5:848–855

143. Demeret C, Vassetzky Y, Méchali M (2001) Chromatin remodeling and DNA replication: From nucleosomes to loop domains. Oncogene 20:3086–3093

144. McNairn AJ, Gilbert DM (2003) Epigenomic replication: linking epigenetics to DNA replication. Bioessays 25:647–656

145. Méchali M (2001) DNA replication origins: From sequence specificity to epigenetics. Nat Rev Genet 2:640–645

146. Arneodo A, d'Aubenton-Carafa Y, Audit B, Brodie of Brodie E-B, Nicolay S, St-Jean P, Thermes C, Touchon M, Vaillant C (2007) DNA in chromatin: From genome-wide sequence analysis to the modeling of replication in mammals. Adv Chem Phys 135:203–252

147. Bulmer M (1991) Strand symmetry of mutation rates in the beta-globin region. J Mol Evol 33:305–310

148. Francino MP, Ochman H (2000) Strand symmetry around the beta-globin origin of replication in primates. Mol Biol Evol 17:416–422

149. Gierlik A, Kowalczuk M, Mackiewicz P, Dudek MR, Cebrat S (2000) Is there replication-associated mutational pressure in the Saccharomyces cerevisiae genome? J Theor Biol 202:305–314

150. Berezney R, Dubey DD, Huberman JA (2000) Heterogeneity of eukaryotic replicons, replicon clusters, and replication foci. Chromosoma 108:471–484

151. Vassilev LT, Burhans WC, DePamphilis ML (1990) Mapping an origin of DNA replication at a single-copy locus in exponentially proliferating mammalian cells. Mol Cell Biol 10:4685–4689

152. Codlin S, Dalgaard JZ (2003) Complex mechanism of site-specific DNA replication termination in fission yeast. EMBO J 22:3431–3440

153. Little RD, Platt TH, Schildkraut CL (1993) Initiation and termination of DNA replication in human rRNA genes. Mol Cell Biol 13:6600–6613

154. Santamaria D, Viguera E, Martinez-Robles ML, Hyrien O, Hernandez P, Krimer DB, Schvartzman JB (2000) Bi-directional replication and random termination. Nucleic Acids Res 28:2099–2107

155. White EJ, Emanuelsson O, Scalzo D, Royce T, Kosak S, Oakeley EJ, Weissman S, Gerstein M, Groudine M, Snyder M, Schübeler D (2004) DNA replication-timing analysis of human chromosome 22 at high resolution and different developmental states. Proc Natl Acad Sci USA 101:17771–17776

156. Woodfine K, Beare DM, Ichimura K, Debernardi S, Mungall AJ, Fiegler H, Collins VP, Carter NP, Dunham I (2005) Replication timing of human chromosome 6. Cell Cycle 4:172–176

157. Raghuraman MK, Winzeler EA, Collingwood D, Hunt S, Wodicka L, Conway A, Lockhart DJ, Davis RW, Brewer BJ, Fangman WL (2001) Replication dynamics of the yeast genome. Science 294:115–121

158. Watanabe Y, Fujiyama A, Ichiba Y, Hattori M, Yada T, Sakaki Y, Ikemura T (2002) Chromosome-wide assessment of replication timing for human chromosomes 11q and 21q: Disease-related genes in timing-switch regions. Hum Mol Genet 11:13–21

159. Costantini M, Clay O, Federico C, Saccone S, Auletta F, Bernardi G (2007) Human chromosomal bands: Nested structure, high-definition map and molecular basis. Chromosoma 116:29–40

160. Schmegner C, Hameister H, Vogel W, Assum G (2007) Isochores and replication time zones: A perfect match. Cytogenet Genome Res 116:167–172

161. Chakalova L, Debrand E, Mitchell JA, Osborne CS, Fraser P (2005) Replication and transcription: shaping the landscape of the genome. Nat Rev Genet 6:669–677

162. Gilbert N, Boyle S, Fiegler H, Woodfine K, Carter NP, Bickmore WA (2004) Chromatin architecture of the human genome: Gene-rich domains are enriched in open chromatin fibers. Cell 118:555–566

163. Hurst LD, Pál C, Lercher MJ (2004) The evolutionary dynamics of eukaryotic gene order. Nat Rev Genet 5:299–310

164. Sproul D, Gilbert N, Bickmore WA (2005) The role of chromatin structure in regulating the expression of clustered genes. Nat Rev Genet 6:775–781

165. MacAlpine DM, Rodriguez HK, Bell SP (2004) Coordination of replication and transcription along a Drosophila chromosome. Genes Dev 18:3094–3105

166. Danis E, Brodolin K, Menut S, Maiorano D, Girard-Reydet C, Méchali M (2004) Specification of a DNA replication origin by a transcription complex. Nat Cell Biol 6:721–730

167. DePamphilis ML (2005) Cell cycle dependent regulation of the origin recognition complex. Cell Cycle 4:70–79

168. Ghosh M, Liu G, Randall G, Bevington J, Leffak M (2004) Transcription factor binding and induced transcription alter chromosomal c-myc replicator activity. Mol Cell Biol 24:10193–10207

169. Lin CM, Fu H, Martinovsky M, Bouhassira E, Aladjem MI (2003) Dynamic alterations of replication timing in mammalian cells. Curr Biol 13:1019–1028

170. Jeon Y, Bekiranov S, Karnani N, Kapranov P, Ghosh S, MacAlpine D, Lee C, Hwang DS, Gingeras TR, Dutta A (2005) Temporal profile of replication of human chromosomes. Proc Natl Acad Sci USA 102:6419–6424

171. Deshpande AM, Newlon CS (1996) DNA replication fork pause sites dependent on transcription. Science 272:1030–1033

172. Takeuchi Y, Horiuchi T, Kobayashi T (2003) Transcription-dependent recombination and the role of fork collision in yeast rDNA. Genes Dev 17:1497–1506

173. Rocha EPC, Danchin A (2003) Essentiality, not expressiveness, drives gene-strand bias in bacteria. Nat Genet 34:377–378

174. Herrick J, Stanislawski P, Hyrien O, Bensimon A (2000) Replication fork density increases during DNA synthesis in X. laevis egg extracts. J Mol Biol 300:1133–1142

175. Zlatanova J, Leuba SH (2003) Chromatin fibers, one-at-a-time. J Mol Biol 331:1–19

176. Tassius C, Moskalenko C, Minard P, Desmadril M, Elezgaray J, Argoul F (2004) Probing the dynamics of a confined enzyme by surface plasmon resonance. Physica A 342:402–409

177. Müller WG, Rieder D, Kreth G, Cremer C, Trajanoski Z, McNally JG (2004) Generic features of tertiary chromatin structure as detected in natural chromosomes. Mol Cell Biol 24: 9359–9370

Books and Reviews

Fractals

Aharony A, Feder J (eds) (1989) Fractals in Physics, Essays in Honour of BB Mandelbrot. Physica D 38. North-Holland, Amsterdam

Avnir D (ed) (1988) The fractal approach to heterogeneous chemistry: surfaces, colloids, polymers. Wiley, New-York

Barabàsi AL, Stanley HE (1995) Fractals concepts in surface growth. Cambridge University Press, Cambridge

Ben Avraham D, Havlin S (2000) Diffusion and reactions in fractals and disordered systems. Cambridge University Press, Cambridge

Bouchaud J-P, Potters M (1997) Théorie des risques financiers. Cambridge University Press, Cambridge

Bunde A, Havlin S (eds) (1991) Fractals and disordered systems. Springer, Berlin

Bunde A, Havlin S (eds) (1994) Fractals in science. Springer, Berlin

Bunde A, Kropp J, Schellnhuber HJ (eds) (2002) The science of disasters: Climate disruptions, heart attacks and market crashes. Springer, Berlin

Family F, Meakin P, Sapoval B, Wood R (eds) (1995) Fractal aspects of materials. Material Research Society Symposium Proceedings, vol 367. MRS, Pittsburg

Family F, Vicsek T (1991) Dynamics of fractal surfaces. World Scientific, Singapore

Feder J (1988) Fractals. Pergamon, New-York

Frisch U (1995) Turbulence. Cambridge University Press, Cambridge

Mandelbrot BB (1982) The Fractal Geometry of Nature. Freeman, San Francisco

Mantegna RN, Stanley HE (2000) An introduction to econophysics. Cambridge University Press, Cambridge

Meakin P (1998) Fractals, scaling and growth far from equilibrium. Cambridge University Press, Cambridge

Peitgen HO, Jürgens H, Saupe D (1992) Chaos and fractals: New frontiers of science. Springer, New York

Peitgen HO, Saupe D (eds) (1987) The science of fractal images. Springer, New-York

Pietronero L, Tosatti E (eds) (1986) Fractals in physics. North-Holland, Amsterdam

Stanley HE, Osbrowski N (eds) (1986) On growth and form: Fractal and non-fractal patterns in physics. Martinus Nijhof, Dordrecht

Stanley HE, Ostrowski N (eds) (1988) Random fluctuations and pattern growth. Kluwer, Dordrecht

Vicsek T (1989) Fractal growth phenomena. World Scientific, Singapore

Vicsek T, Schlesinger M, Matsuchita M (eds) (1994) Fractals in natural science. World Scientific, Singapore

West BJ (1990) Fractal physiology and chaos in medicine. World Scientific, Singapore

West BJ, Deering W (1994) Fractal physiology for physicists: Levy statistics. Phys Rep 246:1–100

Wilkinson GG, Kanellopoulos J, Megier J (eds) (1995) Fractals in geoscience and remote sensing, image understanding research senes, vol 1. ECSC-EC-EAEC, Brussels

Wavelets

Abry P (1997) Ondelettes et turbulences. Diderot Éditeur, Art et Sciences, Paris

Arneodo A, Argoul F, Bacry E, Elezgaray J, Muzy J-F (1995) Ondelettes, multifractales et turbulences: de l'ADN aux croissances cristallines. Diderot Éditeur, Art et Sciences, Paris

Chui CK (1992) An introduction to wavelets. Academic Press, Boston

Combes J-M, Grossmann A, Tchamitchian P (eds) (1989) Wavelets. Springer, Berlin

Daubechies I (1992) Ten lectures on wavelets. SIAM, Philadelphia

Erlebacher G, Hussaini MY, Jameson LM (eds) (1996) Wavelets: Theory and applications. Oxford University Press, Oxford

Farge M, Hunt JCR, Vassilicos JC (eds) (1993) Wavelets, fractals and Fourier. Clarendon Press, Oxford

Flandrin P (1993) Temps-Fréquence. Hermès, Paris

Holschneider M (1996) Wavelets: An analysis tool. Oxford University Press, Oxford

Jaffard S, Meyer Y, Ryan RD (eds) (2001) Wavelets: Tools for science and technology. SIAM, Philadelphia

Lemarie PG (ed) (1990) Les ondelettes en 1989. Springer, Berlin

Mallat S (1998) A wavelet tour in signal processing. Academic Press, New-York

Meyer Y (1990) Ondelettes. Herman, Paris

Meyer Y (ed) (1992) Wavelets and applications. Springer, Berlin

Meyer Y, Roques S (eds) (1993) Progress in wavelets analysis and applications. Éditions Frontières, Gif-sur-Yvette

Ruskai MB, Beylkin G, Coifman R, Daubechies I, Mallat S, Meyer Y, Raphael L (eds) (1992) Wavelets and their applications. Jones and Barlett, Boston

Silverman BW, Vassilicos JC (eds) (2000) Wavelets: The key to intermittent information? Oxford University Press, Oxford

Torresani B (1998) Analyse continue par ondelettes. Éditions de Physique, Les Ulis

DNA and Chromatin

Alberts B, Watson J (1994) Molecular biology of the cell, 3rd edn. Garland Publishing, New-York

Calladine CR, Drew HR (1999) Understanding DNA. Academic Press, San Diego

Graur D, Li WH (1999) Fundamentals of molecular evolution. Sinauer Associates, Sunderland

Hartl DL, Jones EW (2001) Genetics: Analysis of genes and genomes. Jones and Bartlett, Sudbury

Kolchanov NA, Lim HA (1994) Computer analysis of genetic macromolecules: Structure, function and evolution. World Scientific, Singapore

Kornberg A, Baker TA (1992) DNA Replication. WH Freeman, New-York

Lewin B (1994) Genes V. Oxford University Press, Oxford

Sudbery P (1998) Human molecular genetics. Addison Wesley, Singapore

Van Holde, KE (1988) Chromatin. Springer, New-York

Watson JD, Gilman M, Witkowski J, Zoller M (1992) Recombinant DNA. Freeman, New-York

Wolfe AP (1998) Chromatin structure and function, 3rd edn. Academic Press, London

Fractal and Transfractal Scale-Free Networks

Hernán D. Rozenfeld, Lazaros K. Gallos,
Chaoming Song, Hernán A. Makse
Levich Institute and Physics Department,
City College of New York, New York, USA

Article Outline

Glossary

Degree of a node Number of edges incident to the node.

Scale-free network Network that exhibits a wide (usually power-law) distribution of the degrees.

Small-world network Network for which the diameter increases logarithmically with the number of nodes.

Distance The length (measured in number of links) of the shortest path between two nodes.

Box Group of nodes. In a *connected box* there exists a path within the box between any pair of nodes. Otherwise, the box is *disconnected*.

Box diameter The longest distance in a box.

Definition of the Subject

The explosion in the study of complex networks during the last decade has offered a unique view in the structure and behavior of a wide range of systems, spanning many different disciplines [1]. The importance of complex networks lies mainly in their simplicity, since they can represent practically any system with interactions in a unified way by stripping complicated details and retaining the main features of the system. The resulting networks include only *nodes*, representing the interacting agents and *links*, representing interactions. The term 'interactions' is used loosely to describe any possible way that causes two nodes to form a link. Examples can be real physical links, such as the wires connecting computers in the Internet or roads connecting cities, or alternatively they may be virtual links, such as links in WWW homepages or acquaintances in societies, where there is no physical medium actually connecting the nodes.

The field was pioneered by the famous mathematician P. Erdős many decades ago, when he greatly advanced graph theory [27]. The theory of networks would have perhaps remained a problem of mathematical beauty, if it was not for the discovery that a huge number of everyday life systems share many common features and can thus be described through a unified theory. The remarkable diversity of these systems incorporates artificially or man-made technological networks such as the Internet and the World Wide Web (WWW), social networks such as social acquaintances or sexual contacts, biological networks of natural origin, such as the network of protein interactions of Yeast [1,36], and a rich variety of other systems, such as proximity of words in literature [48], items that are bought by the same people [16] or the way modules are connected to create a piece of software, among many others.

The advances in our understanding of networks, combined with the increasing availability of many databases, allows us to analyze and gain deeper insight into the main characteristics of these complex systems. A large number of complex networks share the *scale-free* property [1,28], indicating the presence of few highly connected nodes (usually called hubs) and a large number of nodes with small degree. This feature alone has a great impact on the analysis of complex networks and has introduced a new way of understanding these systems. This property carries important implications in many everyday life problems, such as the way a disease spreads in communities of individuals, or the resilience and tolerance of networks under random and intentional attacks [19,20,21,31,59].

Although the scale-free property holds an undisputed importance, it has been shown to not completely determine the global structure of networks [6]. In fact, two networks that obey the same distribution of the degrees may dramatically differ in other fundamental structural properties, such as in correlations between degrees or in the average distance between nodes. Another fundamental prop-

erty, which is the focus of this article, is the presence of self-similarity or fractality. In simpler terms, we want to know whether a subsection of the network looks much the same as the whole [8,14,29,66]. In spite of the fact that in regular fractal objects the distinction between self-similarity and fractality is absent, in network theory we can distinguish the two terms: in a *fractal network* the number of boxes of a given size that are needed to completely cover the network scales with the box size as a power law, while a *self-similar network* is defined as a network whose degree distribution remains invariant under renormalization of the network (details on the renormalization process will be provided later). This essential result allows us to better understand the origin of important structural properties of networks such as the power-law degree distribution [35,62,63].

Introduction

Self-similarity is a property of fractal structures, a concept introduced by Mandelbrot and one of the fundamental mathematical results of the 20th century [29,45,66]. The importance of fractal geometry stems from the fact that these structures were recognized in numerous examples in Nature, from the coexistence of liquid/gas at the critical point of evaporation of water [11,39,65], to snowflakes, to the tortuous coastline of the Norwegian fjords, to the behavior of many complex systems such as economic data, or the complex patterns of human agglomeration [29,66].

Typically, real world scale-free networks exhibit the small-world property [1], which implies that the number of nodes increases exponentially with the diameter of the network, rather than the power-law behavior expected for self-similar structures. For this reason complex networks were believed to *not* be length-scale invariant or self-similar.

In 2005, C. Song, S. Havlin and H. Makse presented an approach to analyze complex networks, that reveals their self-similarity [62]. This result is achieved by the application of a renormalization procedure which coarse-grains the system into boxes containing nodes within a given size [62,64]. As a result, a power-law relation between the number of boxes needed to cover the network and the size of the box is found, defining a finite self-similar exponent. These fundamental properties, which are shown for the WWW, cellular and protein-protein interaction networks, help to understand the emergence of the scale-free property in complex networks. They suggest a common self-organization dynamics of diverse networks at different scales into a critical state and in turn bring together previously unrelated fields: the statistical physics of complex networks with renormalization group, fractals and critical phenomena.

Fractality in Real-World Networks

The study of real complex networks has revealed that many of them share some fundamental common properties. Of great importance is the form of the degree distribution for these networks, which is unexpectedly wide. This means that the degree of a node may assume values that span many decades. Thus, although the majority of nodes have a relatively small degree, there is a finite probability that a few nodes will have degree of the order of thousands or even millions. Networks that exhibit such a wide distribution $P(k)$ are known as *scale-free* networks, where the term refers to the absence of a characteristic scale in the degree k. This distribution very often obeys a power-law form with a degree exponent γ, usually in the range $2 < \gamma < 4$ [2],

$$P(k) \sim k^{-\gamma} . \tag{1}$$

A more generic property, that is usually inherent in scale-free networks but applies equally well to other types of networks, such as in Erdős–Rényi random graphs, is the *small-world* feature. Originally discovered in sociological studies [47], it is the generalization of the famous 'six degrees of separation' and refers to the very small network diameter. Indeed, in small-world networks a very small number of steps is required to reach a given node starting from any other node. Mathematically this is expressed by the slow (logarithmic) increase of the average diameter of the network, $\bar{\ell}$, with the total number of nodes N, $\bar{\ell} \sim \ln N$, where ℓ is the *shortest* distance between two nodes and defines the distance metric in complex networks [2,12,27,67], namely,

$$N \sim e^{\bar{\ell}/\ell_0} , \tag{2}$$

where ℓ_0 is a characteristic length.

These network characteristics have been shown to apply in many empirical studies of diverse systems [1,2,28]. The simple knowledge that a network has the scale-free and/or small-world property already enables us to qualitatively recognize many of its basic properties. However, structures that have the same degree exponents may still differ in other aspects [6]. For example, a question of fundamental importance is whether scale-free networks are also self-similar or fractals. The illustrations of scale-free networks (see, e. g,. Figs. 1 and 2b) seem to resemble traditional fractal objects. Despite this similarity, Eq. (2) definitely appears to contradict a basic property of fractality: fast increase of the diameter with the system size. More-

over, a fractal object should be self-similar or invariant under a scale transformation, which is again not clear in the case of scale-free networks where the scale has necessarily limited range. So, how is it even possible that fractal scale-free networks exist? In the following, we will see how these seemingly contradictory aspects can be reconciled.

Fractality and Self-Similarity

The classical theory of self-similarity requires a power-law relation between the number of nodes N and the diameter of a fractal object ℓ [8,14]. The fractal dimension can be calculated using either *box-counting* or *cluster-growing* techniques [66]. In the first method the network is covered with N_B boxes of linear size ℓ_B. The fractal dimension or box dimension d_B is then given by [29]:

$$N_B \sim \ell_B^{-d_B}. \tag{3}$$

In the second method, instead of covering the network with boxes, a random seed node is chosen and nodes centered at the seed are grown so that they are separated by a maximum distance ℓ. The procedure is then repeated by choosing many seed nodes at random and the average "mass" of the resulting clusters, $\langle M_c \rangle$ (defined as the number of nodes in the cluster) is calculated as a function of ℓ to obtain the following scaling:

$$\langle M_c \rangle \sim \ell^{d_f}, \tag{4}$$

defining the fractal cluster dimension d_f [29]. If we use Eq. (4) for a small-world network, then Eq. (2) readily implies that $d_f = \infty$. In other words, these networks cannot be characterized by a finite fractal dimension, and should be regarded as infinite-dimensional objects. If this were true, though, local properties in a part of the network would not be able to represent the whole system. Still, it is also well established that the scale-free nature is similar in different parts of the network. Moreover, a graphical representation of real-world networks allows us to see that those systems seem to be built by attaching (following some rule) copies of itself.

The answer lies in the inherent inhomogeneity of the network. In the classical case of a *homogeneous* system (such as a fractal percolation cluster) the degree distribution is very narrow and the two methods described above are fully equivalent, because of this local neighborhood invariance. Indeed, all boxes in the box-covering method are statistically similar with each other as well as with the
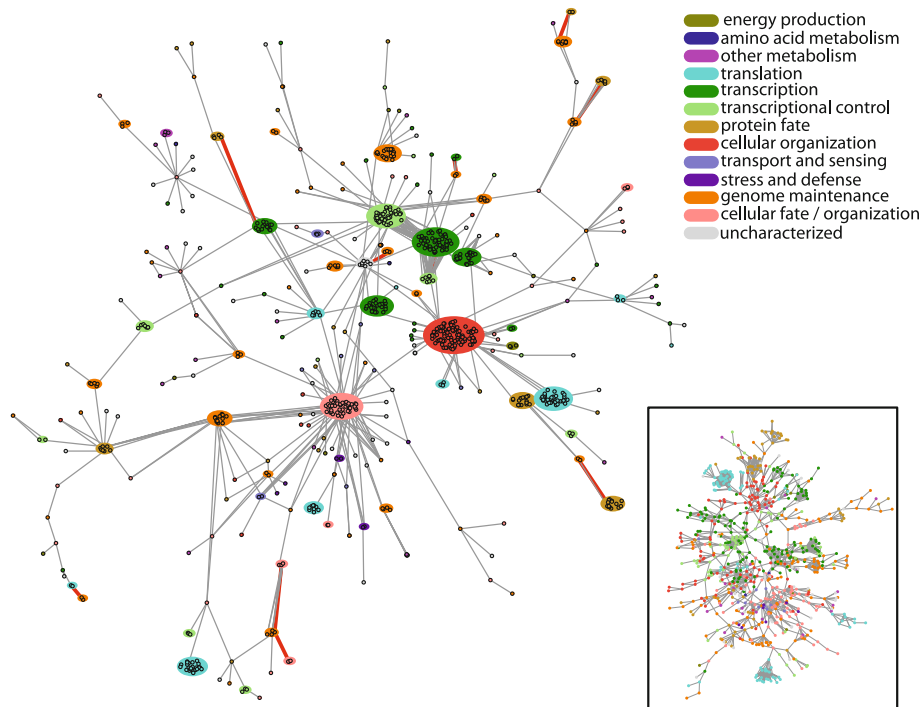
Fractal and Transfractal Scale-Free Networks, Figure 1
Representation of the Protein Interaction Network of Yeast. The colors show different subgroups of proteins that participate in different functionality classes [36]
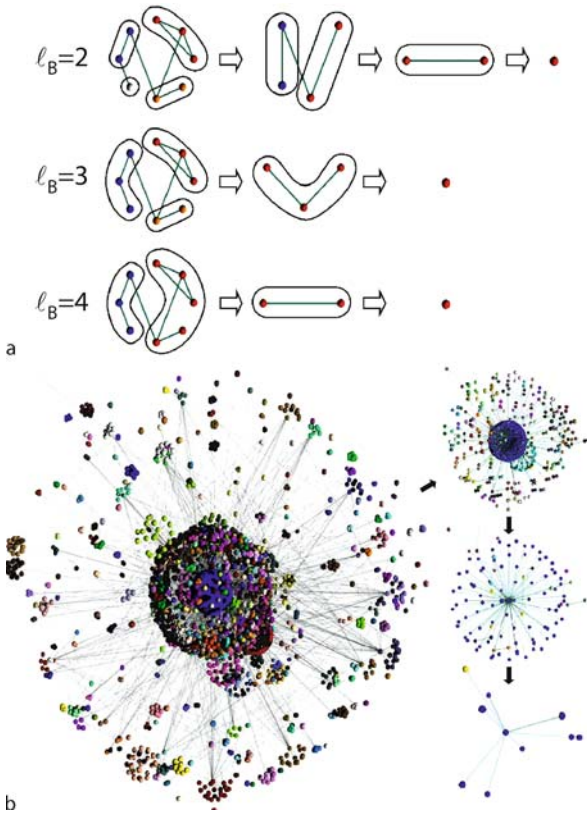
Fractal and Transfractal Scale-Free Networks, Figure 2
The renormalization procedure for complex networks. **a** Demonstration of the method for different ℓ_B and different stages in a network demo. The first column depicts the original network. The system is tiled with boxes of size ℓ_B (different colors correspond to different boxes). All nodes in a box are connected by a minimum distance smaller than the given ℓ_B. For instance, in the case of $\ell_B = 2$, one identifies four boxes which contain the nodes depicted with colors red, orange, white, and blue, each containing 3, 2, 1, and 2 nodes, respectively. Then each box is replaced by a single node; two renormalized nodes are connected if there is at least one link between the unrenormalized boxes. Thus we obtain the network shown in the second column. The resulting number of boxes needed to tile the network, $N_B(\ell_B)$, is plotted in Fig. 3 vs. ℓ_B to obtain d_B as in Eq. (3). The renormalization procedure is applied again and repeated until the network is reduced to a single node (third and fourth columns for different ℓ_B). **b** Three stages in the renormalization scheme applied to the entire WWW. We fix the box size to $\ell_B = 3$ and apply the renormalization for four stages. This corresponds, for instance, to the sequence for the network demo depicted in the second row in part a of this figure. We color the nodes in the web according to the boxes to which they belong

boxes grown when using the cluster-growing technique, so that Eq. (4) can be derived from Eq. (3) and $d_B = d_f$.

In *inhomogeneous* systems, though, the local environment can vary significantly. In this case, Eqs. (3) and (4)

are no longer equivalent. If we focus on the box-covering technique then we want to cover the entire network with the minimum possible number of boxes $N_B(\ell_B)$, where the distance between any two nodes that belong in a box is smaller than ℓ_B. An example is shown in Fig. 2a using a simple 8-node network. After we repeat this procedure for different values of ℓ_B we can plot N_B vs. ℓ_B.

When the box-covering method is applied to real large-scale networks, such as the WWW [2] (http://www.nd.edu/~networks), the network of protein interaction of *H. sapiens* and *E. coli* [25,68] and several cellular networks [38,52], then they follow Eq. (3) with a clear power-law, indicating the *fractal* nature of these systems (Figs. 3a,b,c). On the other hand when the method is applied to other real world networks such as the Internet [24] or the Barabási–Albert network [7], they do not satisfy Eq. (3), which manifests that these networks are *not* fractal.

The reason behind the discrepancy in the fractality of homogeneous and inhomogeneous systems can be better clarified studying the mass of the boxes. For a given ℓ_B value, the average mass of a box $\langle M_B(\ell_B) \rangle$ is

$$\langle M_B(\ell_B) \rangle \equiv \frac{N}{N_B(\ell_B)} \sim \ell_B^{d_B} , \qquad (5)$$

as also verified in Fig. 3 for several real-world networks. On the other hand, the average performed in the cluster growing method (averaging over single boxes without tiling the system) gives rise to an exponential growth of the mass

$$\langle M_c(\ell) \rangle \sim e^{\ell/\ell_1} , \qquad (6)$$

in accordance with the small-world effect, Eq. (2). Correspondingly, the probability distribution of the mass of the boxes M_B using box-covering is very broad, while the cluster-growing technique leads to a narrow probability distribution of M_c.

The topology of scale-free networks is dominated by several highly connected hubs—the nodes with the largest degree—implying that most of the nodes are connected to the hubs via one or very few steps. Therefore, the average performed in the cluster growing method is biased; the hubs are overrepresented in Eq. (6) since almost every node is a neighbor of a hub, and there is always a very large probability of including the same hubs in all clusters. On the other hand, the box covering method is a global tiling of the system providing a flat average over all the nodes, i. e. each part of the network is covered with an equal probability. Once a hub (or any node) is covered, it cannot be covered again.
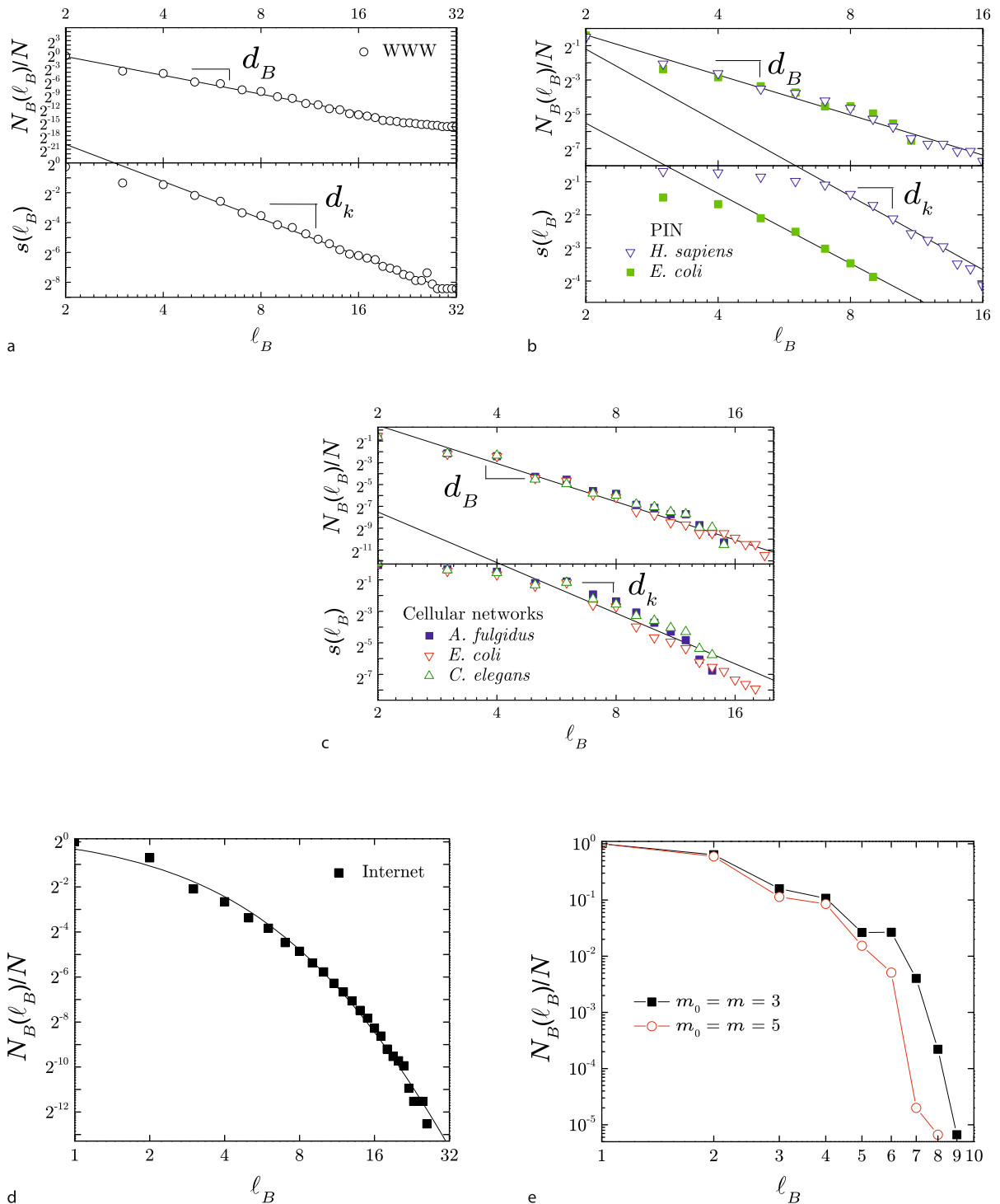
Fractal and Transfractal Scale-Free Networks, Figure 3
Self-similar scaling in complex networks. **a** *Upper panel*: Log-log plot of the N_B vs. ℓ_B revealing the self-similarity of the WWW according to Eq. (3). *Lower panel*: The scaling of $s(\ell_B)$ vs. ℓ_B according to Eq. (9). **b** Same as a but for two protein interaction networks: *H. sapiens* and *E. coli*. Results are analogous to b but with different scaling exponents. **c** Same as a for the cellular networks of *A. fulgidus*, *E. coli* and *C. elegans*. **d** Internet. Log-log plot of $N_B(\ell_B)$. The solid line shows that the internet [24] is *not* a fractal network since it does not follow the power-law relation of Eq. (5). **e** Same as **d** for the Barabási–Albert model network [7] with $m = 3$ and $m = 5$

In conclusion, we can state that the two dominant methods that are routinely used for calculations of fractality and give rise to Eqs. (3) and (4) are not equivalent in scale-free networks, but rather highlight different aspects: box covering reveals the self-similarity, while cluster growth reveals the small-world effect. The apparent contradiction is due to the hubs being used many times in the latter method.

Scale-free networks can be classified into three groups: (i) pure fractal, (ii) pure small-world and (iii) a mixture between fractal and small-world. (i) A fractal network satisfies Eq. (3) at all scales, meaning that for any value of ℓ_B, the number of boxes always follows a power-law (examples are shown in Fig. 3a,b,c). (ii) When a network is a pure small-world, it never satisfies Eq. (3). Instead, N_B follows an exponential decay with ℓ_B and the network cannot be regarded as fractal. Figures 3d and 3e show two examples of pure small-world networks. (iii) In the case of mixture between fractal and small-world, Eq. (3) is satisfied up to some cut-off value of ℓ_B, above which the fractality breaks down and the small-world property emerges. The small-world property is reflected in the plot of N_B vs. ℓ_B as an exponential cut-off for large ℓ_B.

We can also understand the coexistence of the small-world property and the fractality through a more intuitive approach. In a pure fractal network the length of a path between any pair of nodes scales as a power-law with the number of nodes in the network. Therefore, the diameter L also follows a power-law, $L \sim N^{1/d_B}$. If one adds a few shortcuts (links between randomly chosen nodes), many paths in the network are drastically shortened and the small-world property emerges as $L \sim \text{Log} N$. In spite of this fact, for shorter scales, $\ell_B \ll L$, the network still behaves as a fractal. In this sense, we can say that globally the network is small-world, but locally (for short scales) the network behaves as a fractal. As more shortcuts are added, the cut-off in a plot of N_B vs. ℓ_B appears for smaller ℓ_B, until the network becomes a pure small-world for which all paths lengths increase logarithmically with N.

The reasons why certain networks have evolved towards a fractal or non-fractal structure will be described later, together with models and examples that provide additional insight into the processes involved.

Renormalization

Renormalization is one of the most important techniques in modern Statistical Physics [17,39,58]. The idea behind this procedure is to continuously create smaller replicas of a given object, retaining at the same time the essential structural features, and hoping that the coarse-grained copies will be more amenable to analytic treatment.

The idea for renormalizing the network emerges naturally from the concept of fractality described above. If a network is self-similar, then it will look more or less the same under different scales. The way to observe these different length-scales is based on renormalization principles, while the criterion to decide on whether a renormalized structure retains its form is the invariance of the main structural features, expressed mainly through the degree distribution.

The method works as follows. Start by fixing the value of ℓ_B and applying the box-covering algorithm in order to cover the entire network with boxes (see Appendix). In the renormalized network each box is replaced by a single node and two nodes are connected if there existed at least one connection between the two corresponding boxes in the original network. The resulting structure represents the first stage of the renormalized network. We can apply the same procedure to this new network, as well, resulting in the second renormalization stage network, and so on until we are left with a single node.

The second column of the panels in Fig. 2a shows this step in the renormalization procedure for the schematic network, while Fig. 2b shows the results for the same procedure applied to the entire WWW for $\ell_B = 3$.

The renormalized network gives rise to a new probability distribution of links, $P(k')$ (we use a prime $'$ to denote quantities in the renormalized network). This distribution remains invariant under the renormalization:

$$P(k) \rightarrow P(k') \sim (k')^{-\gamma} . \tag{7}$$

Fig. 4 supports the validity of this scale transformation by showing a data collapse of all distributions with the same γ according to (7) for the WWW.

Here, we present the basic scaling relations that characterize renormalizable networks. The degree k' of each node in the renormalized network can be seen to scale with the largest degree k in the corresponding original box as

$$k \rightarrow k' = s(\ell_B)k . \tag{8}$$

This equation defines the scaling transformation in the connectivity distribution. Empirically, it was found that the scaling factor $s(< 1)$ scales with ℓ_B with a new exponent, d_k, as $s(\ell_B) \sim \ell_B^{-d_k}$, so that

$$k' \sim \ell_B^{-d_k} k , \tag{9}$$

This scaling is verified for many networks, as shown in Fig. 3.
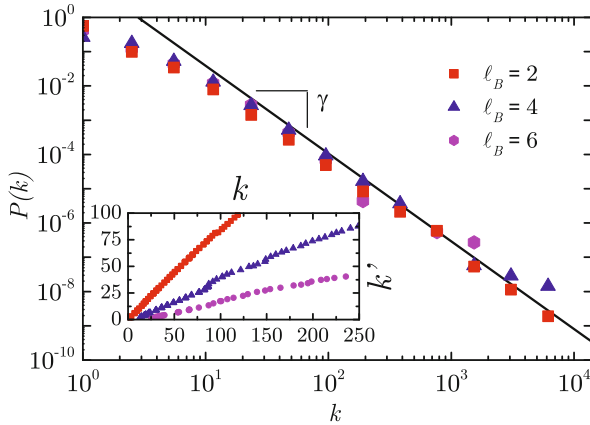
Fractal and Transfractal Scale-Free Networks, Figure 4
Invariance of the degree distribution of the WWW under the renormalization for different box sizes, ℓ_B. We show the data collapse of the degree distributions demonstrating the self-similarity at different scales. The *inset* shows the scaling of $k' = s(\ell_B)k$ for different ℓ_B, from where we obtain the scaling factor $s(\ell_B)$. Moreover, renormalization for a fixed box size ($\ell_B = 3$) is applied, until the network is reduced to a few nodes. It was found that $P(k)$ is invariant under these multiple renormalizations procedures

The exponents γ, d_B, and d_k are not all independent from each other. The proof starts from the density balance equation $n(k)dk = n'(k')dk'$, where $n(k) = NP(k)$ is the number of nodes with degree k and $n'(k') = N'P(k')$ is the number of nodes with degree k' after the renormalization (N' is the total number of nodes in the renormalized network). Substituting Eq. (8) leads to $N' = s^{\gamma-1}N$. Since the total number of nodes in the renormalized network is the number of boxes needed to cover the unrenormalized network at any given ℓ_B we have the identity $N' = N_B(\ell_B)$. Finally, from Eqs. (3) and (9) one obtains the relation between the three indexes

$$\gamma = 1 + \frac{d_B}{d_k} \, . \tag{10}$$

The use of Eq. (10) yields the same γ exponent as that obtained in the direct calculation of the degree distribution. The significance of this result is that the scale-free properties characterized by γ can be related to a more fundamental length-scale invariant property, characterized by the two new indexes d_B and d_k.

We have seen, thus, that concepts introduced originally for the study of critical phenomena in statistical physics, are also valid in the characterization of a different class of phenomena: the topology of complex networks. A large number of scale-free networks are fractals and an even larger number remain invariant under a scale-trans-

formation. The influence of these features on the network properties will be delayed until the sixth chapter, after we introduce some algorithms for efficient numerical calculations and two theoretical models that give rise to fractal networks.

Models: Deterministic Fractal and Transfractal Networks

The first model of a scale-free fractal network was presented in 1979 when N. Berker and S. Ostlund [9] proposed a hierarchical network that served as an exotic example where renormalization group techniques yield exact results, including the percolation phase transition and the $q \to 1$ limit of the Potts model. Unfortunately, in those days the importance of the power-law degree distribution and the concept of fractal and non-fractal complex networks were not known. Much work has been done on these types of hierarchical networks. For example, in 1984, M. Kaufman and R. Griffiths made use of Berker and Ostlund's model to study the percolation phase transition and its percolation exponents [22,37,40,41].

Since the late 90s, when the importance of the power-law degree distribution was first shown [1] and after the finding of C. Song, S. Havlin and H. Makse [62], many hierarchical networks that describe fractality in complex networks have been proposed. These artificial models are of great importance since they provide insight into the origins and fundamental properties that give rise to the fractality and non-fractality of networks.

The Song–Havlin–Makse Model

The correlations between degrees in a network [46,49,50,54] are quantified through the probability $P(k_1, k_2)$ that a node of degree k_1 is connected to another node of degree k_2. In Fig. 5 we can see the degree correlation profile $R(k_1, k_2) = P(k_1, k_2)/P_r(k_1, k_2)$ of the cellular metabolic network of E. *coli* [38] (known to be a fractal network) and for the Internet at the router level [15] (a non-fractal network), where $P_r(k_1, k_2)$ is obtained by randomly swapping the links without modifying the degree distribution.

Figure 5 shows a dramatic difference between the two networks. The network of E. *coli*, that is a fractal network, presents an anti-correlation of the degrees (or dissasortativity [49,50]), which means that mostly high degree nodes are linked to low degree nodes. This property leads to fractal networks. On the other hand, the Internet exhibits a high correlation between degrees leading to a non-fractal network.

With this idea in mind, in 2006 C. Song, S. Havlin and H. Makse presented a model that elucidates the way
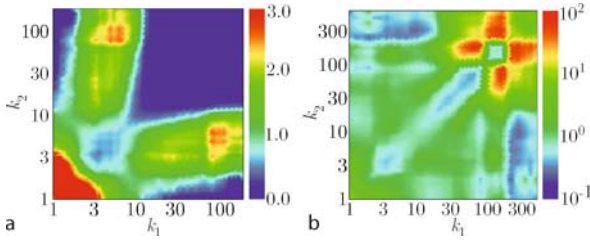
Fractal and Transfractal Scale-Free Networks, Figure 5
Degree correlation profile for a the cellular metabolic network of E. *coli*, and b the Internet at the router level
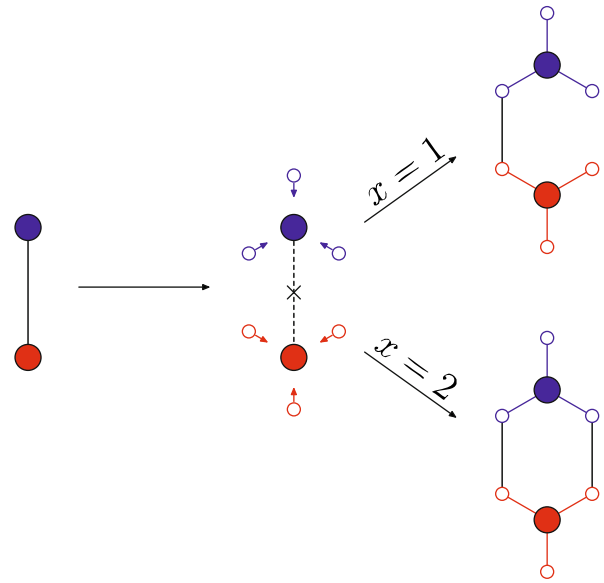

Fractal and Transfractal Scale-Free Networks, Figure 6
The model grows from a small network, usually two nodes connected to each other. During each step and for every link in the system, each endpoint of a link produces *m* offspring nodes (in this drawing $m = 3$). In this case, with probability $e = 1$ the original link is removed and *x* new links between randomly selected nodes of the new generation are added. Notice that the case of $x = 1$ results in a tree structure, while loops appear for $x > 1$

new nodes must be connected to the old ones in order to build a fractal, a non-fractal network, or a mixture between fractal and non-fractal network [63]. This model shows that, indeed, the correlations between degrees of the nodes are a determinant factor for the fractality of a network. This model was later extended [32] to allow loops in the network, while preserving the self-similarity and fractality properties.

The algorithm is as follows (see Fig. 6): In generation $n = 0$, start with two nodes connected by one link. Then, generation $n + 1$ is obtained recursively by attaching m new nodes to the endpoints of each link l of generation n. In addition, with probability e remove link l and add x new links connecting pairs of new nodes attached to the endpoints of l.

The degree distribution, diameter and fractal dimension can be easily calculated. For example, if $e = 1$ (pure fractal network), the degree distribution follows a power-law $P(k) \sim k^{-\gamma}$ with exponent $\gamma = 1 + \log(2m + x)/\log m$ and the fractal dimension is $d_B = \log(2m + x)/\log m$. The diameter L scales, in this case, as power of the number of nodes as $L \sim N^{1/d_B}$ [63,64]. Later, in Sect. "Properties of Fractal and Transfractal Networks", several topological properties are shown for this model network.

(*u*, *v*)-Flowers

In 2006, H. Rozenfeld, S. Havlin and D. ben-Avraham proposed a new family of recursive *deterministic* scale-free networks, the (*u*, *v*)-*flowers*, that generalize both, the original scale-free model of Berker and Ostlund [9] and the pseudo-fractal network of Dorogovstev, Goltsev and Mendes [26] and that, by appropriately varying its two parameters *u* and *v*, leads to either fractal networks or non-fractal networks [56,57]. The algorithm to build the (*u*, *v*)-*flowers* is the following: In generation $n = 1$ one starts with a cycle graph (a ring) consisting of $u + v \equiv w$ links and nodes (other choices are possible). Then, generation $n + 1$ is obtained recursively by replacing each link

by two parallel paths of *u* and *v* links long. Without loss of generality, $u \leq v$. Examples of (1, 3)- and (2, 2)-*flowers* are shown in Fig. 7. The DGM network corresponds to the special case of $u = 1$ and $v = 2$ and the Berker and Ostlund model corresponds to $u = 2$ and $v = 2$.

An essential property of the (*u*, *v*)-*flowers* is that they are self-similar, as evident from an equivalent method of construction: to produce generation $n + 1$, make $w = u + v$ copies of the net in generation *n* and join them at the hubs.

The number of links of a (*u*, *v*)-flower of generation *n* is

$$M_n = (u + v)^n = w^n , \tag{11}$$

and the number of nodes is

$$N_n = \left(\frac{w - 2}{w - 1}\right) w^n + \left(\frac{w}{w - 1}\right) . \tag{12}$$

The degree distribution of the (*u*, *v*)-*flowers* can also be easily obtained since by construction, (*u*, *v*)-*flowers* have only nodes of degree $k = 2^m$, $m = 1, 2, \dots, n$. As in the DGM case, (*u*, *v*)-*flowers* follow a scale-free degree distri-
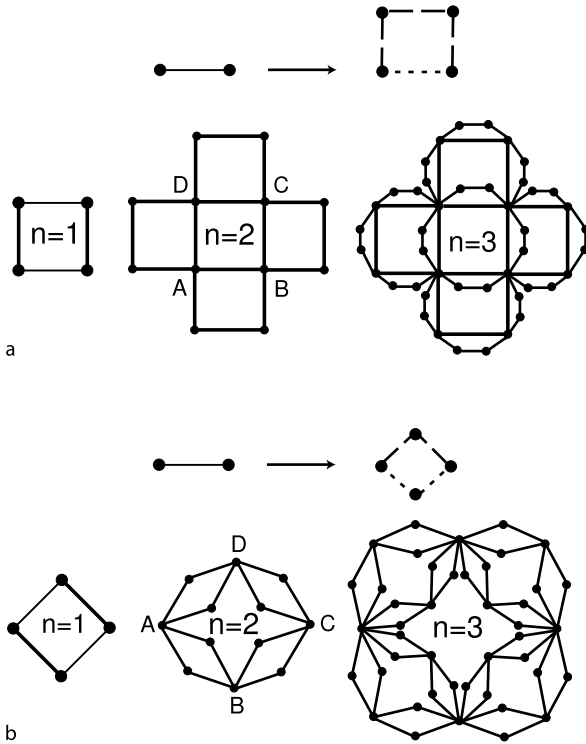
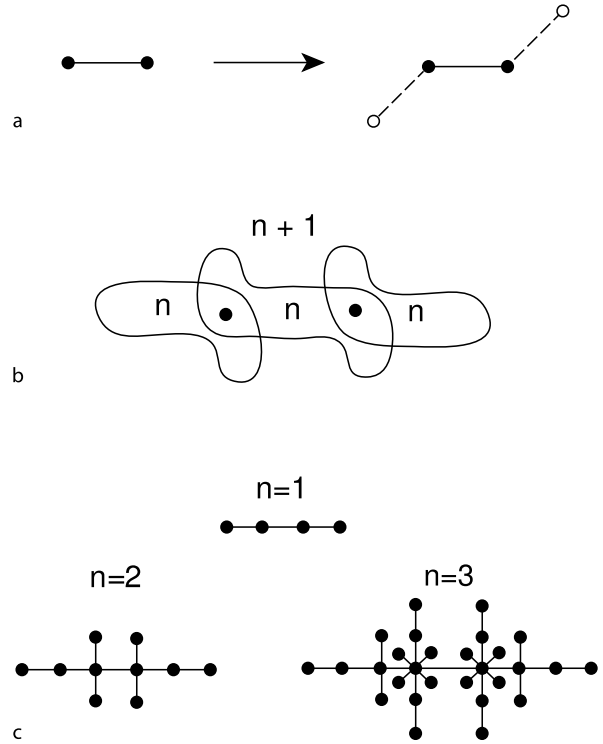**Fractal and Transfractal Scale-Free Networks, Figure 7**
(u, v)-*flowers* with $u + v = 4 (\gamma = 3)$. **a** $u = 1$ (*dotted line*) and $v = 3$ (*broken line*). **b** $u = 2$ and $v = 2$. The graphs may also be iterated by joining four replicas of generation n at the hubs A and B, for **a**, or A and C, for **b**

Fractal and Transfractal Scale-Free Networks, Figure 8
The $(1, 2)$-tree. **a** Each link in generation n is replaced by a chain of $u = 1$ links, to which ends one attaches chains of $v/2 = 1$ links. **b** Alternative method of construction highlighting self-similarity: $u + v = 3$ replicas of generation n are joined at the hubs. **c** Generations $n = 1, 2, 3$

bution, $P(k) \sim k^{-\gamma}$, of degree exponent

$$\gamma = 1 + \frac{\ln(u + v)}{\ln 2} . \tag{13}$$

Recursive scale-free *trees* may be defined in analogy to the flower nets. If v is even, one obtains generation $n + 1$ of a (u, v)-tree by replacing every link in generation n with a chain of u links, and attaching to each of its endpoints chains of $v/2$ links. Figure 8 shows how this works for the $(1, 2)$-tree. If v is odd, attach to the endpoints (of the chain of u links) chains of length $(v \pm 1)/2$. The trees may be also constructed by successively joining w replicas at the appropriate hubs, and they too are self-similar. They share many of the fundamental scaling properties with (u, v)-*flowers*: Their degree distribution is also scale-free, with the same degree exponent as (u, v)-*flowers*.

The self-similarity of (u, v)-*flowers*, coupled with the fact that different replicas meet at a *single* node, makes them amenable to exact analysis by renormalization techniques. The lack of loops, in the case of (u, v)-trees, further simplifies their analysis [9,13,56,57].

Dimensionality of the (u, v)-Flowers   There is a vast difference between (u, v)-nets with $u = 1$ and $u > 1$. If $u = 1$ the diameter L_n of the nth generation flower scales linearly with n. For example, L_n for the $(1, 2)$-flower [26] and $L_n = 2n$ for the $(1, 3)$-flower. It is easy to see that the diameter of the $(1, v)$-flower, for v odd, is $L_n = (v-1)n + (3-v)/2$, and, in general one can show that $L_n \sim (v-1)n$.

For $u > 1$, however, the diameter grows as a power of n. For example, for the $(2, 2)$-flower we find $L_n = 2^n$, and, more generally, the diameter satisfies $L_n \sim u^n$. To summarize,

$$L_n \sim \begin{cases} (v - 1)n & u = 1 , \\ u^n & u > 1 , \end{cases} \quad \text{flowers} . \tag{14}$$

Similar results are quite obvious for the case of (u, v)-trees, where

$$L_n \sim \begin{cases} vn & u = 1 , \\ u^n & u > 1 , \end{cases} \quad \text{trees} . \tag{15}$$

Since $N_n \sim (u + v)^n$ (Eq. (12)), we can recast these relations as

$$L \sim \begin{cases} \ln N & u = 1 , \\ N^{\ln u / \ln(u+v)} & u > 1 . \end{cases} \qquad (16)$$

Thus, (u, v)-nets are *small world* only in the case of $u = 1$. For $u > 1$, the diameter increases as a power of N, just as in *finite*-dimensional objects, and the nets are in fact *fractal*.

For $u > 1$, the change of mass upon the rescaling of length by a factor b is

$$N(bL) = b^{d_B} N(L) , \qquad (17)$$

where d_B is the fractal dimension [8]. In this case, $N(uL) = (u + v)N(L)$, so

$$d_B = \frac{\ln(u + v)}{\ln u} , \quad u > 1 . \qquad (18)$$

Transfinite Fractals Small world nets, such as $(1, v)$-nets, are *infinite*-dimensional. Indeed, their mass (N, or M) increases faster than any power (dimension) of their diameter. Also, note that a naive application of (4) to $u \to 1$ yields $d_f \to \infty$. In the case of $(1, v)$-nets one can use their weak self-similarity to define a new measure of dimensionality, $\tilde{d}_f$, characterizing how mass scales with diameter:

$$N(L + \ell) = e^{\ell \tilde{d}_f} N(L) . \qquad (19)$$

Instead of a multiplicative rescaling of length, $L \mapsto bL$, a slower additive mapping, $L \mapsto L + \ell$, that reflects the small world property is considered. Because the exponent $\tilde{d}_f$ usefully distinguishes between different graphs of infinite dimensionality, $\tilde{d}_f$ has been termed the *transfinite* fractal dimension of the network. Accordingly, objects that are self-similar and have infinite dimension (but finite transfinite dimension), such as the $(1, v)$-nets, are termed transfinite fractals, or *transfractals*, for short.

For $(1, v)$-nets, we see that upon 'zooming in' one generation level the mass increases by a factor of $w = 1 + v$, while the diameter grows from L to $L + v - 1$ (for flowers), or to $L + v$ (trees). Hence their transfractal dimension is

$$\tilde{d}_f = \begin{cases} \frac{\ln(1+v)}{v} & (1, v)\text{-trees} , \\ \frac{\ln(1+v)}{v-1} & (1, v)\text{-flowers} . \end{cases} \qquad (20)$$

There is some arbitrariness in the selection of e as the base of the exponential in the definition (19). However the base is inconsequential for the sake of comparison between dimensionalities of different objects. Also, *scaling relations* between various transfinite exponents hold, irrespective of the choice of base: consider the scaling relation

of Eq. (10) valid for fractal scale-free nets of degree exponent γ [62,63]. For example, in the fractal (u, v)-nets (with $u > 1$) renormalization reduces lengths by a factor $b = u$ and all degrees are reduced by a factor of 2, so $b^{d_k} = 2$. Thus $d_k = \ln 2 / \ln u$, and since $d_B = \ln(u + v)/\ln u$ and $\gamma = 1 + \ln(u + v)/\ln 2$, as discussed above, the relation (10) is indeed satisfied.

For transfractals, renormalization reduces distances by an *additive* length, ℓ, and we express the self-similarity manifest in the degree distribution as

$$P'(k) = e^{\ell \tilde{d}_k} P(e^{-\ell \tilde{d}_k} k) , \qquad (21)$$

where $\tilde{d}_k$ is the transfinite exponent analogous to d_k. Renormalization of the transfractal $(1, v)$-nets reduces the link lengths by $\ell = v - 1$ (for flowers), or $\ell = v$ (trees), while all degrees are halved. Thus,

$$\tilde{d}_k = \begin{cases} \frac{\ln 2}{v} & (1, v)\text{-trees} , \\ \frac{\ln 2}{v-1} & (1, v)\text{-flowers} . \end{cases}$$

Along with (20), this result confirms that the scaling relation

$$\gamma = 1 + \frac{\tilde{d}_f}{\tilde{d}_k} \qquad (22)$$

is valid also for transfractals, and regardless of the choice of base. A general proof of this relation is practically identical to the proof of (10) [62], merely replacing fractal with transfractal scaling throughout the argument.

For scale-free transfractals, following $m = L/\ell$ renormalizations the diameter and mass reduce to order one, and the scaling (19) implies $L \sim m\ell$, $N \sim e^{m\ell \tilde{d}_f}$, so that

$$L \sim \frac{1}{\tilde{d}_f} \ln N ,$$

in accordance with their small world property. At the same time the scaling (21) implies $K \sim e^{m\ell \tilde{d}_k}$, or $K \sim N^{\tilde{d}_k / \tilde{d}_f}$. Using the scaling relation (22), we rederive $K \sim N^{1/(\gamma-1)}$, which is indeed valid for scale-free nets *in general*, be they fractal or transfractal.

Properties of Fractal and Transfractal Networks

The existence of fractality in complex networks immediately calls for the question of what is the importance of such a structure in terms of network properties. In general, most of the relevant applications seem to be modified to a larger or lesser extent, so that fractal networks can be considered to form a separate network sub-class, sharing the main properties resulting from the wide distribution of regular scale-free networks, but at the same time bearing novel properties. Moreover, from a practical point of

view a fractal network can be usually more amenable to analytic treatment.

In this section we summarize some of the applications that seem to distinguish fractal from non-fractal networks.

Modularity

Modularity is a property closely related to fractality. Although this term does not have a unique well-defined definition we can claim that modularity refers to the existence of areas in the network where groups of nodes share some common characteristics, such as preferentially connecting within this area (the 'module') rather than to the rest of the network. The isolation of modules into distinct areas is a complicated task and in most cases there are many possible ways (and algorithms) to partition a network into modules.

Although networks with significant degree of modularity are not necessarily fractals, practically all fractal networks are highly modular in structure. Modularity naturally emerges from the effective 'repulsion' between hubs.

Since the hubs are not directly connected to each other, they usually dominate their neighborhood and can be considered as the 'center of mass' for a given module. The nodes surrounding hubs are usually assigned to this module.

The renormalization property of self-similar networks is very useful for estimating how modular a given network is, and especially for how this property is modified under varying scales of observation. We can use a simple definition for modularity M, based on the idea that the number of links connecting nodes within a module, L_i^{in}, is higher than the number of link connecting nodes in different modules, L_i^{out}.

For this purpose, the boxes that result from the box-covering method at a given length-scale ℓ_B are identified as the network modules for this scale. This partitioning assumes that the minimization of the number of boxes corresponds to an increase of modularity, taking advantage of the idea that all nodes within a box can reach each other within less than ℓ_B steps. This constraint tends to assign the largest possible number of nodes in a given neighborhood within the same box, resulting in an optimized modularity function.

A definition of the modularity function M that takes advantage of the special features of the renormalization process is, thus, the following [32]:

$$M(\ell_B) = \frac{1}{N_B} \sum_{i=1}^{N_B} \frac{L_i^{in}}{L_i^{out}} , \tag{23}$$

where the sum is over all the boxes.

The value of M through Eq. (23) for a given ℓ_B value is of small usefulness on its own, though. We can gather more information on the network structure if we measure M for different values of ℓ_B. If the dependence of M on ℓ_B has the form of a power-law, as if often the case in practice, then we can define the modularity exponent d_M through

$$M(\ell_B) \sim \ell_B^{d_M} . \tag{24}$$

The exponent d_M carries the important information of how modularity scales with the length, and separates modular from non-modular networks. The value of d_M is easy to compute in a d-dimensional lattice, since the number of links within any module scales with its bulk, as $L_i^{in} \sim \ell_B^d$ and the number of links outside the module scale with the length of its interface, i. e. $L_i^{out} \sim \ell_B^{d-1}$. So, the resulting scaling is $M \sim \ell_B$ i. e. $d_M = 1$. This is also the borderline value that separates non-modular structures ($d_M < 1$) from modular ones ($d_M > 1$).

For the Song–Havlin–Makse fractal model introduced in the previous section, a module can be identified as the neighborhood around a central hub. In the simplest version with $x = 1$, the network is a tree, with well-defined modules. Larger values of x mean that a larger number of links are connecting different modules, creating more loops and 'blurring' the discreteness of the modules, so that we can vary the degree of modularity in the network. For this model, it is also possible to analytically calculate the value of the exponent d_M.

During the growth process at step t, the diameter in the network model increases multiplicatively as $L(t + 1) = 3L(t)$. The number of links within a module grows with $2m + x$ (each node on the side of one link gives rise to m new links and x extra links connect the new nodes), while the number of links pointing out of a module is by definition proportional to x. Thus, the modularity $M(\ell_B)$ of a network is proportional to $(2m + x)/x$. Equation (24) can then be used to calculate d_M for the model:

$$\frac{2m + x}{x} \sim 3^{d_M} , \tag{25}$$

which finally yields

$$d_M = \frac{\ln \left(2 \frac{m}{x} + 1\right)}{\ln 3} . \tag{26}$$

So, in this model the important quantity that determines the degree of modularity in the system is the ratio of the growth parameters m/x.

Most of the real-life networks that have been measured display some sort of modular character, i. e. $d_M > 1$, although many of them have values very close to 1. Only in

a few cases we have observed exponents $d_M < 1$. Most interesting is, though, the case of d_M values much larger than 1, where a large degree of modularity is observed and this trend is more pronounced for larger length-scales.

The importance of modularity as described above can be demonstrated in biological networks. There, it has been suggested that the boxes may correspond to functional modules and in protein interaction networks, for example, there may be an evolution drive of the system behind the development of its modular structure.

Robustness

Shortly after the discovery of the scale-free property, the first important application of their structure was perhaps their extreme resilience to removal of random nodes [3,10,19,21,57,60,61]. At the same time such a network was found to be quite vulnerable to an intentional attack, where nodes are removed according to decreasing order of their degree [20,30]. The resilience of a network is usually quantified through the size of the largest remaining connected cluster $S_{max}(p)$, when a fraction p of the nodes has been removed according to a given strategy. At a critical point p_c where this size becomes equal to $S_{max}(p_c) \simeq 0$, we consider that the network has been completely disintegrated. For the random removal case, this threshold is $p_c \simeq 1$, i. e. practically all nodes need to be destroyed. In striking contrast, for intentional attacks p_c is in general of the order of only a few percent, although the exact value depends on the system details.

Fractality in networks considerably strengthens the robustness against intentional attacks, compared to non-fractal networks with the same degree exponent γ. In Fig. 9 the comparison between two such networks clearly shows that the critical fraction p_c increases almost 4 times from $p_c \simeq 0.02$ (non-fractal topology) to $p_c \simeq 0.09$ (fractal topology). These networks have the same γ exponent, the same number of links, number of nodes, number of loops and the same clustering coefficient, differing only in whether hubs are directly connected to each other. The fractal property, thus, provides a way of increasing resistance against the network collapse, in the case of a targeted attack.

The main reason behind this behavior is the dispersion of hubs in the network. A hub is usually a central node that helps other nodes to connect to the main body of the system. When the hubs are directly connected to each other, this central core is easy to destroy in a targeted attack leading to a rapid collapse of the network. On the contrary, isolating the hubs into different areas helps the network to retain connectivity for longer time, since destroying the hubs
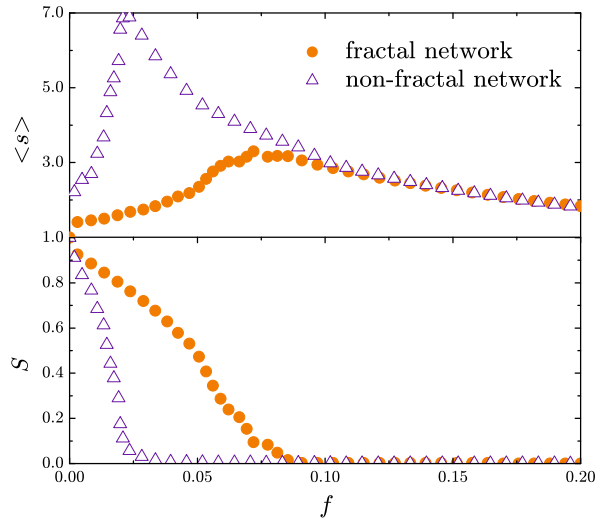
Fractal and Transfractal Scale-Free Networks, Figure 9
Vulnerability under intentional attack of a non-fractal Song–Makse–Havlin network (for $e = 0$) and a fractal Song–Makse–Havlin network (for $e = 1$). The plot shows the relative size of the largest cluster, S, and the average size of the remaining isolated clusters, $\langle s \rangle$ as a function of the removal fraction f of the largest hubs for both networks

now is not similarly catastrophic, with most of the nodes finding alternative paths through other connections.

The advantage of increased robustness derived from the combination of modular and fractal network character, may provide valuable hints on why most biological networks have evolved towards a fractal architecture (better chance of survival against lethal attacks).

Degree Correlations

We have already mentioned the importance of hub-hub correlations or anti-correlations in fractality. Generalizing this idea to nodes of any degree, we can ask what is the joint degree probability $P(k_1, k_2)$ that a randomly chosen link connects two nodes with degree k_1 and k_2, respectively. Obviously, this is a meaningful question only for networks with a wide degree distribution, otherwise the answer is more or less trivial with all nodes having similar degrees. A similar and perhaps more useful quantity is the conditional degree probability $P(k_1|k_2)$, defined as the probability that a random link from a node having degree k_2 points to a node with degree k_1. In general, the following balance condition is satisfied

$$k_2 P(k_1|k_2)P(k_2) = k_1 P(k_2|k_1)P(k_1) . \qquad (27)$$

It is quite straightforward to calculate $P(k_1|k_2)$ for completely uncorrelated networks. In this case, $P(k_1|k_2)$ does

not depend on k_2, and the probability to chose a node with degree k_1 becomes simply $P(k_1|k_2) = k_1 P(k_1)/\langle k_1 \rangle$. In the case where degree-degree correlations are present, though, the calculation of this function is very difficult, even when restricting ourselves to a direct numerical evaluation, due to the emergence of huge fluctuations.

We can still estimate this function, though, using again the self-similarity principle. If we consider that the function $P(k_1, k_2)$ remains invariant under the network renormalization scheme described above, then it is possible to show that [33]

$$P(k_1, k_2) \sim k_1^{-(\gamma-1)} k_2^{-\epsilon} \quad (k_1 > k_2), \tag{28}$$

and similarly

$$P(k_1|k_2) \sim k_1^{-(\gamma-1)} k_2^{-(\epsilon-\gamma+1)} \quad (k_1 > k_2), \tag{29}$$

In the above equations we have also introduced the correlation exponent ϵ, which characterizes the degree of correlations in a network. For example, the case of uncorrelated networks is described by the value $\epsilon = \gamma - 1$.

The exponent ϵ can be measured quite accurately using an appropriate quantity. For this purpose, we can introduce a measure such as

$$E_b(k) \equiv \frac{\int_{bk}^{\infty} P(k|k_2)\mathrm{d}k_2}{\int_{bk}^{\infty} P(k)\mathrm{d}k}, \tag{30}$$

which estimates the probability that a node with degree k has neighbors with degree larger than bk, and b is an arbitrary parameter that has been shown not to influence the results. It is easy to show that

$$E_b(k) \sim \frac{k^{1-\epsilon}}{k^{1-\gamma}} = k^{-(\epsilon-\gamma)}. \tag{31}$$

This relation allows us to estimate ϵ for a given network, after calculating the quantity $E_b(k)$ as a function of k.

The above discussion can be equally applied to both fractal and non-fractal networks. If we restrict ourselves to fractal networks, then we can develop our theory a bit further. If we consider the probability $\mathcal{E}(\ell_B)$ that the largest degree node in each box is connected directly with the other largest degree nodes in other boxes (after optimally covering the network), then this quantity scales as a power law with ℓ_B:

$$\mathcal{E}(\ell_B) \sim \ell_B^{-d_e}, \tag{32}$$

where d_e is a new exponent describing the probability of hub-hub connection [63]. The exponent ϵ, which describes correlations over any degree, is related to d_e, which

refers to correlations between hubs only. The resulting relation is

$$\epsilon = 2 + \frac{d_e}{d_k} = 2 + (\gamma - 1)\frac{d_e}{d_B}. \tag{33}$$

For an infinite fractal dimension $d_B \to \infty$, which is the onset of non-fractal networks that cannot be described by the above arguments, we have the limiting case of $\epsilon = 2$. This value separates fractal from non-fractal networks, so that fractality is indicated by $\epsilon > 2$. Also, we have seen that the line $\epsilon = \gamma - 1$ describes networks for which correlations are minimal. Measurements of many real-life networks have verified the above statements, where networks with $\epsilon > 2$ having been clearly characterized as fractals with alternate methods. All non-fractal networks have values of $\epsilon < 2$ and the distance from the $\epsilon = \gamma - 1$ line determines how much stronger or weaker the correlations are, compared to the uncorrelated case.

In short, using the self-similarity principle makes it possible to gain a lot of insight on network correlations, a notoriously difficult task otherwise. Furthermore, the study of correlations can be reduced to the calculation of a single exponent ϵ, which is though capable of delivering a wealth of information on the network topological properties.

Diffusion and Resistance

Scale-free networks have been described as objects of infinite dimensionality. For a regular structure this statement would suggest that one can simply use the known diffusion laws for $d = \infty$. Diffusion on scale-free structures, however, is much harder to study, mainly due to the lack of translational symmetry in the system and different local environments. Although exact results are still not available, the scaling theory on fractal networks provides the tools to better understand processes, such as diffusion and electric resistance.

In the following, we describe diffusion through the average first-passage time T_{AB}, which is the average time for a diffusing particle to travel from node A to node B. At the same time, assuming that each link in the network has an electrical resistance of 1 unit, we can describe the electrical properties through the resistance between the two nodes A and B, R_{AB}.

The connection between diffusion (first-passage time) and electric networks has long been established in homogeneous systems. This connection is usually expressed through the Einstein relation [8]. The Einstein relation is of great importance because it connects a *static quantity* R_{AB} with a *dynamic quantity* T_{AB}. In other words, the be-

havior of a diffusing particle can be inferred by simply having knowledge of a static topological property of the network.

In any renormalizable network the scaling of T and R follows the form:

$$\frac{T'}{T} = \ell_{\mathrm{B}}^{-d_w} , \quad \frac{R'}{R} = \ell_{\mathrm{B}}^{-\zeta} , \tag{34}$$

where $T'(R')$ and $T(R)$ are the first-passage time (resistance) for the renormalized and original networks, respectively. The dynamical exponents d_w and ζ characterize the scaling in any lattice or network that remains invariant under renormalization. The Einstein relation relates these two exponents through the dimensionality of the substrate d_{B}, according to:

$$d_w = \zeta + d_{\mathrm{B}} . \tag{35}$$

The validity of this relation in inhomogeneous complex networks, however, is not yet clear. Still, in fractal and transfractal networks there are many cases where this relation has been proved to be valid, hinting towards a wider applicability. For example, in [13,56] it has been shown that the Einstein Relation [8] in (u, v)-*flowers* and (u, v)-trees is valid for any u and v, that is for both fractal and transfractal networks. In general, in terms of the scaling theory we can study diffusion and resistance (or conductance) in a similar manner [32].

Because of the highly inhomogeneous character of the structure, though, we are interested in how these quantities behave as a function of the end-node degrees k_1 and k_2 when they are separated by a given distance ℓ. Thus, we are looking for the full dependence of $T(\ell; k_1, k_2)$ and $R(\ell; k_1, k_2)$. Obviously, for lattices or networks with narrow degree distribution there is no degree dependence and those results should be a function of ℓ only.

For self-similar networks, we can rewrite Eq. (34) above as

$$\frac{T'}{T} = \left(\frac{N'}{N}\right)^{d_w/d_{\mathrm{B}}} , \quad \frac{R'}{R} = \left(\frac{N'}{N}\right)^{\zeta/d_{\mathrm{B}}} , \tag{36}$$

where we have taken into account Eq. (3). This approach offers the practical advantage that the variation of N'/N is larger than the variation of ℓ_{B}, so that the exponents calculation can be more accurate. To calculate these exponents, we fix the box size ℓ_{B} and we measure the diffusion time T and resistance R between any two points in a network before and after renormalization. If for every such pair we plot the corresponding times and resistances in T' vs. T and R' vs. R plots, as shown in Fig. 10, then all these points fall in a narrow area, suggesting a constant value for the ratio T'/T over the entire network. Repeating this procedure
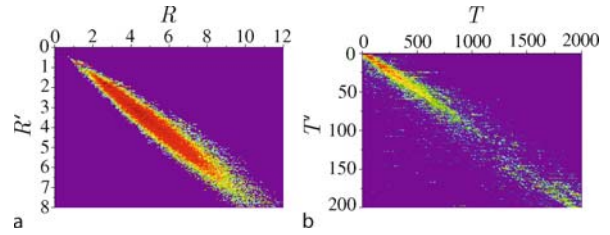
Fractal and Transfractal Scale-Free Networks, Figure 10
Typical behavior of the probability distributions for the resistance R vs. R' and the diffusion time T vs. T', respectively, for a given ℓ_{B} value. Similar plots for other ℓ_{B} values verify that the ratios of these quantities during a renormalization stage are roughly constant for all pairs of nodes in a given biological network
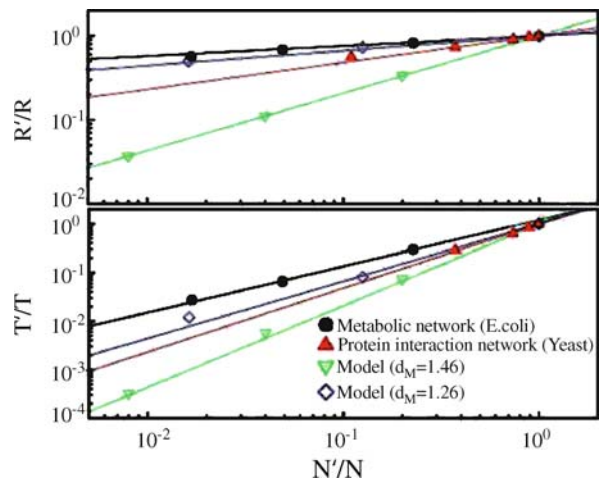
Fractal and Transfractal Scale-Free Networks, Figure 11
Average value of the ratio of resistances R/R' and diffusion times T/T', as measured for different ℓ_{B} values (each point corresponds to a different value of ℓ_{B}). Results are presented for both biological networks, and two fractal network models with different d_M values. The slopes of the curves correspond to the exponents ζ/d_{B} (top panel) and d_w/d_{B} (bottom panel)

for different ℓ_{B} values yields other ratio values. The plot of these ratios vs. N'/N (Fig. 11) finally exhibits a power-law dependence, verifying Eq. (36). We can then easily calculate the exponents d_w and ζ from the slopes in the plot, since the d_{B} exponent is already known through the standard box-covering methods. It has been shown that the results for many different networks are consistent, within statistical error, with the Einstein relation [32,56].

The dependence on the degrees k_1, k_2 and the distance ℓ can also be calculated in a scaling form using the self-similarity properties of fractal networks. After renormalization, a node with degree k in a given network, will have a degree $k' = \ell_{\mathrm{B}}^{-d_k} k$ according to Eq. (9). At the

same time all distances ℓ are scaled down according to $\ell' = \ell/\ell_B$. This means that Eqs. (36) can be written as

$$R'(\ell'; k_1', k_2') = \ell_B^{-\zeta} R(\ell; k_1, k_2) \tag{37}$$

$$T'(\ell'; k_1', k_2') = \ell_B^{-d_w} T(\ell; k_1, k_2) . \tag{38}$$

Substituting the renormalized quantities we get:

$$R'(\ell_B^{-1}\ell; \ell_B^{-d_k} k_1, \ell_B^{-d_k} k_2) = \ell_B^{-\zeta} R(\ell; k_1, k_2) . \tag{39}$$

The above equation holds for all values of ℓ_B, so we can select this quantity to be $\ell_B = k_2^{1/d_k}$. This constraint allows us to reduce the number of variables in the equation, with the final result:

$$R\left(\frac{\ell}{k_2^{\frac{1}{d_k}}}; \frac{k_1}{k_2}, 1\right) = k_2^{\frac{-\zeta}{d_k}} R(\ell_B; k_1, k_2) . \tag{40}$$

This equation suggests a scaling for the resistance R:

$$R(\ell; k_1, k_2) = k_2^{\frac{\zeta}{d_k}} f_R\left(\frac{\ell}{k_2^{\frac{1}{d_k}}}, \frac{k_1}{k_2}\right), \tag{41}$$

Fractal and Transfractal Scale-Free Networks, Figure 12
Rescaling of a the resistance and b the diffusion time according to Eqs. (41) and (42) for the protein interaction network of yeast (upper symbols) and the Song–Havlin–Makse model for $e = 1$ (lower filled symbols). The data for PIN have been vertically shifted upwards by one decade for clarity. Each symbol corresponds to a fixed ratio k_1/k_2 and the different colors denote a different value for k_1. Inset: Resistance R as a function of distance ℓ, before rescaling, for constant ratio $k_1/k_2 = 1$ and different k_1 values

where $f_R()$ is an undetermined function. All the above arguments can be repeated for the diffusion time, with a similar expression:

$$T(\ell; k_1, k_2) = k_2^{\frac{d_w}{d_k}} f_T\left(\frac{\ell}{k_2^{\frac{1}{d_k}}}, \frac{k_1}{k_2}\right), \tag{42}$$

where the form of the right-hand function may be different. The final result for the scaling form is Eqs. (41) and (42), which is also supported by the numerical data collapse in Fig. 12. Notice that in the case of homogeneous networks, where there is almost no k-dependence, the unknown functions in the rhs reduce to the forms $f_R(x, 1) = x^\zeta$, $f_T(x, 1) = x^{d_w}$, leading to the well-established classical relations $R \sim \ell^\zeta$ and $T \sim \ell^{d_w}$.

Future Directions

Fractal networks combine features met in fractal geometry and in network theory. As such, they present many unique aspects. Many of their properties have been well-studied and understood, but there is still a great amount of open and unexplored questions remaining to be studied.

Concerning the structural aspects of fractal networks, we have described that in most networks the degree distribution $P(k)$, the joint degree distribution $P(k_1, k_2)$ and a number of other quantities remain invariant under renormalization. Are there any quantities that are not invariable, and what would their importance be?

Of central importance is the relation of topological features with functionality. The optimal network covering leads to the partitioning of the network into boxes. Do these boxes carry a message other than nodes proximity? For example, the boxes could be used as an alternative definition for separated communities, and fractal methods could be used as a novel method for community detection in networks [4,5,18,51,53].

The networks that we have presented are all static, with no temporal component, and time evolution has been ignored in all our discussions above. Clearly, biological networks, the WWW, and other networks have grown (and continue to grow) from some earlier simpler state to their present fractal form. Has fractality always been there or has it emerged as an intermediate stage obeying certain evolutionary drive forces? Is fractality a stable condition or growing networks will eventually fall into a non-fractal form?

Finally, we want to know what is the inherent reason behind fractality. Of course, we have already described how hub-hub anti-correlations can give rise to fractal networks. However, can this be directly related to some un-

derlying mechanism, so that we gain some information on the process? In general, in Biology we already have some idea on the advantages of adopting a fractal structure. Still, the question remains: why fractality exists in certain networks and not in others? Why both fractal and non-fractal networks are needed? It seems that we will be able to increase our knowledge for the network evolutionary mechanisms through fractality studies.

In conclusion, a deeper understanding of the self-similarity, fractality and transfractality of complex networks will help us analyze and better understand many fundamental properties of real-world networks.

Acknowledgments

Appendix: The Box Covering Algorithms

The estimation of the fractal dimension and the self-similar features in networks have become standard properties in the study of real-world systems. For this reason, in the last three years many box covering algorithms have been proposed [64,69]. This section presents four of the main algorithms, along with a brief discussion on the advantages and disadvantages that they offer.

Recalling the original definition of box covering by Hausdorff [14,29,55], for a given network G and box size ℓ_B, a box is a set of nodes where all distances ℓ_{ij} between any two nodes i and j in the box are smaller than ℓ_B. The minimum number of boxes required to cover the entire network G is denoted by N_B. For $\ell_B = 1$, each box encloses only 1 node and therefore, N_B is equal to the size of the network N. On the other hand, $N_B = 1$ for $\ell_B \geq \ell_B^{max}$, where ℓ_B^{max} is the diameter of the network plus one.

The ultimate goal of a box-covering algorithm is to find the *minimum* number of boxes $N_B(\ell_B)$ for any ℓ_B. It has been shown that this problem belongs to the family of NP-hard problems [34], which means that the solution cannot be achieved in polynomial time. In other words, for a relatively large network size, there is no algorithm that can provide an exact solution in a reasonably short amount of time. This limitation requires treating the box covering problem with approximations, using for example optimization algorithms.

The Greedy Coloring Algorithm

The box-covering problem can be mapped into another NP-hard problem [34]: the graph coloring problem.

An algorithm that approximates well the optimal solution of this problem was presented in [64]. For an arbitrary value of ℓ_B, first construct a dual network G', in which two nodes are connected if the distance between them in G (the original network) is greater or equal than ℓ_B. Figure 13 shows an example of a network G which yields such a dual network G' for $\ell_B = 3$ (upper row of Fig. 13).

Vertex coloring is a well-known procedure, where labels (or colors) are assigned to each vertex of a network, so that no edge connects two identically colored vertices. It is clear that such a coloring in G' gives rise to a natural box covering in the original network G, in the sense that vertices of the same color will necessarily form a box since the distance between them must be less than ℓ_B. Accordingly, the minimum number of boxes $N_B(G)$ is equal to the minimum required number of colors (or the chromatic number) in the dual network G', $\chi(G')$.

In simpler terms, (a) if the distance between two nodes in G is greater than ℓ_B these two neighbors cannot belong in the same box. According to the construction of G', these two nodes will be connected in G' and thus they cannot have the same color. Since they have a different color they will not belong in the same box in G. (b) On the contrary, if the distance between two nodes in G is less than ℓ_B it is possible that these nodes belong in the same box. In G' these two nodes will not be connected and it is allowed
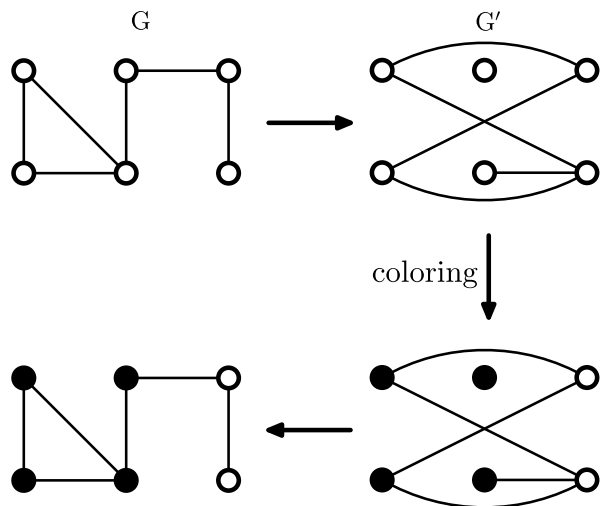
Fractal and Transfractal Scale-Free Networks, Figure 13
Illustration of the solution for the network covering problem via mapping to the graph coloring problem. Starting from *G* (*upper left panel*) we construct the dual network *G'* (*upper right panel*) for a given box size (here $\ell_B = 3$), where two nodes are connected if they are at a distance $\ell \geq \ell_B$. We use a greedy algorithm for vertex coloring in *G'*, which is then used to determine the box covering in *G*, as shown in the plot

for these two nodes to carry the same color, i. e. they may belong to the same box in G, (whether these nodes will actually be connected depends on the exact implementation of the coloring algorithm).

The algorithm that follows both constructs the dual network G' and assigns the proper node colors for all ℓ_B values in one go. For this implementation a two-dimensional matrix $c_{i\ell}$ of size $N \times \ell_B^{max}$ is needed, whose values represent the color of node i for a given box size $\ell = \ell_B$.

1. Assign a unique id from 1 to N to all network nodes, without assigning any colors yet.
2. For all ℓ_B values, assign a color value 0 to the node with id=1, i. e. $c_{1\ell} = 0$.
3. Set the id value $i = 2$. Repeat the following until $i = N$.
   (a) Calculate the distance ℓ_{ij} from i to all the nodes in the network with id j less than i.
   (b) Set $\ell_B = 1$
   (c) Select one of the unused colors $c_{j\ell_{ij}}$ from all nodes $j < i$ for which $\ell_{ij} \geq \ell_B$. This is the color $c_{i\ell_B}$ of node i for the given ℓ_B value.
   (d) Increase ℓ_B by one and repeat (c) until $\ell_B = \ell_B^{max}$.
   (e) Increase i by 1.

The results of the greedy algorithm may depend on the original coloring sequence. The quality of this algorithm was investigated by randomly reshuffling the coloring sequence and applying the greedy algorithm several times and in different models [64]. The result was that the probability distribution of the number of boxes N_B (for all box sizes ℓ_B) is a narrow Gaussian distribution, which indicates that almost any implementation of the algorithm yields a solution close to the optimal.

Strictly speaking, the calculation of the fractal dimension d_B through the relation $N_B \sim \ell_B^{-d_B}$ is valid only for the minimum possible value of N_B, for any given ℓ_B value, so any box covering algorithm must aim to find this minimum N_B. Although there is no rule to determine when this minimum value has been actually reached (since this would require an exact solution of the NP-hard coloring problem) it has been shown [23] that the greedy coloring algorithm can, in many cases, identify a coloring sequence which yields the optimal solution.

Burning Algorithms

This section presents three box covering algorithms based on more traditional breadth-first search algorithm.

A box is defined as *compact* when it includes the maximum possible number of nodes, i. e. when there do not exist any other network nodes that could be included in this box. A *connected* box means that any node in the box
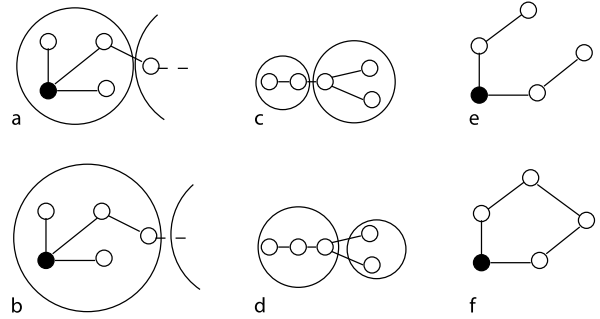
Fractal and Transfractal Scale-Free Networks, Figure 14
Our definitions for a box that is **a** non-compact for $\ell_B = 3$, i. e. could include more nodes, **b** compact, **c** connected, and **d** disconnected (the nodes in the *right box* are not connected in the box). **e** For this box, the values $\ell_B = 5$ and $r_B = 2$ verify the relation $\ell_B = 2r_B + 1$. **f** One of the pathological cases where this relation is not valid, since $\ell_B = 3$ and $r_B = 2$

can be reached from any other node in this box, without having to leave this box. Equivalently, a *disconnected* box denotes a box where certain nodes can be reached by other nodes in the box only by visiting nodes outside this box. For a demonstration of these definitions see Fig. 14.

Burning with the Diameter ℓ_B, and the Compact-Box-Burning (CBB) Algorithm

The basic idea of the CBB algorithm for the generation of a box is to start from a given box center and then expand the box so that it includes the maximum possible number of nodes, satisfying at the same time the maximum distance between nodes in the box ℓ_B. The CBB algorithm is as follows (see Fig. 15):

1. Initially, mark all nodes as uncovered.
2. Construct the set C of all yet uncovered nodes.
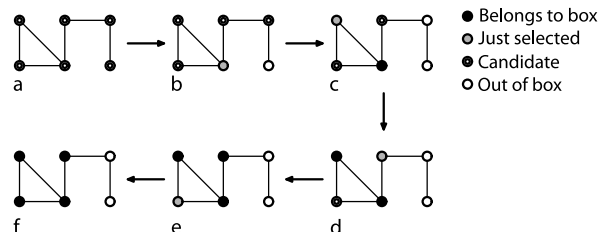
Fractal and Transfractal Scale-Free Networks, Figure 15
Illustration of the CBB algorithm for $\ell_B = 3$. **a** Initially, all nodes are candidates for the box. **b** A random node is chosen, and nodes at a distance further than ℓ_B from this node are no longer candidates. **c** The node chosen in b becomes part of the box and another candidate node is chosen. The above process is then repeated until the box is complete

3. Choose a random node p from the set of uncovered nodes C and remove it from C.
4. Remove from C all nodes i whose distance from p is $\ell_{pi} \geq \ell_B$, since by definition they will not belong in the same box.
5. Repeat steps (3) and (4) until the candidate set is empty.
6. Repeat from step (2) until all the network has been covered.
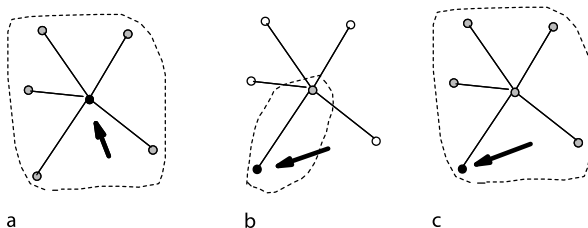
Random Box Burning

In 2006, J. S. Kim et al. presented a simple algorithm for the calculation of fractal dimension in networks [42,43,44]:

1. Pick a randomly chosen node in the network as a seed of the box.
2. Search using breath-first search algorithm until distance l_B from the seed. Assign all newly burned nodes to the new box. If no new node is found, discard and start from (1) again.
3. Repeat (1) and (2) until all nodes have a box assigned.

This Random Box Burning algorithm has the advantage of being a fast and simple method. However, at the same time there is no inherent optimization employed during the network coverage. Thus, this simple Monte-Carlo method is almost certain that will yield a solution far from the optimal and one needs to implement many different realizations and only retain the smallest number of boxes found out of all these realizations.

Burning with the Radius r_B, and the Maximum-Excluded-Mass-Burning (MEMB) Algorithm

A box of size ℓ_B includes nodes where the distance between any pair of nodes is less than ℓ_B. It is possible, though, to grow a box from a given central node, so that all nodes in the box are within distance less than a given box radius r_B (the maximum distance from a central node). This way, one can still recover the same fractal properties of a network. For the original definition of the box, ℓ_B corresponds to the box diameter (maximum distance between any two nodes in the box) plus one. Thus, ℓ_B and r_B are connected through the simple relation $\ell_B = 2r_B + 1$. In general this relation is exact for loopless configurations, but in general there may exist cases where this equation is not exact (Fig. 14).

The MEMB algorithm always yields the optimal solution for non scale-free homogeneous networks, since the choice of the central node is not important. However, in inhomogeneous networks with wide-tailed degree distribution, such as scale-free networks, this algorithm fails to achieve an optimal solution because of the presence of hubs.

The MEMB, as a difference from the Random Box Burning and the CBB, attempts to locate some *optimal central* nodes which act as the burning origins for the boxes. It contains as a special case the choice of the hubs as centers of the boxes, but it also allows for low-degree nodes to be burning centers, which sometimes is convenient for finding a solution closer to the optimal.

In the following algorithm we use the basic idea of box optimization, in which each box covers the maximum possible number of nodes. For a given burning radius r_B, we define the *excluded mass* of a node as the number of uncovered nodes within a chemical distance less than r_B. First, calculate the excluded mass for all the uncovered nodes. Then, seek to cover the network with boxes of maximum excluded mass. The details of this algorithm are as follows (see Fig. 17):

1. Initially, all nodes are marked as uncovered and non-centers.
2. For all non-center nodes (including the already covered nodes) calculate the excluded mass, and select the node p with the maximum excluded mass as the next center.
3. Mark all the nodes with chemical distance less than r_B from p as covered.
4. Repeat steps (2) and (3) until all nodes are either covered or centers.

Notice that the excluded mass has to be updated in each step because it is possible that it has been modified during this step. A box center can also be an already covered node, since it may lead to a larger box mass. After the above procedure, the number of selected centers coincides with the number of boxes N_B that completely cover the network.

a   b   c

Fractal and Transfractal Scale-Free Networks, Figure 16
Burning with the radius r_B from **a** a hub node or **b** a non-hub node results in very different network coverage. In **a** we need just one box of $r_B = 1$ while in **b** 5 boxes are needed to cover the same network. This is an intrinsic problem when burning with the radius. **c** Burning with the maximum distance ℓ_B (in this case $\ell_B = 2r_B + 1 = 3$) we avoid this situation, since independently of the starting point we would still obtain $N_B = 1$

Fractal and Transfractal Scale-Free Networks, Figure 17
Illustration of the MEMB algorithm for $r_B = 1$. *Upper row*: Calculation of the box centers **a** We calculate the excluded mass for each node. **b** The node with maximum mass becomes a center and the excluded masses are recalculated. **c** A new center is chosen. Now, the entire network is covered with these two centers. *Bottom row*: Calculation of the boxes **d** Each box includes initially only the center. Starting from the centers we calculate the distance of each network node to the closest center. **e** We assign each node to its nearest box

Fractal and Transfractal Scale-Free Networks, Figure 18
Comparison of the distribution of N_B for 10^4 realizations of the four network covering methods presented in this paper. Notice that three of these methods yield very similar results with narrow distributions and comparable minimum values, while the random burning algorithm fails to reach a value close to this minimum (and yields a broad distribution)

However, the non-center nodes have not yet been assigned to a given box. This is performed in the next step:

1. Give a unique box id to every center node.
2. For all nodes calculate the "central distance", which is the chemical distance to its nearest center. The central distance has to be less than r_B, and the center identification algorithm above guarantees that there will always exist such a center. Obviously, all center nodes have a central distance equal to 0.
3. Sort the non-center nodes in a list according to increasing central distance.
4. For each non-center node i, at least one of its neighbors has a central distance less than its own. Assign to i the same id with this neighbor. If there exist several such neighbors, randomly select an id from these neighbors. Remove i from the list.
5. Repeat step (4) according to the sequence from the list in step (3) for all non-center nodes.

Comparison Between Algorithms

The choice of the algorithm to be used for a problem depends on the details of the problem itself. If connected boxes are a requirement, MEMB is the most appropriate algorithm; but if one is only interested in obtaining the fractal dimension of a network, the greedy-coloring or the random box burning are more suitable since they are the fastest algorithms.

As explained previously, any algorithm should intend to find the optimal solution, that is, find the minimum number of boxes that cover the network. Figure 18 shows the performance of each algorithm. The greedy-coloring, the CBB and MEMB algorithms exhibit a narrow distribution of the number of boxes, showing evidence that they cover the network with a number of boxes that is close to the optimal solution. Instead, the Random Box Burning returns a wider distribution and its average is far above the average of the other algorithms. Because of the great ease and speed with which this technique can be implemented, it would be useful to show that the average number of covering boxes is overestimated by a fixed proportionality constant. In that case, despite the error, the predicted number of boxes would still yield the correct scaling and fractal dimension.

Bibliography

1. Albert R, Barabási A-L (2002) Rev Mod Phys 74:47; Barabási A-L (2003) Linked: how everything is connected to everything else and what it means. Plume, New York; Newman MEJ (2003) SIAM Rev 45:167; Dorogovtsev SN, Mendes JFF (2002) Adv Phys 51:1079; Dorogovtsev SN, Mendes JFF (2003) Evolution of networks: from biological nets to the internet and WWW. Oxford University Press, Oxford; Bornholdt S, Schuster HG (2003) Handbook of graphs and networks. Wiley-VCH, Berlin; Pastor-Satorras R, Vespignani A (2004) Evolution and structure of the internet. Cambridge University Press, Cambridge; Amaral LAN, Ottino JM (2004) Complex networks – augmenting the framework for the study of complex systems. Eur Phys J B 38: 147–162
2. Albert R, Jeong H, Barabási A-L (1999) Diameter of the world wide web. Nature 401:130–131

3. Albert R, Jeong H, Barabási AL (2000) Nature 406:p378
4. Bagrow JP, Bollt EM (2005) Phys Rev E 72:046108
5. Bagrow JP (2008) Stat Mech P05001
6. Bagrow JP, Bollt EM, Skufca JD (2008) Europhys Lett 81:68004
7. Barabási A-L, Albert R (1999) Sience 286:509
8. ben-Avraham D, Havlin S (2000) Diffusion and reactions in fractals and disordered systems. Cambridge University Press, Cambridge
9. Berker AN, Ostlund S (1979) J Phys C 12:4961
10. Beygelzimer A, Grinstein G, Linsker R, Rish I (2005) Physica A Stat Mech Appl 357:593–612
11. Binney JJ, Dowrick NJ, Fisher AJ, Newman MEJ (1992) The theory of critical phenomena: an introduction to the renormalization group. Oxford University Press, Oxford
12. Bollobás B (1985) Random graphs. Academic Press, London
13. Bollt E, ben-Avraham D (2005) New J Phys 7:26
14. Bunde A, Havlin S (1996) Percolation I and Percolation II. In: Bunde A, Havlin S (eds) Fractals and disordered systems, 2nd edn. Springer, Heidelberg
15. Burch H, Chewick W (1999) Mapping the internet. IEEE Comput 32:97–98
16. Butler D (2006) Nature 444:528
17. Cardy J (1996) Scaling and renormalization in statistical physics. Cambridge University Press, Cambridge
18. Clauset A, Newman MEJ, Moore C (2004) Phys Rev E 70:066111
19. Cohen R, Erez K, ben-Avraham D, Havlin S (2000) Phys Rev Lett 85:4626
20. Cohen R, Erez K, ben-Avraham D, Havlin S (2001) Phys Rev Lett 86:3682
21. Cohen R, ben-Avraham D, Havlin S (2002) Phys Rev E 66:036113
22. Comellas F Complex networks: deterministic models physics and theoretical computer science. In: Gazeau J-P, Nesetril J, Rovan B (eds) From Numbers and Languages to (Quantum) Cryptography. 7 NATO Security through Science Series: Information and Communication Security. IOS Press, Amsterdam. pp 275–293. 348 pags. ISBN 1-58603-706-4
23. Cormen TH, Leiserson CE, Rivest RL, Stein C (2001) Introduction to algorithms. MIT Press, Cambridge
24. Data from SCAN project. The Mbone. http://www.isi.edu/scan/scan.html Accessed 2000
25. Database of Interacting Proteins (DIP) http://dip.doe-mbi.ucla.edu Accessed 2008
26. Dorogovtsev SN, Goltsev AV, Mendes JFF (2002) Phys Rev E 65:066122
27. Erdős P, Rényi A (1960) On the evolution of random graphs. Publ Math Inst Hung Acad Sci 5:17–61
28. Faloutsos M, Faloutsos P, Faloutsos C (1999) Comput Commun Rev 29:251–262
29. Feder J (1988) Fractals. Plenum Press, New York
30. Gallos LK, Argyrakis P, Bunde A, Cohen R, Havlin S (2004) Physica A 344:504–509
31. Gallos LK, Cohen R, Argyrakis P, Bunde A, Havlin S (2005) Phys Rev Lett 94:188701
32. Gallos LK, Song C, Havlin S, Makse HA (2007) PNAS 104:7746
33. Gallos LK, Song C, Makse HA (2008) Phys Rev Lett 100:248701
34. Garey M, Johnson D (1979) Computers and intractability: a guide to the theory of NP-completeness. W.H. Freeman, New York
35. Goh K-I, Salvi G, Kahng B, Kim D (2006) Phys Rev Lett 96:018701
36. Han J-DJ et al (2004) Nature 430:88–93
37. Hinczewski M, Berker AN (2006) Phys Rev E 73:066126

38. Jeong H, Tombor B, Albert R, Oltvai ZN, Barabási A-L (2000) Nature 407:651–654
39. Kadanoff LP (2000) Statistical physics: statics, dynamics and renormalization. World Scientific Publishing Company, Singapore
40. Kaufman M, Griffiths RB (1981) Phys Rev B 24:496(R)
41. Kaufman M, Griffiths RB (1984) Phys Rev B 24:244
42. Kim JS, Goh K-I, Salvi G, Oh E, Kahng B, Kim D (2007) Phys Rev E 75:016110
43. Kim JS, Goh K-I, Kahng B, Kim D (2007) Chaos 17:026116
44. Kim JS, Goh K-I, Kahng B, Kim D (2007) New J Phys 9:177
45. Mandelbrot B (1982) The fractal geometry of nature. W.H. Freeman and Company, New York
46. Maslov S, Sneppen K (2002) Science 296:910–913
47. Milgram S (1967) Psychol Today 2:60
48. Motter AE, de Moura APS, Lai Y-C, Dasgupta P (2002) Phys Rev E 65:065102
49. Newman MEJ (2002) Phys Rev Lett 89:208701
50. Newman MEJ (2003) Phys Rev E 67:026126
51. Newman MEJ, Girvan M (2004) Phys Rev E 69:026113
52. Overbeek R et al (2000) Nucl Acid Res 28:123–125
53. Palla G, Barabási A-L, Vicsek T (2007) Nature 446:664–667
54. Pastor-Satorras R, Vázquez A, Vespignani A (2001) Phys Rev Lett 87:258701
55. Peitgen HO, Jurgens H, Saupe D (1993) Chaos and fractals: new frontiers of science. Springer, New York
56. Rozenfeld H, Havlin S, ben-Avraham D (2007) New J Phys 9:175
57. Rozenfeld H, ben-Avraham D (2007) Phys Rev E 75:061102
58. Salmhofer M (1999) Renormalization: an introduction. Springer, Berlin
59. Schwartz N, Cohen R, ben-Avraham D, Barabasi A-L, Havlin S (2002) Phys Rev E 66:015104
60. Serrano MA, Boguna M (2006) Phys Rev Lett 97:088701
61. Serrano MA, Boguna M (2006) Phys Rev E 74:056115
62. Song C, Havlin S, Makse HA (2005) Nature 433:392
63. Song C, Havlin S, Makse HA (2006) Nature Phys 2:275
64. Song C, Gallos LK, Havlin S, Makse HA (2007) J Stat Mech P03006
65. Stanley HE (1971) Introduction to phase transitions and critical phenomena. Oxford University Press, Oxford
66. Vicsek T (1992) Fractal growth phenomena, 2nd edn. World Scientific, Singapore Part IV
67. Watts DJ, Strogatz SH (1998) Collective dynamics of "small-world" networks. Nature 393:440–442
68. Xenarios I et al (2000) Nucl Acids Res 28:289–291
69. Zhoua W-X, Jianga Z-Q, Sornette D (2007) Physica A 375:741–752

Freeway Traffic Management and Control

A. HEGYI[1], T. BELLEMANS[2], B. DE SCHUTTER[1]
[1] TU Delft, Delft, The Netherlands
[2] Hasselt University, Diepenbeek, Belgium

Article Outline

Glossary

Glossary

Feedforward control The block diagram of a feedforward control structure is shown in Fig. 1 [4]. The behavior of process **P** can be influenced by the control inputs. As a result the outputs (measurements or observations) show a given behavior. The controller **C** determines the control inputs in order to reach a given desired behavior of the outputs, taking into account the disturbances that act on the process. In the feedforward structure the controller **C** translates the desired behavior and the measured disturbances into control actions for the process.

The term feedforward refers to the fact that the direction of the information flow in the system contains no loops, i. e., it propagates only "forward".

The main advantages of a feedforward controller are that the complete system is stable if the controller and the process are stable, and that its design is in general simple.

Feedback control In Fig. 2 the feedback control structure is shown [4]. In contrast to the feedforward control structure, here the behavior of the outputs is coupled back to the controller (hence the name feedback). This structure is also often referred to as "closed-loop" control.

The main advantages of a feedback controller over a feedforward controller are that (1) it may have a quicker response (resulting in better performance), (2) it may correct undesired offsets in the output, (3) it may suppress unmeasurable disturbances that are ob-
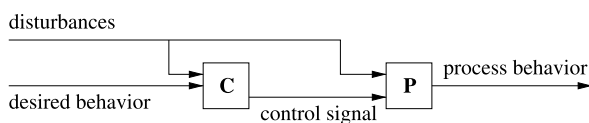
Freeway Traffic Management and Control, Figure 1
The feedforward control structure
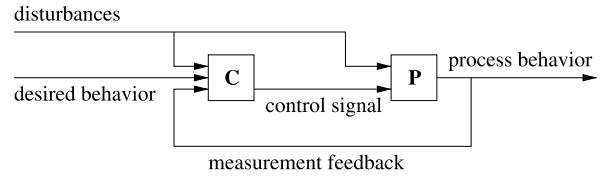
Freeway Traffic Management and Control, Figure 2
The feedback control structure

servable through the output only, and (4) it may stabilize an unstable system.

Optimal control Optimal control is a control methodology that formulates a control problem in terms of a *performance function*, also called an *objective function* [73]. This function expresses the performance of the system over a given period of time, and the goal of the controller is to find the control signals that result in optimal performance. Depending on the mathematical description of the control problem there exist several methods for the optimization of the control input including analytic and numerical approaches. Optimal control can be considered as a feedforward control approach.

Model predictive control Model predictive control (MPC) is an extension of the optimal control framework [15,79]. In Fig. 3 the block diagram of MPC is shown.

In MPC, at each time step k the optimal control signal is computed (by numerical optimization) over a prediction horizon of N_p steps. A control horizon $N_c (< N_p)$ can be selected to reduce the number of variables and to improve the stability of the system. Beyond the control horizon the control signal is usually taken to be constant. From the resulting optimal control signal only the first sample of the computed control signal is applied to the process. In the next time step $k + 1$, a new optimization is performed with a prediction horizon that is shifted one time step ahead, and of the resulting control signal again only the first sample is applied, and so on. This scheme, called rolling horizon, allows for updating the state from measurements, or even for updating the model in every iteration step.

In other words, MPC is equivalent to optimal control extended with feedback. The advantage of updating the state through feedback is that this results in a controller that has a low sensitivity to prediction errors. Regularly updating the prediction model results in an adaptive control system, which could be useful in situations where the model significantly changes, such as in case of incidents or changing weather conditions.
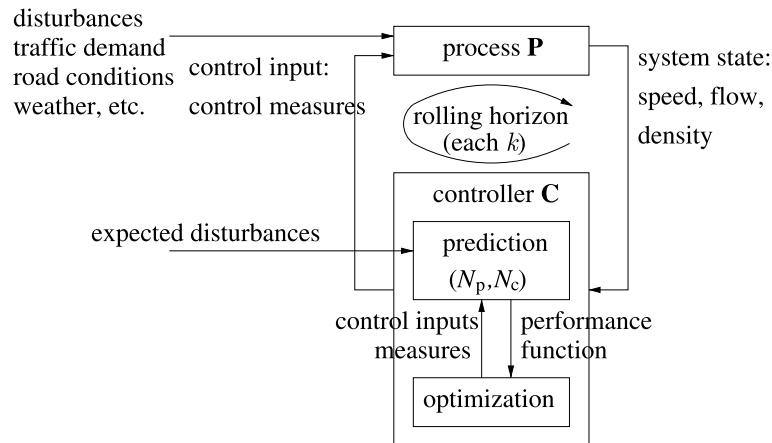
Freeway Traffic Management and Control, Figure 3
The model predictive control (MPC) structure

Acronyms and Abbreviations

MPC   Model Predictive Control
OD   Origin-Destination
ADAS Advanced Driver Assistance Systems
AHS   Automated Highway System
IVHS Intelligent Vehicle/Highway System

Definition of the Subject

The goal of this chapter is to provide an overview of dynamic traffic control techniques described in the literature and applied in practice. *Dynamic traffic control* is the term to indicate a collection of tools, procedures, and methods that are used to intervene in traffic in order to improve the traffic flow on the short term, i. e., ranging from minutes to hours. The nature of the improvement may include increased safety, higher traffic flows, shorter travel times, more stable traffic flows, more reliable travel times, or reduced emissions and noise production.

The tools used for this purpose are in general changeable signs (including traffic signals, dynamic speed limit signs, and changeable message signs), radio broadcast messages, or human traffic controllers at the location of interest. Moreover, currently the possibilities of assisting, informing, and guiding drivers via in-car systems are also being explored.

The term *dynamic traffic management* includes besides dynamic traffic control also the management of emergency services and non-automated procedures (such as the implementation of predefined traffic control scenarios during special events), typically performed in traffic control centers. However, in this chapter the focus is on automatic control methods. Furthermore, this chapter deals exclu-

sively with dynamic freeway traffic control techniques. Given the differences in traffic operation (e. g., higher speed limits) and in traffic infrastructure (e. g., intersections versus on-ramps and off-ramps), the control measures that can be implemented for urban and for freeway traffic differ. The interested reader is referred to [28,35,90] for an overview of urban traffic control.

Introduction

The number of vehicles and the need for transportation is continuously growing, and nowadays cities around the world face serious traffic congestion problems: Almost every weekday morning and evening during rush hours the capacity of many main roads is exceeded. Traffic jams do not only cause considerable costs due to unproductive time losses; they also augment the probability of accidents and have a negative impact on the environment (air pollution, lost fuel) and on the quality of life (health problems, noise, stress).

One solution to the ever growing traffic congestion problem is to extend the road network. Extending the freeway infrastructure is rather expensive, and in many countries this option is currently not considered to be a viable solution. Moreover, in densely populated areas building new roads is sometimes even unfeasible due to spacial limitations. Furthermore, there are often also other socio-economic objectives to be achieved, such as environmental objectives, which are considered alongside the objective of reducing congestion. Dynamic traffic control is an alternative that aims at increasing the safety and efficiency of the existing traffic networks without the necessity of adding new road infrastructure.

Since the 1960s traffic control is applied on freeway systems. However, during the last decades there have been developments in traffic science, traffic technology, control theory, and in the typical traffic patterns that all have consequences for the most appropriate traffic control approach. These developments will be discussed in the next sections.

The Need for Network-Oriented Automatic Traffic Control: Developments

The increasing complexity of the congested traffic patterns and the increasing availability of traffic control measures motivate the increasing usage of automatic traffic control systems and the increasing interest in network-oriented control over the last decades. The interest in network-oriented control from the practitioners' point of view is also motivated from policies aiming at socioeconomic goals, such as the efficient transport over important network corridors.

The fact that the length, the duration, and the number of traffic jams continues to grow has consequences for traffic control. When there are more locations with congestion, the available control measures have to solve more problems, which implies a higher complexity. Since nowadays the chances are higher that a vehicle encounters more than one traffic jam during one trip, the traffic control measures influencing a vehicle in one traffic jam will also influence the other jam(s) that the vehicle encounters. Therefore, the spatial interrelations between traffic situations at different locations in the network get stronger, and consequently the interrelations between the traffic control measures at different locations in the network also get stronger. These interrelations may differ per situation (and depend on, e. g., network topology, traffic demand, etc.) and the control measures may even counteract each other. For the various traffic management agencies or local governments that are responsible for different parts of the traffic network, this means that there is a need for a stronger cooperation and agreement on how the common network goals should be achieved. Similarly, for the automatic control methods, coordinated control strategies are required in these cases, to ensure that all available control measures serve the same objective, or at least that they do not counteract each other.

Another development is that freeways are equipped with more and more traffic control measures. The increasing number of control measures augments the controllability of the freeways. However, with this development the number of possible combinations of control measures also increases drastically, which in its turn in-creases the complexity of the dynamic traffic management problem.

On modern freeways often a large amount of data is available on-line and off-line. This data can serve as a basis for choices about appropriate control measures given the actual and expected traffic situation. However, the available data is currently not fully utilized either by traffic control center operators, whose actions are typically based on heuristic reasoning, or by automatic control measures, which mostly use local data only. Traffic data may also contain information about the current disturbances of the network (incidents, weather influences, unexpected demands) and information about the traffic system at a network level (about route choice and origin-destination relationships). The origin-destination (OD) matrix describes the traffic demand (vehicles per hour) appearing at each origin in a traffic network towards each destination in the network. An OD matrix may be time-varying, and can be calculated at different levels of temporal aggregation, e. g., hourly, peak or interpeak, 24 hour, etc. Methods have been developed to estimate such OD relationships from traffic measurements, and to estimate the traffic state (e. g., speeds, flows and densities) in networks that are incompletely equipped with detectors [43,124]. The methods can be used to supply a traffic control system with more accurate data, leading to better control actions.

These developments motivate the application of automatic control systems that can handle complex traffic scenarios, multiple control measures, and a large amount of data, and that can benefit from the network-oriented information by selecting appropriate control measures for given OD patterns and disturbances.

Regardless of these developments the three effects that cause the majority of suboptimal traffic network performances – in terms of travel time – have remained the same. Therefore, the primary goal of traffic control was and is to resolve the following effects/issues:

- **Bottlenecks.** Typical bottlenecks are freeway sections with an on-ramp, bridges, tunnels, curves, grades, and merge and diverge areas. The performance degradation typically originates from the phenomenon that the maximum achievable outflow from a traffic jam created at a bottleneck is often lower than the capacity of the road. This phenomenon is often called the capacity drop. A special case of a bottleneck is an upstream propagating jam that is growing at the tail by the incoming vehicles and resolving at the head by the leaving vehicles. A moving jam can be a serious bottleneck as it could reduce the maximum outflow to 70% of the capacity [59], while the capacities of the other bottle-

necks are in the range of 85–100% [37,59]. Dynamic traffic control measures may help to prevent a traffic breakdown at a bottleneck, or to improve the flow when a breakdown has occurred.

- **Suboptimal route choice.** In a dynamically changing network with jams, incidents, and road works the driver may not always be informed sufficiently well to make the optimal route choice. Furthermore, even if each individual driver has found the quickest (or in general, least costly) route to his or her destination, it may not lead to optimal performance at the network level, as known from the famous example of Braess [11]. Systems that influence the drivers' route choice may contribute to a better performance for the users, the network, or even both.
- **Blocking.** The tail of a traffic jam on a freeway may propagate so far upstream that it blocks traffic on a route that is not leading over the bottleneck that has caused the jam. A typical case is when a traffic jam created on the freeway at an on-ramp propagates back to an upstream off-ramp and blocks the traffic that wants to leave via the off-ramp. Figure 4 illustrates a situation where off-ramp traffic is blocked by a jam originating from the downstream on-ramp. All control measures that can limit the length of a traffic jam may in principle be applied to prevent blocking.

Automatic traffic control strategies try to optimize traffic network performance. A simplified, idealized description of the operation of traffic in the network links is given by what is known in traffic theory as the fundamental diagram [81]. The fundamental diagram describes steady-state traffic operation on a homogeneous freeway (i. e., the spatial gradients of speed, flow and density are equal to zero) as illustrated in Fig. 5. For low traffic densities, the relation between traffic density and traffic flow is nearly linear. For traffic densities smaller than the critical density ρ_{crit}, the traffic flow on the freeway increases with increasing traffic density (Fig. 5), despite the fact that the average speed decreases with increasing traffic density. Traffic operation is in a stable regime for traffic densities lower than the critical density. The maximal flow that can be achieved
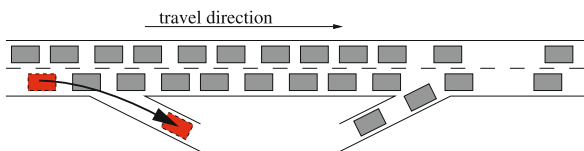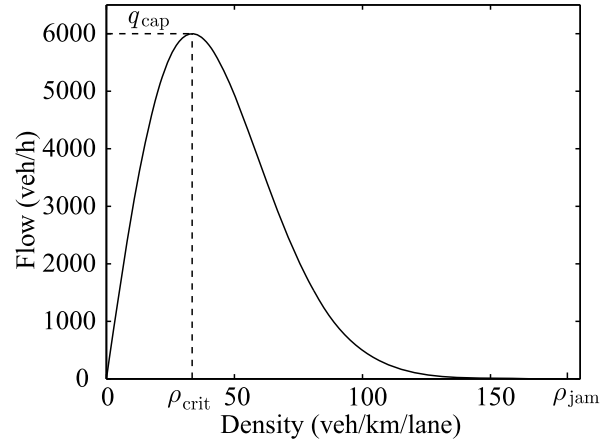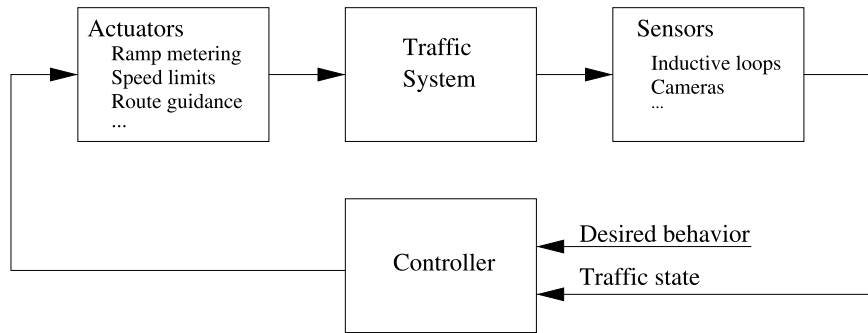

Freeway Traffic Management and Control, Figure 5
A flow-density fundamental diagram for a three lane freeway. As long as the traffic density on the freeway is smaller than the critical density ρ_{crit}, the traffic flow on the freeway increases with increasing traffic density. If the traffic density reaches the critical density, the flow is maximal and equal to the freeway capacity q_{cap}. If the traffic density further increases, the traffic flow on the freeway starts to decrease with increasing traffic density until the traffic stalls at the jam density ρ_{jam}

on the freeway, the capacity q_{cap}, is reached for a traffic density equal to the critical density and the resulting average speed of the vehicles is called the critical speed. If the critical density is exceeded, the average speed continues to decrease and the traffic flow decreases with increasing density. For traffic densities higher than the critical density, congestion sets in and an unstable traffic regime results. Typical values of ρ_{crit} and q_{cap} for a three-lane highway are 33.5 vehicles per kilometer and per lane and 6000 vehicles per hour respectively [94].

Based on the discussion of the fundamental diagram presented above, it can be observed that automatic control strategies can try to prevent or to reduce congestion by steering the state of traffic operation towards the stable region of operation. Here the fundamental diagram is merely presented as a seminal approach to traffic state analysis. However, in the literature alternative approaches to traffic state analysis have been reported. e. g., Kerner [60] has proposed the three phase theory, introducing the concepts "wide moving jams" and "synchronized traffic". The work by Treiber et al. [117] and Lee et al. [70] has added additional traffic states to the analysis such as "oscillating congested traffic" and "homogeneous congested traffic".

Automatic Traffic Control

Dynamic traffic management systems typically operate according to the feedback control concept known from con-

Freeway Traffic Management and Control, Figure 4
Congestion caused by high on-ramp demand could also result in the blocking of an upstream off-ramp

Freeway Traffic Management and Control, Figure 6
Schematic representation of the dynamic traffic management control loop. Based on the measurements provided by the sensors the controller determines the control signals sent to the actuators. Since the control loop is closed, the deviations from the desired traffic system behavior are observed and appropriate control actions are taken

trol systems theory, as shown in Fig. 6. The traffic sensors provide information about the current traffic state, such as speed, flow, density, or occupancy. The controller determines appropriate control signals that are sent to the actuators (depending on the system the changes in the control signal may be implemented instantly or may need to be phased in). The reaction of the traffic system is measured by the sensors again, which closes the control loop. If the new measurements show a deviation from the desired behavior (caused, e. g., by unforeseen disturbances), the new control signals are adapted accordingly. Note that there also exist traffic control systems that have a feedforward structure, e. g., the demand-capacity ramp metering approach that will be discussed in Sect. "Ramp Metering".

We define an "appropriate control" signal in terms of a control objective. From the network operator's point of view typical objectives are:

- **Efficiency.** Efficiency is often expressed in terms of throughput or travel time. This objective is shared by the network operators and the individual drivers. Nevertheless, situations may arise when the network operator and the individual driver have conflicting interests, e. g., minimizing the total travel time in a network (network optimum) may be conflicting with individually minimizing the travel times (user optimum). This will be discussed in more detail in Sect. "Route Guidance".
- **Safety.** In traffic control, safety may be a direct goal of the control or it may be a constraint (boundary condition) that should be satisfied. For example, dynamic speed limits or variable message signs may reduce the speed limit or give a warning under adverse weather conditions or poor visibility conditions in order to improve safety. Other systems may have other primary

goals, such as improving the flow, and in these cases the control systems are often still required to be safety-neutral compared with the situation without control. There may also be an interaction between safety and efficiency, which has to be taken into account in the design of the control system. This interaction may be related to the following three processes. First, a safer traffic system in general results in fewer accidents and therefore more often higher flows may occur. Second, if congestion is prevented by an appropriate control method safety may be increased due to the more homogeneous flows. Third, lower speeds and densities in general positively influence safety. More specifically, Brownfield et al. [13] observed that for freeway sites, the accident rate in congested conditions was nearly twice the rate in uncongested conditions. However, the proportion of accidents that were serious or fatal was lower in congested conditions than in uncongested conditions. Hence, depending on the network operator's definition of safety, safety and efficiency may be conflicting or non-conflicting objectives.

- **Network reliability.** Even if not every traffic jam can be prevented, it is valuable for drivers when the travel time to their destinations is predictable, since good arrival time estimations make departure time choices easier. Therefore, improving the network reliability/predictability serves the economic efficiency of the network and improves driver comfort. Traffic control in general can improve reliability (predictability) by aiming at the realization of predicted travel times, or the reverse, by predicting realizable travel times, or both. Furthermore, network reliability can be improved by measures that aim at synchronization of the traffic demand and the capacity supply of the network, and at a better distribution traffic of flows over the

network. For a more elaborate discussion on network reliability we refer the interested reader to [6,21,78, 113,114].

- **Low fuel consumption, low air and noise pollution.** In general, congestion contributes to less smooth journeys (more deceleration-acceleration movements), which increases emissions. In or near urban areas the environmental effects of traffic may be considered more important than, efficiency, for example, which can result in a different trade-off between the two objectives. An example of such a trade-off is between travel speed and air pollution [3]. The typical measure for these purposes is speed limitation.

Another important aspect of a traffic control system are the constraints due to physical, technical, or policy-based limitations. Such constraints may include minimum and maximum ramp metering rates, maximum on-ramp queue length, minimum and maximum dynamic speed limit values, etc. The automatic traffic controller is required to cope with these constraints.

Chapter Overview

The remainder of this chapter is organized as follows. We discuss the sensor technologies used in the context of freeway traffic control in Sect. "Sensor Technologies". In Sect. "Traffic Flow Modeling" we address traffic flow models, which play an important role in the design and evaluation of traffic control strategies. Next, the most frequently used freeway control measures are discussed in Sect. "Freeway Traffic Control Measures". While in Sect. "Freeway Traffic Control Measures" the focus is on the individual control measures, in Sect. "Network-Oriented Traffic Control Systems" we discuss the approaches to combine and to integrate several control measures for network-oriented control. We conclude in Sect. "Future Directions" by considering the new developments that are expected to play a role in future freeway traffic control systems.

Sensor Technologies

In order to implement traffic responsive freeway control, traffic measurements need to be collected at different locations throughout the freeway network. This section first deals with the most common traffic variables and traffic sensors to collect them. Next, the need for traffic demand and traffic routing data, and the way these data can be obtained using common traffic measurements, is briefly addressed. New data collection technologies, that will play an important role in future freeway control systems are discussed in Sect. "Future Directions".

Measurements

Traditionally the following traffic variables are measured to determine the traffic state on a freeway [81]: the traffic flow or the traffic intensity on the freeway, the average speed of the vehicles, the traffic density, the occupancy level of the freeway, the time headways (and in some cases the distance headways) and the speed variance. Note that occupancy is defined as the relative time (percentage) that the traffic sensor is occupied by a vehicle. In practice, it is often used as a surrogate measure for traffic density since it is directly related to density (as long as the average vehicle length is constant) and can be measured more easily than density.

Depending on the application and on the traffic measurement system, several levels of detail can be distinguished. The traffic variables can either be measured for every freeway lane separately or a value averaged over all lanes of the freeway can be obtained. Some measurement systems allow for a classification of the vehicles in categories based on their size (e.g., trucks versus cars) and provide the traffic variables per category. Furthermore, instantaneous values of the measured traffic variables can be provided or values averaged over a time period. The period over which the measurements are averaged can range from seconds over minutes to hours. As a rule of thumb, one can assume that the higher the level of detail of the data collected, the higher the cost of the measurement system involved.

In real-life situations, the measurements that are provided by the traffic sensors will contain measurement errors. These errors include incidental missing values, incidental faulty measurements, biased measurements, and missing values over a period of time. Given the importance of the traffic measurements in the dynamic traffic management control loop (Fig. 6), these errors need to be detected. Depending on the application, the controller may or may not be able to deal with erroneous or missing values. Techniques to estimate missing values that have been reported in the literature include reference days [24], multiple imputation [77], time series analysis [24], and Kalman and particle filtering [43,124].

For the purpose of freeway traffic control the most commonly measured traffic variables are the traffic flow, the speed, and the occupancy of the freeway traffic. The choice of these variables is influenced by their importance in traffic theory as well as by the ease by which they can be measured with most common traffic detector technologies.

There exists a wide variety of technologies [58] to measure traffic variables such as, pneumatic sensors, inductive

loops, cameras, ultrasonic sensors, microwave sensors, active and passive infrared sensors, passive acoustic arrays, and magnetometers.

Inductive loops are the most widespread detection systems to date and were introduced as traffic detection systems in the 60s [85]. The main advantages of inductive loops are their wide application range, the flexible design, and the availability of the common traffic variables. The main disadvantages of inductive loop technology are the sensitivity to wear and tear due to physical stress on the loops induced by traffic, the susceptibility of the loops to damage by road maintenance works, and the special installation and maintenance requirements (such as lane closure during maintenance) [58].

Traffic detection using video cameras emerged during the 80s and is a non-intrusive technology that is becoming more and more popular [85]. A fixed camera is mounted above the freeway and its images are sent to a video processing unit that extracts the desired variables using image processing algorithms. Camera traffic detection technology can provide the common traffic variables, is less prone to wear and tear by traffic, and generally requires less lane closures for maintenance and reconfiguration. The main disadvantages of traffic cameras are the higher upfront cost of the installation compared to inductive loops and the dependence on visibility conditions (such as fog, heavy snow, sunlight shining directly into the camera could heavily impair the quality of the images taken by the traffic cameras). Other factors that adversely affect the detection accuracy are vibrations caused by wind and traffic, lack of contrast between vehicle and road color, and varying lighting conditions (such as during dusk and dawn) [33,63,64].

In addition to the registration of the traditional traffic variables, video camera technology is also applied to register travel times on corridors or to obtain information regarding the routes followed throughout the network. This information can be obtained by tracking the vehicles at strategic locations (such as at large junctions or at the entrances and the exits of the area under consideration) using automated license plate recognition algorithms, for example. These systems consist of video image processing units connected to video cameras monitoring the traffic. Often, these systems are implemented at sites where the hardware to register the vehicles is already available (e. g., automated toll booths) [19,111].

Estimation

In order to coordinate or to integrate traffic control measures, the spatial aspect of the traffic network needs to be taken into account such that the impact of control mea-

sures on distant parts of the traffic network can be accounted for. However, the traffic sensors discussed above are traffic sensors that are localized in space, and as a consequence, they only yield information on the evolution of the traffic state on the freeway through time and at the sensor locations. Implementing traffic detectors very densely on the freeway in order to register the traffic states for every freeway section would be inconvenient and costly. However, data fusion techniques (such as extended Kalman filtering or particle filtering) allow one to combine traffic measurements scattered over the traffic network in order to obtain network traffic state estimation [43,124,126]. Traffic state estimation and prediction can be used in the implementation of control measures as is illustrated by simulation in [9].

The similarities between traffic flows and flows of compressible fluids have since long been documented in the literature [75,102] (see also Sect. "Traffic Flow Modeling" below). However, when it comes to the routing of traffic flows in networks, a major difference emerges; while the particles in a fluid have no predetermined destination, the vehicles traveling through a traffic network are traveling from a particular origin to a particular destination. Hence, the destination of the vehicles constrains the alternative routes that can be chosen. It is clear that route guidance control measures have an impact on the routing process. However, since travel times are typically an important factor in the routing process of well-informed travelers, other traffic measures, in combination with the traffic demands, influence the routing behavior as well. In order to assess the impact of traffic measures on the traffic states in the traffic network, the traffic demands (OD matrices) and the routing process need to be modeled. As the traffic demand cannot be measured directly, it needs to be estimated using the traffic measurements in the traffic network. Several techniques to estimate the traffic demands (OD matrices) that correspond to the measured traffic states in the traffic network have been developed. For an overview of the literature, the interested reader is referred to [76]. The route choice process, which influences the impact of control measures through rerouting effects, is also the subject of research [57].

Traffic Flow Modeling

Traffic flow models can be classified according to various criteria such as *area of application*, *level of detail*, *deterministic versus stochastic* [49].

An example of the application of traffic flow models for the design of traffic control measures is model predictive control, which makes use of an internal prediction model

in order to find the best traffic control measures to be applied to the real traffic process. Since these models are operated in real-time, and are often used to evaluate several control scenarios, they should allow for fast execution on a computer.

For the assessment of traffic control strategies often a simulation model is used instead of (or before) a real-world test. Simulation has several advantages. Above all, simulation is cheaper and faster, and it does not require real human drivers as test subjects. It also provides an environment where the unpredictable disturbances of a field test, such as weather influences, traffic demand variations, and incidents, can be excluded, or if necessary simulations can be repeated under exactly the same disturbance scenario.

Since none of the available traffic models perfectly describes the real traffic behavior, one has to keep in mind the intended application, when making the choice between the available traffic flow models. As Papageorgiou [89] argues for macroscopic traffic flow models an important criterion is that the model should have sufficient descriptive power to reproduce all important phenomena for the intended application. Similar arguments are also used by Kerner [60] but for different phenomena.

Traffic models can also be classified according to the level of detail with which they describe the traffic process:

- **Microscopic** models describe the behavior of individual vehicles. Important aspects of microscopic models are the so-called *car following* and *lane changing* behavior. Car following and lane changing behavior is generally described as a function of the distance to and (relative) speed of the surrounding vehicles, and the desired speed. Since the vehicles are modeled individually in microscopic traffic models, it is easy to assign different characteristics to each vehicle. These characteristics can be related to the driving style of the driver (aggressive, patient), vehicle type (car, truck), its destination, and route choice.

  A special type of microscopic traffic models are the *cellular-automaton* models [60,80,130] in which the freeway is discretized into cells of about 7.5 m length. Each cell can contain only one vehicle and the traffic dynamics is described by a probabilistic model of the hopping behavior of the vehicles through the cells.

  In general, it is difficult to calibrate microscopic models with real traffic data, due to the large number of parameters in this type of models and the poor availability of appropriate traffic data.

  We refer the interested reader to [2] for an extensive comparison of commercial microscopic simulation models and to [49] for a more theoretical overview.

- **Mesoscopic** models do not track individual vehicles, but describe the behavior of individual vehicles in probabilistic terms. Examples of mesoscopic models are headway distribution models [12] and gas-kinetic models [48]. Typically, these models are not used for traffic control.

- **Macroscopic** models use a high level of aggregation without distinguishing between individual vehicle behavior. Instead, traffic is described in aggregate terms as average speed, average flow, and average density.

  Macroscopic traffic flow modeling started when Lighthill and Whitham [75] presented in 1955 a model based on the analogy between traffic flows and flows in rivers. Independently of Lighthill and Whitham one year later Richards [102] published a similar model. Therefore, this model is usually referred to as the Lighthill–Whitham–Richards (LWR) model.

  Since 1955 a large variety of macroscopic traffic flow models have evolved from the LWR model with differences in the order of the model, the phenomena that they (re)produce (such as capacity drop, stop-and-go waves, and other congestion phenomena or patterns), and the effects of heterogeneous traffic (cars and trucks, etc.) [26,44,48,99].

  Another approach has been followed by Kerner [60] who developed a qualitative traffic flow theory based on empirical observation. This theory distinguishes three so-called traffic phases: free-flow, synchronized flow, and jammed traffic, and describes the transition between these phases qualitatively in probabilistic terms.

A last classification that is relevant in the context of traffic control is whether the model is deterministic or stochastic. Deterministic models define a relationship between model inputs, variables, and outputs that typically describe the average behavior of traffic. Stochastic models describe traffic behavior in terms of relationships between random variables, e. g., random reaction time of drivers, randomness in equilibrium speed-density (or car following) relationships, route choice, etc. These stochastic effects can reproduce phenomena such as the creation of traffic jams by random fluctuations in traffic flows [109], and can be used for the stochastic evaluation of traffic control approaches. Another application of stochastic traffic flow models is in the area of state estimation, which is an essential part of control approaches such as optimal control or model predictive control [43,124].

Freeway Traffic Control Measures

In this section we give an overview of control measures that are used or could be used to improve traffic perfor-

mance. We focus on control measures that are currently applied, or could be applied in the near future, such as ramp metering, speed limits, and route guidance. For each control measure we present the principle of operation including the control approaches, and the existing field tests and simulation results. At the end of this section some other traffic control measures are presented that may also be used to improve the performance of traffic systems.

Ramp Metering

Principle of Operation   Ramp metering is one of the most investigated and applied freeway traffic control measures. A ramp metering set-up is implemented as a traffic signal that is placed at the on-ramp of a freeway as shown in Fig. 7. The required metering rate is implemented by appropriately choosing the phase lengths of the traffic signal. Several ramp metering implementations can be distinguished [20], e. g., single-lane with one vehicle per green ramp metering, single-lane with multiple vehicles per green ramp metering (bulk metering), and dual-lane ramp metering.

Ramp metering can be used in two modes: the *traffic spreading mode* and the *traffic restricting mode*. In the traffic spreading mode ramp metering smooths the merging process of on-ramp traffic by breaking the platoons and by spreading the on-ramp traffic demand over time as observed by Elefteriadou [31]. This mitigates the shock waves that can occur under high traffic density conditions. In this application the metering rate equals the average arrival rate of the vehicles.

Freeway Traffic Management and Control, Figure 7
Ramp metering at the freeway A13 in Delft, The Netherlands. One car may pass per green phase. To prevent red-light running the control is enforced

Restrictive ramp metering can be used for three different purposes:

- **Prevention of breakdowns.** When traffic is dense, ramp metering can prevent a traffic breakdown on the freeway by adjusting the metering rate such that the density on the freeway remains below the critical value. Preventing a traffic breakdown has not only the advantage that it results in a higher flow, but also that it prevents the creation of a jam that could block the off-ramp upstream the on-ramp (as shown in Fig. 4). These effects are studied in detail by Papageorgiou and Kotsialos [92]. Daganzo [27] has quantified the role of ramp metering in avoiding the activation of freeway gridlocks.
- **Influencing route choice.** Ramp metering can be implemented to influence the traffic demand and traffic routing. The impact of ramp metering on the traffic state and on the travel times is taken into account by the drivers in their routing behavior [132]. Banks [5] has described a theory to apply ramp metering to influence traffic routing to avoid freeway bottlenecks. Based on a similar idea Middelham [83] has performed a synthetic study on the route choice effects of ramp metering.
- **Localization of traffic jams.** According to Kerner [60] ramp metering can prevent the backpropagation of traffic jams and shock waves occurring at on-ramps. This could be beneficial on the network level since it could localize the traffic jam, and it could also be beneficial to the traffic throughput.

The control strategies that have been developed for restrictive ramp metering can be classified as static or dynamic, fixed-time or traffic-responsive, and local or coordinated.

Fixed-time strategies use (time-dependent) fixed metering rates that are determined off-line based on historical demands. This approach was first suggested by Wattleworth [127], and was extended to a dynamic traffic model by Papageorgiou [87]. The disadvantage of fixed-time strategies is that they do not take into account the day-to-day variations in the traffic demand or the variations in the demand during a period with a constant metering rate, which may result in underutilization of the freeway or inability to prevent congestion.

Traffic-responsive strategies solve these issues by adjusting on-line the metering rate as a function of the prevailing traffic conditions. These strategies also aim at the same objectives as the fixed-time strategies, but use direct traffic measurements instead of historical data to prevent or to reduce congestion. One of the best known strategies

is the *demand-capacity* strategy [91]:

$$q_{ramp}(k) = \begin{cases} q_{cap} - q_{in}(k-1) & \text{if} \quad o_{meas}(k-1) \le o_{cr} \\ q_{r,min} & \text{otherwise} \end{cases}$$

with $q_{ramp}(k)$ the admitted ramp flow at time step k, q_{cap} the downstream freeway capacity, $q_{in}(k)$ the freeway flow measured upstream of the on-ramp at time step k, $q_{r,min}$ the minimal on-ramp flow during congestion, $o_{meas}(k)$ the occupancy downstream the on-ramp at time step k, and o_{cr} is the critical occupancy (at which the flow is maximal). Since the traffic state on the freeway cannot be determined based on the measurement of the traffic flow alone, the downstream occupancy is measured in order to determine whether congestion is present ($o_{meas}(k-1) > o_{cr}$) or not.

A similar strategy is occupancy-based ramp metering, where the upstream traffic flow measurement from demand-capacity ramp metering is replaced by an occupancy measurement. The measured occupancies are related to traffic flows based on historical measurements. Next, the demand-capacity approach described above can be applied [18]. A common disadvantage of both demand-capacity formulations is that they have an (open-loop) feedforward structure, which is known to perform poorly under unknown disturbances and cannot guarantee a zero offset in the output under steady-state conditions.

A better approach is to use a (closed-loop) feedback structure, because it allows for controller formulations that can reject disturbances and have zero steady-state error. ALINEA [96] is such a ramp metering strategy and its control law is defined as follows:

$$q_{ramp}(k) = q_{ramp}(k-1) + K(\hat{o} - o_{meas}(k)) \ ,$$

where $q_{ramp}(k)$ is the metered on-ramp flow at time step k, K is a positive constant, $\hat{o}$ is a set-point for the occupancy, and $o_{meas}(k)$ is the measured occupancy on the freeway downstream of the on-ramp at time step k. ALINEA tries to maintain the occupancy on the freeway equal to a set-point $\hat{o}$, which is chosen in the region of stable operation. Given the probabilistic nature of traffic operation, the set-point $\hat{o}$ is often chosen somewhat smaller than the critical occupancy in order to guarantee free-flow traffic operation.

More advanced ramp metering strategies are the traffic-responsive coordinated strategies such as MET-ALINE [95], FLOW [55], or methods that use optimal control [66] or model predictive control [7].

The ramp metering strategies discussed above attempt to conserve free-flow traffic on the freeway. However, given the probabilistic nature of traffic operation, congestion can set in at lower or higher densities than the

critical density. Based on these insights, Kerner [60,61] defined a *congested-pattern control approach* to ramp metering called ANCONA. The basic idea of ANCONA is to allow congestion to set in, but to keep congested traffic conditions to the minimum level possible. Once congestion sets in, ANCONA tries to reestablish free-flow conditions on the freeway by reducing the on-ramp metering rate. Kerner claims that, by allowing congestion to set in, ANCONA utilizes the available freeway capacity better. The control rule of ANCONA is given by [60]:

$$q_{ramp}(k) = \begin{cases} q_1 & \text{if} \quad v_{det}(k) \le v_{cong} \\ q_2 & \text{if} \quad v_{det}(k) > v_{cong} \end{cases} \ ,$$

where $q_{ramp}(k)$ is the on-ramp flow at time step k, q_1 and q_2 are heuristically determined constant flows with $q_1 < q_2$, $v_{det}(k)$ is the traffic speed on the freeway just upstream of the on-ramp at time step k, and v_{cong} is the speed threshold that separates the free and the synchronized (locally congested) traffic flow phases.

Field Tests and Simulation Studies Several field tests and simulation studies have shown the effectiveness of ramp metering. In Paris on the Boulevard Périphérique and in Amsterdam several ramp metering strategies have been tested [97]. The demand-capacity, occupancy, and ALINEA strategies were applied in the field tests at a single ramp in Paris. It was found that ALINEA was clearly superior to the other two in all the performance measures (total time spent, total traveled distance, mean speed, mean congestion duration). At the Boulevard Périphérique in Paris the multi-variable (coordinated) feedback strategy MET-ALINE was also applied and was compared with the local feedback strategy ALINEA. Both strategies resulted in approximately the same performance improvement [95]. One of the largest field tests was conducted in the Twin Cities metropolitan area of Minnesota. In this area 430 operational ramp meters were shut down to evaluate their effectiveness. The results of this test show that ramp metering not only serves the purposes of improving traffic flow and traffic safety, but also improves travel time reliability [16,72].

A number of studies have simulated ramp metering for different transportation networks and traffic scenarios, with different control approaches, and with the use of microscopic and macroscopic traffic flow models [39,45,66,94,96,112]. Generally the total network travel time is considered as the performance measure and is improved by about 0.39–30% when using ramp metering.

The validation of ANCONA versus ALINEA is performed by simulation by Kerner [61], who found that AN-CONA in some cases can lead to higher flows.

Also note that Kerner has criticized modeling approaches to simulations of freeway traffic control strategies in [62] (see also [98] for some comments).

Dynamic Speed Limits

Dynamic speed limits are used to reduce the maximum speed on freeways according to given performance, safety, or environmental criteria. An example of a dynamic speed limit gantry is shown in Fig. 8.

Principle of Operation   The working principle of speed limit systems can be categorized based on their intended effects: improving safety, improving traffic flow, or their environmental effects, such as reducing noise or air pollution.

It is generally accepted that speed reduction on freeways leads to improved safety [106,116,128]. Lower speeds in general are associated with lower crash rates and with a lower impact in case of a collision. If the environmental conditions or traffic conditions are such that the posted maximum speeds are considered to be unsafe, the speed limit can be lowered to match the given conditions. Dynamic speed limits may function as a warning that an incident or jam is present ahead.

In the literature, basically two approaches to dynamic speed limit control can be found for flow improvement. The first emphasizes the homogenization effect [1,38,69,109,122], whereas the second is more focused

Freeway Traffic Management and Control, Figure 8
A variable speed limit gantry on the A13 freeway near Overschie, The Netherlands. In this particular case the maximum speed limit is 80 km/h due to environmental reasons, and the limit may drop to 70 km/h or 50 km/h in case of a downstream jam

on preventing traffic breakdown by reducing the flow by means of speed limits [23,42,71].

- The basic idea of homogenization is that speed limits can reduce the speed (and/or density) differences, by which a more stable (and safer) flow can be achieved. The homogenizing approach typically uses speed limits that are above the critical speed (i. e., the speed that corresponds to the maximal flow). So, these speed limits do not limit the traffic flow, but only slightly reduce the average speed (and slightly increase the density). In theory this approach can increase the time to breakdown slightly [109], but it cannot suppress or resolve shock waves. An extended overview of speed limit systems that aim at reducing speed differentials is given by Wilkie [128].

- The traffic breakdown prevention approach focuses more on preventing too high densities, and also allows speed limits that are lower than the critical speed in order to limit the inflow to these areas. By resolving the high-density areas (bottlenecks) higher flow can be achieved in contrast to the homogenization approach [42].

Currently, the main purpose of most of the existing practical dynamic speed limit systems is to increase safety by lowering the speed limits in potentially dangerous situations, such as upstream of congested areas or during adverse weather conditions [106,116,128]. Although these systems primarily aim at safety, in general they also have a positive effect on the flow, due to the fact that preventing accidents results in a higher flow. There are also some examples of practical systems that are designed with the purpose of flow improvement [101,104] – with varying success. These practical systems in general use a switching scheme based on traffic conditions, weather conditions, visibility conditions, or pavement conditions [100,129].

Several control methodologies are used in the literature to find a control law for speed control, such as multi-layer control [74], sliding-mode control [71], and optimal control [1]. In [29] optimal control is approximated by a neural network in a rolling horizon framework. Other authors use, or simplify their control law to, a control logic where the switching between the speed limit values is based on traffic volume, speed, or density [38,69,71,109,122]. We refer the interested reader for further reading about the various control methodologies to the references at the end of this chapter.

Some authors recognize the importance of anticipation in the speed control scheme. A pseudo-anticipative scheme is used in [71] by switching between speed limits based on the density of the neighboring downstream seg-

ment. Explicit predictions are used in [1,29] and this is the only approach that results in a significant flow improvement. The heuristic algorithm proposed in [128] also contains anticipation to shock waves being formed.

Another concept of dynamic speed limits is their use in combination with ramp metering to prevent a breakdown on the freeway at the on-ramp and to prevent the ramp queue to propagate back to the urban network [41] by taking over the flow limitation function from the ramp metering when the ramp queue has reached its limit.

Field Tests and Simulation Studies   Field data evaluations show that in general homogenization results in a more stable and safer traffic flow, but no significant improvement of traffic volume is expected nor measured [104,122]. Since the introduction of speed control on the M25 in the United Kingdom, an increase of flow of 1.5% per year is reported for the morning peaks, but no improvement is found in the afternoon peaks [101].

The effect of dynamic speed limits on traffic behavior strongly depends on whether the speed limits are enforced or not, and on whether the speed limits are advisory or mandatory, which also determines the suitability for a certain application. Most application oriented studies [110,122,128] enforce speed limits, except for [69]. Enforcement is usually accepted by the drivers if the speed limit system leads to a more stable traffic flow.

Route Guidance

Principle of Operation   Route guidance systems assist drivers in choosing their route when alternative routes exist to their destination. The systems typically display traffic information such as congestion length, the delay on the alternative routes, or the travel time to the next common point on the alternative routes (an example is given in Fig. 9). Recently, in-car navigation system manufacturers have shown interest in providing route advice taking the traffic jams and travel times on the alternative routes into account.

In route guidance the notions of *system optimum* and *user equilibrium* (or *user optimum*) play an important role. The system optimum is achieved when the vehicles are guided such that the total costs of all drivers (typically the total travel time) is minimized. However, the system optimum does not necessarily minimize the travel time (or some generalized cost measure) for each individual driver. So, some drivers may select another route that has a shorter individual travel time (lower cost). The traffic network is in user equilibrium when on each utilized route the costs are equal, and on routes that are not utilized the

Freeway Traffic Management and Control, Figure 9
A route guidance system in The Netherlands showing traffic jams lengths on alternative routes to Schiphol Airport (Photo courtesy of Peek Traffic B.V.)

cost is higher than that on the utilized routes. This means that no driver has the possibility to find another route that reduces his or her individual cost.

The cost function is typically defined as the travel time, either as the *predicted travel time* or as the *instantaneous travel time*. The predicted travel time is the time that the driver will experience when he or she drives along the given route, while the instantaneous travel time is the travel time determined based on the current speeds on the route. In a dynamic setting the speeds in the network may change during a trip, and consequently the instantaneous travel time may be different from the predicted travel time.

Papageorgiou and Messmer [93] have developed a theoretical framework for route guidance in traffic networks. Three different traffic control problems are formulated: an optimal control problem to achieve the system optimum (minimize the total time that is spent in the network), an optimal control problem to achieve a user optimum (equalize travel times), and a feedback control problem to achieve a user optimum (equalize travel times). The resulting control strategies are demonstrated on a test network with six pairs of alternative routes. The feedback control strategy is tested with instantaneous travel times and results in a user equilibrium for most alternative routes, and the resulting total time spent by the vehicles in the network is very close to the system optimum.

Wang et al. [125] combine the advantages of a feedback approach (relatively simple, robust, fast) and predicted travel times. The resulting *predictive feedback* controller is compared with optimal control and with a feedback controller based on instantaneous travel times.

When the disturbances are known the simulation results show that the predictive feedback results in nearly optimal splitting rates, and is clearly superior to the feedback based on instantaneous travel times. The robustness of the feedback approach is shown for several cases: incorrectly predicted demand, an (unpredictable) incident, and an incorrect compliance rate.

Field Tests and Simulation Studies In several studies it is assumed that the turning rates can be directly manipulated by route guidance messages [93,125]. In the case of in-car systems it is plausible that by giving direct route advice to individual drivers the splitting rates can be influenced sufficiently. However, in the case of route guidance by variable message signs the displayed messages do not directly determine the splitting rates: The drivers make their own decisions. Therefore, empirical studies about the reaction of drivers to dynamic route information messages, and the effectiveness of route guidance can provide useful information.

Kraan et al. [68] present an extensive evaluation of the impact on network performance of variable message signs on the freeway network around Amsterdam, The Netherlands. Several performance indicators are compared before and after the installation of 14 new variable message signs (of which seven are used as incident management signs and seven as dynamic route information signs). The performance indicators, such as the total traveled distance, the total congestion length and duration, and the instantaneous travel time delay are compared for alternative routes and for most locations a small but statistically significant improvement is found. The day-to-day standard deviation of these indicators decreased after the installation of the variable message signs, which indicates that the travel times have become more reliable.

Another field test is reported by Diakaki et al. [30] in which a combination of route guidance, ramp metering, and urban traffic control is applied to the M8 corridor network in Glasgow, UK. The applied control methodology resulted in an increased network throughput and in a reduced average travel time.

Other Control Measures

Besides ramp metering, dynamic speed limits, and route guidance, there are also other dynamic traffic control measures that can potentially improve the traffic performance. In this section we describe a selection of such measures, and describe in which situations they are useful (cf. [84]).

- **Peak lanes.** During peak hours the hard shoulder lane of a freeway (which is normally used only by vehi-

cles in emergency) is opened for traffic. Whether the lane is opened or closed is communicated by variable message signs showing a green arrow or a red cross. Due to the extra lane the capacity of the road is increased, which could prevent congestion. The disadvantage of using the emergency lane as a normal lane is that the safety may be reduced. For this reason, often extra conditions ensuring safety are required, such as the creation of emergency refuges adjacent to the hard shoulder lane, or the requirement that emergency services should be able to access the incident location over or through the guard rail. Furthermore, there may be CCTV surveillance or vehicle patrols to detect incidents early. This traffic control measure is useful where the additional capacity prevents congestion and the downstream infrastructure can accommodate the increased traffic flow.

- **Dedicated lanes.** During congestion the shoulder lane may be opened for dedicated vehicles, such as public transport, freight transport, or high occupancy vehicles (with more than two passengers). This reduces the hindrance that congestion causes to these vehicles. Furthermore, public transport can be made more reliable and thus more attractive by this measure. A dedicated freight transport lane increases the stability and homogeneity of the traffic flow.

- **Tidal flow.** Tidal flow allows one to use a freeway lane in the one or the other direction. Depending on the direction of the highest traffic demand the direction of operation is determined. This direction is communicated by a variable message sign showing a red cross or a green arrow. This traffic control measure is useful when the traffic demand is typically not high in both directions simultaneously.

- **The "keep your lane" directive.** When the "keep your lane" directive is displayed, the drivers are not allowed (not recommended) to change lanes. This results in less disturbances in the freeway traffic flow, which may prevent congestion. This traffic control measure is useful when the traffic flow is nearly unstable (close to the critical density) and may be a good alternative to homogenizing speed limits.

Network-Oriented Traffic Control Systems

The integration of traffic control measures in freeway networks is essential in order to ensure that the control actions taken at different locations in the network reinforce rather than counteract or even cancel each other. While in the previous section individual traffic control measures were discussed along with the most prevalent local con-

trol strategies, this section explicitly considers the integration of several control measures in a freeway network context.

Although the focus in this chapter is on automatic control systems, it must be noted that in practice in traffic control centers there is also often a human controller with "oversight" of the system as a safeguard against problems with the system and in view of the complexity of the control problem.

In network-oriented traffic control two ingredients play an important role: *coordination* and *prediction*. Since in a dense network the effect of a local control measure could also influence the traffic flows in more distant parts of the network, the control measures should be coordinated such that they serve the same objectives. Taking into account the effects of control measures on distant parts of the network often also involves prediction, due to the fact that the effect of a control measure has a delay that is at least the travel time between the two control measures in the downstream direction, and at least the propagation time of shock waves in the upstream direction. An advantage of control systems that use explicit predictions is that by anticipating on predictable future events the control system can also *prevent* problems instead of only *reacting* to them. However, it must be noted that while all network-oriented control approaches apply some form of coordination, many approaches do not explicitly make use of predictions.

Network-oriented traffic control has several advantages compared to local control since it ensures that local traffic problems are solved with the aim of achieving an improvement on the network level. For example, solving a local traffic jam only can have as consequence that the vehicles run faster into another (downstream) jam, whereas still the same amount of vehicles have to pass the downstream bottleneck (with a given capacity). In such a case, the average travel time on the network level will still be the same, regardless of whether or not the jam is solved. However, a global approach would take into account both jams and, if possible, solve both of them.

Furthermore, network-oriented control approaches can utilize network-related historical information. For example, if dynamic OD data is available, control on the network level can take advantage of the predictions of the flows in the network. Local controllers are not able to optimize the network performance even if the dynamic OD data is available, because the effect of the control actions on downstream areas is not taken into account.

In the literature basically three approaches exist for coordinating traffic control measures: model-based optimal control methods, knowledge-based methods, and methods that use simple feedback or switching logic, for which the parameters are optimized. In some approaches different methods are combined in a hierarchical control structure. We discuss these approaches in the following subsections.

Model-Based Control Methods

Model-based traffic control techniques use a traffic flow model for predicting the future behavior of the traffic system based on

- the current state of traffic,
- the expected traffic demand on the network level, possibly including OD relationships and external influences, such as weather conditions,
- the planned traffic control measures.

Since the first two items cannot be influenced (except for the possibility that based on real-time congestion information people cancel their planned trip, change the departure time, or travel via another modality), the future performance of the traffic system is optimized by selecting an appropriate scenario for the traffic control measures. Methods that use *optimal control* or *model predictive control* explicitly take the complex nonlinear nature of traffic into account. For example, they take into account the fact that the effect of ramp metering on distant on-ramps will be delayed by the (time-varying) travel time between the two on-ramps. In general, the other existing methods (such as knowledge-based methods, or control parameter optimization) do not explicitly take this kind of delay into account. Furthermore, other advantages of the model-based methods are that traffic demand predictions can be utilized, constraints on the ramp metering rate and the ramp queues can be included easily, and a user-supplied objective function can be optimized.

Optimal control has been successfully applied in simulation studies to integrated control of ramp metering and freeway-to-freeway control [66], to route guidance [47] and to integration of ramp metering and route guidance [66,67]. In [66,67] the integrated controller performed better than route guidance or ramp metering alone.

The model predictive control (MPC) approach is an extension of the optimal control, which uses a rolling horizon framework. This results in a closed-loop (feedback) controller, which has the advantage that it can handle demand prediction errors and disturbances (such as incidents). MPC is computationally more efficient than optimal control due to the shorter prediction and control horizons, and it can be made adaptive by updating the prediction model on-line.

MPC-based control has been applied in simulations to coordinated ramp metering [8], to integrated control of ramp metering and dynamic speed limits [41], and to integrated control of ramp metering and route guidance [57,121]. An illustration of the ability of MPC-based traffic control to deal with a model mismatch was given in [9]. In [7], it was illustrated by simulation of a simple proof-of-concept network that MPC can be implemented to account for the rerouting behavior of vehicles due to changing travel times caused by applying ramp metering.

Knowledge-Based Methods

Knowledge-based traffic control methods typically describe the knowledge about the traffic system in combination with the control system in terms that are comprehensible for humans. Given the current traffic situation the knowledge-based system generates a solution (control measure) via reasoning mechanisms. A typical motivation for these systems is to help traffic control center operators to find good (not necessarily the best) combinations of control measures. The operators often suffer from cognitive overload by the large number of possible actions (control measures) or by time pressure in case of incidents. The possibility for the operators to track the reasoning path of the knowledge-based system makes these systems attractive and more convincing.

An example of a knowledge-based system is the TRYS system [25,46,86], which uses knowledge about the physical structure of the network, the typical traffic problems, and about effects of the available control measures. The TRYS system has been installed in traffic control centers in Madrid and Barcelona, Spain.

Another knowledge-based system is the freeway incident management system [36] developed in Massachusetts, which assists in the management of non-recurrent congestion. The system contains a knowledge base and a reasoning mechanism to guide the traffic operators through the appropriate questions to manage incidents. Besides incident detection and verification, the system assists in notifying the necessary agencies (such as, ambulance, clean-up forces, towing company) and in applying the appropriate diversion measures. The potential benefits (reduced travel times by appropriate diversion) are illustrated by a case study on the Massachusetts Turnpike. The knowledge-based expert system called freeway real-time expert-system demonstration [103,131] has similar functionalities and is illustrated by applying it to a section of the Riverside Freeway (SR-91) in Orange County, California.

Control Parameter Optimization Methods

Allessandri and Di Febbraro [1] follow another approach: A relatively simple control law is used for speed limit control and ramp metering, and the parameters of the control law are found by simulating a large number of scenarios and optimizing the average performance. In [1] a dynamic speed limit switching scheme is developed. The speed limits switch between approximately 70 km/h and 90 km/h, and the switching is based on the density of the segment to be controlled and two thresholds (to switch up and to switch down). The switching scheme uses a hysteresis loop to prevent too frequent switching. Optimizing the thresholds for several objectives resulted in a slight increase of the average throughput, a decrease of the sum of squared densities – which can be considered as a measure of inhomogeneity (since a non-uniform distribution of vehicles over a freeway stretch results in a higher sum of squared densities) – and a small decrease of the total time spent by the vehicles in the network.

Hierarchical Control

The increasing number of traffic control measures that need to be controlled in a network-control context, as well as their interactions, drastically increases the computational complexity of computing the optimal control signals. Hierarchical control was introduced by some authors in order to tackle this problem [22,65,88]. In hierarchical control the controlled process is partitioned in several subprocesses, and the control task is performed by a high-level controller and several low-level controllers. The high-level controller determines centrally the set-points or trajectories representing the desired behavior of the subprocesses. The low-level controllers are used to steer the subprocesses according to the set-points or trajectories supplied by the high-level controller. Usually, the high-level controller operates at a slower time scale than the low-level controllers. Hierarchical systems do not only enable coordination of control for large networks, but they also provide high reliability and robustness [50].

Future Directions

Although there is a large interest in developing freeway traffic control systems, there is by no means a consensus about the most suitable approaches or methods. One of the reasons is that traffic phenomena, such as traffic breakdown and jam resolution, are not perfectly understood [60] and different views lead to different approaches. In addition, technological developments such as advanced

sensor technologies and intelligent vehicles open new possibilities that enable or require new control approaches.

Advanced Sensor Technologies

Given the complexity of traffic state estimation, traffic demand estimation, and the collection of routing information based on conventional traffic measurement data, new data collection methods are being investigated.

Instead of registering vehicles at certain locations using hardware on the freeway network, floating car data can be collected. The collection of floating car data, where individual vehicles are tracked during their journey through the network, provides valuable route choice and traffic demand information. The evolution in mobile computing and in mobile communication has enabled the incorporation of these technologies in the field of traffic data collection, allowing more detailed and more cost-effective data collection. In contrast to the traditional data collection methods that were discussed in Sect. "Measurements", this section deals with two data collection methodologies that are enabled by mobile computing and communication.

Cell phone service providers collect data regarding the base station each cell phone connects to and the time instant the connection is initiated. Since cell phones regularly connect to their current base station and since the location of these base stations is known, information about the journey of the cell phone can be extracted from the service provider's database. By monitoring a large number of cell phones, and more in particular their hands-off processes when hopping from one base station to the next, an impression of the traffic speeds and the travel times can be obtained [108].

The global positioning system (GPS) is well-suited for tracking probe vehicles through space and time in order to obtain route information and travel times [14,115]. With the further miniaturization of electronics, the processing power available in mainstream navigation units and mobile data communication facilities (e. g., GPRS) the cost of instrumenting fleets of probe vehicles decreases. For example, fleets of taxis, buses, and trucks can be used as probe vehicles as they are often readily equipped with GPS and data communication technology. When dealing with probe vehicles, care must be taken to ensure that the number of probe vehicles is large enough in order to be able to accurately determine the traffic state [56].

Although the technologies presented above are readily available and have been used in the past, their structural deployment as a source for traffic measurements for dynamic traffic control systems still needs to break through. Some issues that may determine whether floating car data becomes a viable option for large-scale data collection are the accuracy of the data obtained, privacy concerns related to registering the whereabouts of individuals, operational communication and computation costs, and standardized mobile or in-vehicle availability of communication and GPS functionality.

Intelligent Vehicles and Traffic Control

We now discuss recent and future developments in connection with intelligent vehicles that can further improve the performance of traffic management and control systems by offering better and more accurate ways to collect traffic data and to apply traffic control measures.

Advanced Driver Assistance Systems   The increasing demand for safer passenger cars has stimulated the development of advanced driver assistance systems (ADAS). An ADAS is a control system that uses environment sensors to improve comfort and traffic safety by assisting the driver. Some examples of ADAS are cruise control, forward collision warning, lane departure warning, parking systems, and pre-crash systems for belt-pretensioning [10]. Although traffic management is not the primary goal of ADAS, they can contribute to a better traffic performance [120], either in a more passive way by avoiding incidents and by providing smoother traffic flows, or in an active way by coordination and communication with neighboring vehicles and roadside infrastructure.

The increasing market penetration and use of ADAS and of other in-car navigation, telecommunication, and information systems offer an excellent opportunity to implement a next level of traffic control and management, which shifts away from the road-side traffic management toward a vehicle-oriented traffic management. In this context both inter-vehicle management and road-side/vehicle traffic management and interaction can be considered. The goal is to use the additional measures and control handles offered by intelligent vehicles and to develop control and management methods to substantially improve traffic performance in terms of safety, throughput, reliability, environment, and robustness.

Some examples of new traffic control measures that are made possible by intelligent vehicles are cooperative adaptive cruise control [119] (allowing one to control intervehicle distances), intelligent speed adaptation [17] (allowing one to better and more dynamically control vehicle speeds), and route guidance [82] (where the traffic control centers could on the one hand get data about planned routes and destinations, and on the other hand also send real-time information and control data to the on-board

route planners, for instance, to warn about current and predicted congestion and possibly also to spread the traffic flows more evenly over the network).

These individual ADAS-based traffic control measures could be integrated with roadside traffic control measures such as ramp metering, traffic signals, lanes closures, shoulder lane openings, etc. The actual control strategy could then also make use of a model-based control approach such as MPC.

Cooperative Vehicle-Infrastructure Systems   The new intelligent-vehicle technologies allow communication and coordination between vehicles and the roadside infrastructure and among vehicles themselves. This results in cooperative vehicle-infrastructure systems, which can also be seen as a first step towards fully automated highway systems, which will be discussed below. CVIS (Cooperative Vehicle-Infrastructure Systems) [52] is a European research project that aims to design, develop, and test technologies that allow communication between the cars and with the roadside infrastructure, which improves road safety and efficiency, and reduces environmental impact. This project allows drivers to influence the traffic control system directly and also to get information about the quickest route to their destination, speed limits on the road, as well as warning messages via wireless technologies.

Automated Highway Systems   ADAS and cooperative vehicle-infrastructure systems can even be extended several steps further towards complete automation. Indeed, one approach to augment the throughput on highways is to implement a fully automated system called Automated Highway System (AHS) or Intelligent Vehicle/Highway System (IVHS) [40,123], in which cars travel on the highway in platoons with small distances (say, 2 m) between vehicles within the platoon, and much larger distances (say, 30–60 m) between different platoons. Due to the very short intra-platoon distances this approach requires automated distance-keeping since human drivers cannot react fast enough to guarantee adequate safety. So in AHS every vehicle contains an automated system that can take over the driver's responsibilities in steering, braking, and throttle control. Due to the short spacing between the vehicles within the platoons, the throughput of the highway can increase, allowing it to carry as much as twice or three times as many vehicles as in the present situation. The other major advantages of the platooning system are increased safety and fuel efficiency. Safety is increased by the automation and close coordination between the vehicles, and is enhanced by the small relative speed between the cars in the platoon. Because the cars in the platoon travel together at the same speed, a small distance apart, even high accelerations and decelerations cannot cause a severe collision between the cars (due to the small relative speeds). The short spacing between the vehicles also produces a significant reduction in aerodynamic drag for the vehicles, which leads to improvements in fuel economy and emissions reductions.

Automated platooning has been investigated very thoroughly within the PATH program [54,105]. Related programs are the Japanese Dolphin framework [118] and the Auto21 Collaborative Driving System framework [51, 53].

Although certain authors argue that only full automation can achieve significant capacity increases on highways and thus reduce the occurrences of traffic congestion [123], AHS do not appear to be feasible on the short term. The AHS approach requires major investments to be made by both the traffic authority and the constructors and owners of the vehicles. Since few decisions are left to the driver, and since the AHS assumes almost complete control over the vehicles, which drive at high speeds and at short distances from each other, a strong psychological resistance to this traffic congestion policy is to be expected. Another important question is how the transition of the current highway system to an AHS-based system should occur, and – once it has been installed – what has to be done with vehicles that are not yet equipped for AHS. So before such systems can be implemented, many financial, legislative, political and organizational issues still have to be resolved [34].

Conclusion

In this chapter we have presented an overview of freeway traffic control theory and practice. In this context we have discussed traffic measurements and estimation, individual traffic control measures, and the approaches behind them that relate the control signals to the given traffic situation. The trend of the ever-increasing traffic demands and the appearance of new control technologies have led to the new field of network-oriented traffic control systems. Although there have been many interesting publications about the theory and practice of integrated traffic control, several challenges remain, such as the integration of traffic state estimation and dynamic OD information in the control approaches.

The lively research in freeway traffic control shows that this field is still practically relevant and theoretically challenging. Facing these challenges can be expected to lead to new freeway traffic control approaches in theory and

practice resulting in higher freeway performance in terms of efficiency, reliability, safety and environmental effects. Furthermore, future developments in the field of in-car systems and advanced sensor technologies are expected to enable new traffic management approaches that may measure and control traffic in more detail and with higher performance.

Bibliography

Primary Literature

1. Alessandri A, Di Febbraro A, Ferrara A, Punta E (1999) Non-linear optimization for freeway control using variable-speed signaling. IEEE Trans Veh Technol 48(6):2042–2052
2. Algers S, Bernauer E, Boero M, Breheret L, Di Taranto C, Dougherty M, Fox K, Gabard J-F (2000) SMARTEST – Final report for publication. Technical report, ITS, University of Leeds. http://www.its.leeds.ac.uk/projects/smartest. Accessed on 22 July 2008
3. André M, Hammarström U (2000) Driving speeds in europe for pollutant emissions estimation. Transp Res Part D 5(5):321–335
4. Åström KJ, Wittenmark B (1997) Computer Controlled Systems, 3rd edn. Prentice Hall, Upper Saddle River
5. Banks JH (2005) Metering ramps to divert traffic around bottlenecks: Some elementary theory. Transp Res Rec 1925:12–19
6. Bell MGH (2000) A game theory approach to measuring the performance reliability of transport networks. Transp Res Part B 34(6):533–545
7. Bellemans T (2003) Traffic Control on Motorways. Ph.D. Thesis, Katholieke Universiteit Leuven, Leuven. ftp://ftp.esat.kuleuven.ac.be/pub/SISTA/bellemans/PhD/03-82.pdf. Accessed on 22 July 2008
8. Bellemans T, De Schutter B, De Moor B (2006) Model predictive control for ramp metering of motorway traffic: A case study. Control Eng Practice 14:757–767
9. Bellemans T, De Schutter B, Wets G, De Moor B (2006) Model predictive control for ramp metering combined with extended kalman filter-based traffic state estimation. In: Proceedings of the 2006 IEEE Intelligent Transportation Systems Conference (ITSC 2006), Toronto, Canada, pp 406–411
10. Bishop R (2005) Intelligent vehicle technology and trends. In: Walker J (ed) Artech House ITS Library. Artech House, Boston
11. Braess D (1968) Über ein Paradoxon aus der Verkehrsplanung. Unternehmensforsch 12:258–268
12. Branston D (1976) Models of single lane time headway distributions. Transp Sci 10:125–148
13. Brownfield J, Graham A, Eveleigh H, Maunsell F, Ward H, Robertson S, Allsop R (2003) Congestion and accident risk. In: Technical report, Road Safety Research Report no. 44. Department for Transport, London
14. Byon Y-J, Shalaby A, Abdulhai B (2006) Travel time collection and traffic monitoring via GPS technologies. In: Proceedings of the 2006 IEEE Intelligent Transportation Systems Conference (ITSC 2006), Toronto, Canada, pp 677–682
15. Camacho EF, Bordons C (1995) Model Predictive Control in the Process Industry. Springer, Berlin
16. Cambridge Systematics, Inc. (2001) Twin Cities Ramp Meter Evaluation – Final Report. Cambridge Systematics, Inc., Oakland. Prepared for the Minnesota Department of Transportation
17. Carsten O, Tate F (2001) Intelligent speed adaptation: The best collision avoidance system? In: Proceedings of the 17th International Technological Conference on the Enhanced Safety of Vehicles, Amsterdam, Netherlands, pp 1–10
18. Carvell JD, Balke K Jr, Ullman J, Fitzpatrick K, Nowlin L, Brehmer C (1997) Freeway management handbook. Technical report. Federal Highway Administration, Department of Transport (FHWA, DOT), Washington DC, Report No. FHWA-SA-97-064
19. Castello P, Coelho C, Ninno ED (1999) Traffic monitoring in motorways by real-time number plate recognition. In: Proceedings of the 10th International Conference on Image Analysis and Processing, Venice, Italy, pp 1129–1131
20. Chaudhary NA, Messer CJ (2000) Ramp metering technology and practice. Technical Report 2121-1. Texas Transportation Institute, The Texas A&M University System, College Station
21. Chen A, Yang H, Lo HK, Tang WH (2002) Capacity reliability of a road network: An assessment methodology and numerical results. Transp Res Part B 36(3):225–252
22. Chen OJ, Hotz AF, Ben-Akiva ME (1997) Development and evaluation of a dynamic ramp metering control model. In: Proceedings of the 8th IFAC/IFIP/IFORS Symposium on Transportation Systems, Chania, Greece, June 1997, pp 1162–1168
23. Chien C-C, Zhang Y, Ioannou PA (1997) Traffic density control for automated highway systems. Automatica 33(7):1273–1285
24. Cools M, Moons E, Wets G (2007) Investigating the effect of holidays on daily traffic counts: A time series approach. Transp Res Record 2019:22–31
25. Cuena J, Hernández J, Molina M (1995) Knowledge-based models for adaptive traffic management systems. Transp Res Part C 3(5):311–337
26. Daganzo CF (1994) The cell transmission model: A dynamic representation of highway traffic consistent with the hydrodynamic theory. Transp Res Part B 28B(4):269–287
27. Daganzo CF (1996) The nature of freeway gridlock and how to prevent it. In: Lesort JB (ed) Proceedings of the 13th International Symposium of Transportation and Traffic Theory, Lyon, France
28. Daganzo CF (1997) Fundamentals of Transportation and Traffic Operations. Pergamon, Kidlington
29. Di Febbraro A, Parisini T, Sacone S, Zoppoli R (2001) Neural approximations for feedback optimal control of freeway systems. IEEE Trans Veh Technol 50(1):302–312
30. Diakaki C, Papageorgiou M, McLean T (2000) Integrated traffic-responsive urban corridor control strategy in Glasgow, Scotland. Transp Res Rec 1727:101–111
31. Elefteriadou L (1997) Freeway merging operations: A probabilistic approach. In: Proceedings of the 8th International Federation of Automatic Control (IFAC) Symposium on Transportation Systems, Chania, Greece, pp 1351–1356
32. Federal Highway Administration, US Department of Transportation (1995) Detection Technology for IVHS – Final Report Addendum. Technical report. Federal Highway Administration, US Department of Transportation, Washington, Contract DTFH61-91-C-00076, Report No. FHWA-RD-96-109
33. Federal Highway Administration, US Department of Trans-

portation (1995) Detection Technology for IVHS – Task L Final Report. Technical report. Federal Highway Administration, US Department of Transportation, Washington, Contract DTFH61-91-C-00076, Report No. FHWA_RD-95-100

34. Fenton RE (1994) IVHS/AHS: Driving into the future. IEEE Control Syst Mag 14(6):13–20

35. Gartner NH, Improta G (eds) (1995) Urban Traffic Networks – Dynamic Flow Modeling and Control. Springer, Berlin

36. Gupta A, Maslanka VJ, Spring GS (1992) Development of prototype knowledge-based expert system for managing congestion on massachusets turnpike. Transp Res Rec (1358): 60–66

37. Hall FL, Agyemang-Duah K (1991) Freeway capacity drop and the definition of capacity. Transp Res Rec (1320):91–98

38. Hardman EJ (1996) Motorway speed control strategies using SISTM. In: Proceedings of the 8th International Conference on Road Traffic Monitoring and Control. IEE Conference Publication, no 422. London, 23–25 April 1996, pp 169–172

39. Hasan M, Jha M, Ben-Akiva M (2002) Evaluation of ramp control algorithms using microscopic traffic simulation. Transp Res Part C 10:229–256

40. Hedrick JK, Tomizuka M, Varaiya P (1994) Control issues in automated highway systems. IEEE Control Syst Mag 14(6):21–32

41. Hegyi A, De Schutter B, Hellendoorn J (2005) Model predictive control for optimal coordination of ramp metering and variable speed limits. Transp Res Part C 13(3):185–209

42. Hegyi A, De Schutter B, Hellendoorn J (2005) Optimal coordination of variable speed limits to suppress shock waves. IEEE Trans Intell Transp Syst 6(1):102–112

43. Hegyi A, Girimonte D, Babuška R, De Schutter B (2006) A comparison of filter configurations for freeway traffic state estimation. In: Proceedings of the International IEEE Conference on Intelligent Transportation Systems 2006, Toronto, Canada, 17–20 September 2006, pp 1029–1034

44. Helbing D (1997) Verkehrsdynamik – Neue physikalische Modellierungskonzepte. Springer, Berlin

45. Hellinga B, Van Aerde M (1995) Examining the potential of using ramp metering as a component of an ATMS. Transp Res Record 1494:75–83

46. Hernández J, Cuena J, Molina M (1999) Real-time traffic management through knowledge-based models: The TRYS approach. In: ERUDIT tutorial on Intelligent Traffic Management Models. Helsinki, Finland, 3 August 1999. http://www.erudit.de/erudit/events/tc-c/tut990803.htm. Accessed on 23 July 2008

47. Hoogendoorn SP (1997) Optimal control of dynamic route information panels. In: Proceedings of the 4th World Congress on Intelligent Transportation Systems, IFAC Transportation Systems, Chania, Greece, pp 399–404

48. Hoogendoorn SP, Bovy PHL (2000) Continuum modelling of multiclass traffic flow. Transp Res Part B 34:123–146

49. Hoogendoorn SP, Bovy PHL (2001) State-of-the-art of vehicular traffic flow modelling. J Syst Control Eng – Proc Inst Mech Eng, Part I 215(I4):283–303

50. Hotz A, Much C, Goblick T, Corbet E, Waxman A, Ashok K, Ben-Akiva M, Koutsopoulos H (1992) A distributed, hierarchical system architecture for advanced traffic management systems and advanced traffic information systems. In: Proceedings of the Second Annual Meeting of IVHS AMERICA, Newport Beach

51. http://www.auto21.ca/index.php. Accessed on 23 July 2008

52. http://www.cvisproject.org/. Accessed on 23 July 2008

53. http://www.damas.ift.ulaval.ca/projets/auto21/en/index.html. Accessed on 23 July 2008

54. http://www.path.berkeley.edu/. Accessed on 23 July 2008

55. Jacobson L, Henry K, Mehyar O (1989) Real-time metering algorithm for centralized control. Transp Res Record 1232: 17–26

56. Jiang G, Gang L, Cai Z (2006) Impact of probe vehicles sample size on link travel time estimation. In: Proceedings of the 2006 IEEE Intelligent Transportation Systems Conference (ITSC 2006), Toronto, Canada, 2006, pp 891–896

57. Karimi A, Hegyi A, De Schutter B, Hellendoorn H, Middelham F (2004) Integration of dynamic route guidance and freeway ramp metering using model predictive control. In: Proceedings of the 2004 American Control Conference (ACC 2004), Boston, MA, USA, 30 June–2 July 2004, pp 5533–5538

58. Kell JH, Fullerton IJ, Mills MK (1990) Traffic Detector Handbook, Number FHWA-IP-90-002, 2nd edn. Federal Highway Administration, US Department of Transportation, Washington DC

59. Kerner BS (2002) Empirical features of congested patterns at highway bottlenecks. Transp Res Record (1802):145–154

60. Kerner BS (2004) The Physics of Traffic. Understanding Complex Systems. Springer, Berlin

61. Kerner BS (2007) Control of spatiotemporal congested traffic patterns at highway bottlenecks. IEEE Trans Intell Transp Syst 8(2):308–320

62. Kerner BS (2007) On-ramp metering based on three-phase traffic theory. Traffic Eng Control 48(1):28–35

63. Klein LA, Kelley MR, Mills MK (1997) Evaluation of overhead and in-ground vehicle detector technologies for traffic flow measurement. J Test Evaluation (JTEVA) 25(2):215–224

64. Klein LA, Mills MK, Gibson DRP (2006) Traffic Detector Handbook, Number FHWA-HRT-06-108, 3rd edn. Federal Highway Administration, US Department of Transportation, Washington DC

65. Kotsialos A, Papageorgiou M (2005) A hierarchical ramp metering control scheme for freeway networks. In: Proceedings of the American Control Conference, Portland, OR, USA, 8–10 June 2005, pp 2257–2262

66. Kotsialos A, Papageorgiou M, Middelham F (2001) Optimal coordinated ramp metering with advanced motorway optimal control. In: Proceedings of the 80th Annual Meeting of the Transportation Research Board, vol 3125. Washington DC

67. Kotsialos A, Papageorgiou M, Mangeas M, Haj-Salem H (2002) Coordinated and integrated control of motorway networks via non-linear optimal control. Transp Res C 10(1):65–84

68. Kraan M, van der Zijpp N, Tutert B, Vonk T, van Megen D (1999) Evaluating networkwide effects of variable message signs in the netherlands. Transp Res Record 1689:60–67

69. Kühne RD (1991) Freeway control using a dynamic traffic flow model and vehicle reidentification techniques. Transp Res Record 1320:251–259

70. Lee HY, Lee H-W, Kim D (1999) Empirical phase diagram of traffic flow on highways with on-ramps. In: Helbing D, Herrmann HJ, Schreckenberg M, Wolf DE (eds) Traffic and Granular Flow '99. Springer, Berlin

71. Lenz H, Sollacher R, Lang M (2001) Standing waves and the influence of speed limits. In: Proceedings of the European Control Conference 2001, Porto, Portugal, pp 1228–1232

72. Levinson D, Zhang L (2006) Ramp meters on trial: Evidence

from the twin cities metering holiday. Transp Res Part A 40A:810–828

73. Lewis FL (1992) Applied Optimal Control and Estimation. Prentice Hall, Upper Saddle River

74. Li PY, Horowitz R, Alvarez L, Frankel J, Robertson AM (1995) Traffic flow stabilization. In: Proceedings of the American Control Conference, Seattle, Washington, June 1995, pp 144–149

75. Lighthill MJ, Whitham GB (1955) On kinematic waves, II. A theory of traffic flow on long crowded roads. Proc Royal Soc 229A(1178):317–345

76. Lin P-W, Chang G-L (2007) A generalized model and solution algorithm for estimation of the dynamic freeway origin-destination matrix. Transp Res Part B, 41B:554–572

77. Little R, Rubin D (1987) Handbook of Transportation Science. Wiley, New York

78. Lo HK, Luo XW, Siu BWY (2006) Degradable transport network: Travel time budget of travelers with heterogeneous risk aversion. Transp Res Part B 40(9):792–806

79. Maciejowski JM (2002) Predictive Control with Constraints. Prentice Hall, Harlow

80. Maerivoet S, De Moor B (2005) Cellular automata models of road traffic. Phys Rep 419(1):1–64

81. May AD (1990) Traffic Flow Fundamentals. Prentice Hall, Englewood Cliffs

82. McDonald M, Hounsell NB, Njoze SR (1995) Strategies for route guidance systems taking account of driver response. In: Proceedings of 6th Vehicle Navigation and Information Systems Conference, Seattle, WA, July 1995, pp 328–333

83. Middelham F (1999) A synthetic study of the network effects of ramp metering. Technical report. Transport Research Centre (AVV), Dutch Ministry of Transport, Public Works and Water Management, Rotterdam

84. Middelham F (2003) State of practice in dynamic traffic management in The Netherlands. In: Proceedings of the 10th IFAC Symposium on Control in Transportation Systems (CTS 2003), Tokyo, Japan, August 2003

85. Middleton D, Gopalakrishna D, Raman M (2002) Advances in traffic data collection and management. Technical Report BAT-02-006. Texas Transportation Institute, Cambridge Systematics Inc., Washington. http://www.itsdocs.fhwa.dot.gov/JPODOCS/REPTS_TE/13766.html. Accessed on 23 July 2008

86. Molina M, Hernández J, Cuena J (1998) A structure of problem-solving methods for real-time decision support in traffic control. Int J Human-Computer Stud 49:577–600

87. Papageorgiou M (1980) A new approach to time-of-day control based on a dynamic freeway traffic model. Transp Res Part B, 14B:349–360

88. Papageorgiou M (1983) A hierarchical control system for freeway traffic. Transp Res Part B, 17B(3):251–261

89. Papageorgiou M (1998) Some remarks on macroscopic traffic flow modelling. Transp Res Part A 32(5):323–329

90. Papageorgiou M (ed) (1991) Concise Encyclopedia of Traffic & Transportation Systems. Pergamon

91. Papageorgiou M, Kotsialos A (2000) Freeway ramp metering: An overview. In: Proceedings of the 3rd Annual IEEE Conference on Intelligent Transportation Systems (ITSC 2000), Dearborn, Michigan, USA, October 2000, pp 228–239

92. Papageorgiou M, Kotsialos A (2002) Freeway ramp metering: An overview. IEEE Trans Intell Transp Syst 3(4):271–280

93. Papageorgiou M, Messmer A (1991) Dynamic network traf-

fic assignment and route guidance via feedback regulation. Transp Res Rec 1306:49–58

94. Papageorgiou M, Blosseville J-M, Hadj-Salem H (1990) Modelling and real-time control of traffic flow on the southern part of Boulevard Périphérique in Paris: Part I: Modelling. Transp Res Part A, 24A(5):345–359

95. Papageorgiou M, Blosseville J-M, Hadj-Salem H (1990) Modelling and real-time control of traffic flow on the southern part of Boulevard Périphérique in Paris: Part II: Coordinated on-ramp metering. Transp Res Part A, 24A(5):361–370

96. Papageorgiou M, Hadj-Salem H, Blosseville J-M (1991) ALINEA: A local feedback control law for on-ramp metering. Transp Res Rec (1320):58–64

97. Papageorgiou M, Hadj-Salem H, Middelham F (1997) ALINEA local ramp metering – summary of field results. Transp Res Rec 1603:90–98

98. Papageorgiou M, Wang Y, Kosmatopoulos E, Papamichail I (2007) ALINEA maximizes motorway throughput – An answer to flawed criticism. Traffic Eng Control 48(6):271–276

99. Payne HJ (1971) Models of freeway traffic and control. Simul Counc Proc 1:51–61

100. Rees I (1995) Orbital decongestant. Highways 63(5):17–18

101. Rees T, Harbord B, Dixon C, Abou-Rhame N (2004) Speed-control and incident detection on the M25 controlled motorway (summary of results 1995–2002). Technical Report PPR033, TRL (UK's Transport Research Laboratory), Wokingham

102. Richards PI (1956) Shock waves on the highway. Oper Res 4:42–51

103. Ritchie SG (1990) A knowledge-based decision support architecture for advanced traffic management. Transp Res Part A, 24A(1):27–37

104. Schik P (2003) Einfluss von Streckenbeeinflussungsanlagen auf die Kapazität von Autobahnabschnitten sowie die Stabilität des Verkehrsflusses. Ph.D. Thesis, Universität Stuttgart

105. Shladover SE, Desoer CA, Hedrick JK, Tomizuka M, Walrand J, Zhang WB, McMahon DH, Peng H, Sheikholesham S, McKeown N (1991) Automatic vehicle control developments in the PATH program. IEEE Trans Veh Technol 40(1):114–130

106. Sisiopiku VP (2001) Variable speed control: Technologies and practice. In: Proceedings of the 11th Annual Meeting of ITS America, Miami, pp 1–11

107. Slotine JE, Li W (1991) Applied Nonlinear Control. Prentice Hall, New Jersey

108. Smith BL, Zhang H, Fontaine MD, Green MW (2004) Wireless location technology based traffic monitoring: Critical assessment and evaluation of an early-generation system. J Transp Eng 130(5):576–584

109. Smulders S (1990) Control of freeway traffic flow by variable speed signs. Transp Res Part B, 24B(2):111–132

110. Smulders SA, Helleman DE (1998) Variable speed control: State-of-the-art and synthesis. In: Road Transport Information and Control, number 454 in Conference Publication, IEE, 21–23 April 1998, pp 155–159

111. Soriguera F, Thorson L, Robuste F (2007) Travel time measurement using toll infrastructure. In: Proceedings of the 86th Annual Meeting of the Transportation Research Board, CDROM paper 07-1389.pdf

112. Taylor CJ, Young PC, Chotai A, Whittaker J (1998) Nonminimal state space approach to multivariable ramp metering control of motorway bottlenecks. IEE Proc – Control Theory Appl 146(6):568–574

113. Taylor MAP (1999) Dense network traffic models, travel time reliability and traffic management. I: General introduction. J Adv Transp 33(2):218–233

114. Taylor MAP (1999) Dense network traffic models, travel time reliability and traffic management. II: Application to network reliability. J Adv Transp 33(2):235–251

115. Taylor MAP, Woolley JE, Zito R (2000) Integration of the global positioning system and geographical information systems for traffic congestion studies. Transp Res Part C, 8C:257–285

116. Transportation Research Board (1998) Managing speed: Review of current practice for setting and enforcing speed limits. In: TRB Special Report 254, Transportation Research Board, Committee for Guidance on Setting and Enforcing Speed Limits. National Academy Press, Washington

117. Treiber M, Hennecke A, Helbing D (2000) Congested traffic states in empirical observations and microscopic simulation. Phys Rev E 62:1805–1824

118. Tsugawa S, Kato S, Tokuda K, Matsui T, Fujii H (2000) An architecture for cooperative driving of automated vehicles. In: Proceedings of the IEEE Intelligent Transportation Symposium, Dearborn, MI, USA 2000, pp 422–427

119. Vahidi A, Eskandarian A (2003) Research advances in intelligent collision avoidance and adaptive cruise control. IEEE Trans Intell Transp Syst 4(3):143–153

120. van Arem B, van Driel CJG, Visser R (2006) The impacts of cooperative adaptive cruise control on traffic-flow characteristics. IEEE Trans Intell Transp Syst (4):429–436

121. van den Berg M, Bellemans T, De Schutter B, De Moor B, Hellendoorn J (2005) Control of traffic with anticipative ramp metering. In: Proceedings of the 84th Annual Meeting of the Transportation Research Board, Washington DC, 2005 CDROM paper 05-0252, pp 166–174

122. van den Hoogen E, Smulders S (1994) Control by variable speed signs: Results of the Dutch experiment. In: Proceedings of the 7th International Conference on Road Traffic Monitoring and Control, IEE Conference Publication No. 391, London, England, 26–28 April 1994, pp 145–149

123. Varaiya P (1993) Smart cars on smart roads: Problems of control. IEEE Trans Autom Control 38(2):195–207

124. Wang Y, Papageorgiou M (2005) Real-time freeway traffic state estimation based on extended Kalman filter: A general approach. Transp Res Part B 39(2):141–167

125. Wang Y, Papageorgiou M, Messmer A (2003) A predictive feedback routing control strategy for freeway network traffic. In: Proceedings of the 82nd Annual Meeting of the Transportation Research Board, Washington DC, January 2003

126. Wang Y, Papageorgiou M, Messmer A (2006) RENAISSANCE A unified macroscopic model-based approach to real-time freeway network traffic surveillance. Transp Res Part C, 14C:190–212

127. Wattleworth JA (1965) Peak-period analysis and control of a freeway system. Highw Res Rec 157:1–21

128. Wilkie JK (1997) Using variable speed limit signs to mitigate speed differentials upstream of reduced flow locations. Technical report. Department of Civil Engineering, Texas A&M University, College Station, Prepared for CVEN 677 Advanced Surface Transportation Systems, Report No. SWUTC/97/72840-00003-2

129. Williams B (1996) Highway control. IEE Rev 42(5):191–194

130. Wolf DE (1999) Cellular automata for traffic simulations. Physica A 263:438–451

131. Zhang H, Ritchie SG (1994) Real-time decision-support system for freeway management and control. J Comput Civ Eng 8(1):35–51

132. Zhang L (2007) Traffic diversion effect of ramp metering at the individual and system levels. In: Proceedings of the 86th Annual Meeting of the Transportation Research Board, Washington DC, January 2007, CDROM paper 07-2087

Books and Reviews

We refer the interested reader to the following references
in the various fields that have been discussed in this chapter:

Control: General introduction [4], optimal control [73], model predictive control [15,79], nonlinear control [107],

Traffic flow modeling: General overviews [48,49], cell transmission model [26], Kerner's three-phase theory [60], microscopic simulation models [2], cellular automata [80],

Ramp metering: Overviews of ramp metering strategies [18,91], field test and simulation studies [39],

Speed limit systems: Overviews of practical speed limit systems [106,128],

Intelligent vehicles: Overview [10],

Sensor technologies: Overviews [32,33,63,64]

Functional Genomics for Characterization of Genome Sequences

ADAM M. DEUTSCHBAUER[1], LARS M. STEINMETZ[2]
[1] Lawrence Berkeley National Laboratory, Berkeley, USA
[2] European Molecular Biology Laboratory, Heidelberg, Germany

Article Outline

Glossary

Tiling microarray A microarray containing probes representing the entire genome in an unbiased, uniform pattern. The resolution of a tiling microarray is determined by the length of probes and the amount of overlap between adjacent probes to the genome.

Non-coding RNA (ncRNA) RNA that is not translated into protein. The term ncRNA encompasses different classes of RNA including small nucleolar RNA (snoRNA), small inhibitory RNA (siRNA), ribosomal RNA (rRNA), transfer RNA (tRNA) and microRNA (miRNA).

Pyrosequencing A DNA sequencing method that utilizes enzymatic reactions and light detection as a readout for base incorporation.

Complex trait A heritable trait conditioned by multiple genetic and/or environmental factors.

RNA interference (RNAi) The process by which double strand RNA (dsRNA) targets and degrades a complementary transcript.

Epigenetics A class of heritable traits that are stable over multiple cell divisions but are not associated with DNA sequence changes.

Phylogenetics The study of evolutionary relationships among organisms.

ChIP-chip Shorthand terminology for chromatin-immunoprecipitation with microarray detection; a technique to identify regions of DNA bound by a protein of interest. Following immunoprecipitation of a protein-DNA hybrid, the DNA is released from the protein, labeled, and detected on a DNA microarray.

Genotyping The process of measuring the genetic differences (genotype) between individuals within a population.

Linkage disequilibrium The non-random association of two or more alleles caused primarily by population structure and the absence of recombination in a region of the chromosome.

Epistasis A type of genetic interaction between two or more loci in which the phenotype of the mutant combination deviates from the expectation based on the phenotype of the single mutants.

Chromatin The tightly bundled complex of DNA and nucleosomes (made up of histone proteins) that packages the nuclear genome in eukaryotic cells.

Definition of the Subject

Complete genome sequences have been determined for hundreds of organisms ranging in complexity from bacteria to human. However, sequence data alone is currently of limited use for identifying the functional elements of a genome and for elucidating how these elements interact to control physiological processes.

The field of functional genomics aims to meet these challenges using the sequence data as a blueprint. In a broad sense, functional genomics is defined as the large-scale experimental study of gene function and interactions. In this article, we introduce the reader to the techniques and challenges in functional genomics in the context of an ever more common case scenario: given a complete genome sequence, how does one figure out what the sequence means? By highlighting the relevant literature, we go through the current of the art as to what one would do to arrive at function given a genome sequence. We focus primarily on nucleic acid based technologies though connect in the Future Directions to similar analyzes at the protein and metabolite level.

Introduction

The field of functional genomics is relatively new compared to other molecular biology disciplines such as biochemistry, classical genetics, and cell biology. The technological advances and milestones that laid the groundwork for the completion of genome sequences have occurred primarily over the last 25 years (Fig. 1). In particular, advances related to DNA sequencing deserve special attention as sequence data is the backbone of functional genomics and the ability to generate large amounts of this data has paved the way for many of the techniques described in this article. Foremost among these advances were the development of chain termination based DNA sequencing by Sanger and colleagues [99] and the automation of DNA sequencing by Hood and coworkers [109]. The ability to automate sequencing ushered in a new era of biology, one in which the completion of genomes was technically feasible.

The drive to sequence complete genomes was initiated in the late 1980's with the establishment of the Human Genome Project. However, given that the size of the human genome is immense at 3 billion nucleotides and the optimal techniques for completing a genome sequence were unknown at the time, it was necessary to develop whole genome sequencing in model organisms. In 1995, the first genome of a free living organism, the bacterium *Haemophilus influenzae*, was published [35]. In subsequent years, the genomes of the model genetic organisms *Saccharomyces cerevisiae* (yeast) [45], *Caenorhabditis elegans* (worm) (*C. elegans* Sequencing Consortium 1998), *Drosophila melanogaster* (fruit fly) [2], and mouse [123] were completed by international consortia. Paramount to the success of these projects was the development of automation and robotics for handling large numbers of reactions and software for the analysis and assembly of vast amounts of sequence data. The lessons learned by sequencing model organisms laid the technological foundation for the completion of the draft human sequence
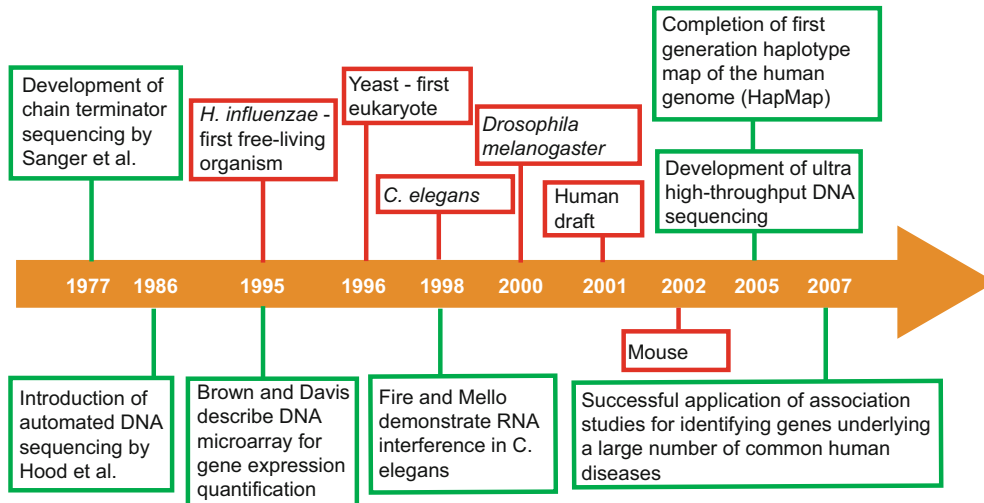
Functional Genomics for Characterization of Genome Sequences, Figure 1
Landmarks in functional genomics. The timeline indicates select technological breakthroughs (in *green boxes*) in functional genomics and highlights the genome completion date of important model organisms (in *red boxes*). References are in parentheses

in 2001 [73]. In some respects, the finishing of the human genome marked the end of the first generation of DNA sequencing. Today, advances in sequencing technology promise a second generation of DNA sequencing, one in which the majority of the earth's living (and some extinct) organisms will have complete genome sequences in the near future (see Future Directions).

The initial analysis of the early genomes was greeted with two harsh realizations. First, current algorithms for gene finding, partly based on the availability of many complete genome sequences, were not yet available. The result was that many predicted genes were later demonstrated not to be genes and conversely, many currently predicted genes were missed by these early, somewhat arbitrary gene prediction methods. Second, the majority of predicted genes identified from the sequence alone had no known function. For example, over 60% of the ∼6000 yeast genes identified from the genome sequence were uncharacterized in 1996 despite 20 years of active molecular genetics research [45]. To fill this void, higher-throughput, so called "post-genomic" techniques aimed at the elucidation of gene function at a genome-scale were developed. The subsequent application of these post-genomic techniques to model organisms marked the beginning of modern functional genomics. Today, functional genomic tools such as microarrays for gene expression [103] and whole-genome reverse genetics have increased the percentage of characterized yeast genes to over 80% [92].

Here we describe some major topics in functional genomics in the context of how one would analyze a newly completed genome sequence today. We begin with a description of *in silico* comparative genomic approaches for finding genes, assigning putative gene function, and inferring regulatory elements. We then describe microarray-based gene expression profiling and high-throughput mutant analysis before ending with a discussion of intraspecific genetic variation. In each instance we highlight recent literature to illustrate the utility and shortcomings of different techniques. To avoid organism-bias and to reflect the universal challenges associated with assigning gene function based on DNA sequence data, we discuss literature from a range of organisms, spanning bacteria to human. Finally, in the Future Directions, we link our discussion of functional genomics to future challenges including data integration, genetic interactions, and the impact of technology.

Computational and Comparative Genomics

A newly completed genome sequence is simply a file containing millions (or even billions) of A, T, C, and G nucleotides. Taken in isolation, the task of identifying functional elements from this file of nucleotides can seem like an insurmountable problem. While the field of functional genomics aims to uncover the functionality of the genome through experiment, one can learn a great deal about a new genome sequence using purely computational tools. Inherent in the development, application, and success of computational genomic analyzes are the wealth of already completed genome sequences with

which one can compare a genome sequence for similarities and differences (http://www.ncbi.nlm.nih.gov/sites/entrez?db=Genome). This field of research, comparative genomics, has two broad aims: to identify functional elements such as genes and regulatory sequences and to infer the evolutionary mechanisms that gave rise to different species [51]. Here we concentrate on computational and comparative genomics in the context of finding functional DNA elements and direct the reader to additional reading on the use of comparative genomics to infer evolutionary relationships in prokaryotes [1], yeast [26], *Drosophila* [17], and human [14].

Concepts

The premise of comparative genomics is relatively straightforward. DNA sequences that are shared between two distinct species are likely to be functional. From an evolutionary perspective, these are genomic regions that are conserved since the species' last common ancestor and represent genomic regions that are more intolerant to mutation. Conversely, since mutations are assumed to accumulate in genomic regions that are under no functional constraint, sequences that are significantly diverged since the last common ancestor represent nonfunctional regions. In practice, identifying conserved (functional) versus nonconserved (nonfunctional) genomic regions is a steep challenge due to a myriad of complicating factors including non-uniform mutation rates both within [125] and between [72] genomes and the fact that the premise is not always true; even highly diverged regions between species can retain similar functions [34]. Furthermore, the success of a comparative genomic analysis is dependent on the marriage of the right biological question and the use of an appropriate phylogenetic range of organisms. For instance, one would have a difficult time identifying conserved gene regulatory elements by comparing worm and human orthologous genes. Worm and human diverged too long ago to make such an analysis fruitful. Conversely, in order to detect positively selected genes in the human lineage a comparison between human and other closely related primate genomes is most appropriate (for example [29]).

Applications

The protein-coding gene is a principle functional unit of the genome. Therefore, computationally predicting genes is usually the first task in any genome-level analysis. While software exists to predict genes using single genomes and gene structure modeling, usually a comparative approach to gene finding is utilized based on the knowl-

edge that genes are more likely to be conserved in related species [82] . In the yeast *Saccharomyces cerevisiae*, the initial prediction of the number of genes was over 6000 [45]. By comparing the genome sequence of *S. cerevisiae* to that of other related *Saccharomyces* species, an additional ∼50 small protein-coding genes were predicted to exist based on high sequence conservation while ∼500 of the originally annotated genes were deemed not to be protein-coding genes based on frameshift and nonsense mutations in the orthologs of closely related species [18,68]. Similar approaches are being applied in mammals, where high-genome complexity and extensive gene splicing make the task of gene identification challenging. In one example, gene prediction software [70] based on the sequence conservation between genes was used to predict over 1000 genes in mouse and human [48].

There is growing appreciation for the important regulatory and structural functions of non-coding RNA (ncRNA). Unfortunately, the ability to computationally identify ncRNA is complicated by the absence of defined sequence signatures similar to what is available for protein-coding genes [27]. However, if ncRNA are functionally relevant, then they should be under evolutionary constraint and conserved at least structurally across related species. Using this rationale, programs to detect ncRNA have been developed using comparisons across species [95]. For instance, an algorithm that combines RNA structure prediction and comparative sequence analysis identified both known and novel conserved ncRNA when applied to the genomes of human, mouse, rat, zebrafish, and pufferfish [122].

Genes by themselves are static stretches of DNA which encode information on the chromosome. It is the regulation of these genes that guides development, maintains homeostasis, and contributes significantly to the diversity in nature. Therefore, the identification of the non-coding sequences that transcription factors bind to regulate gene expression is an active area of research. However, the challenges associated with computationally identifying functional regulatory elements are substantial. Most importantly, the DNA sequences that regulate transcription are often small and degenerate. Therefore, the identification of these elements is often hampered by a large number of false positive results. Even more critical than for identifying protein coding regions, identifying regulatory elements requires comparative genomics across an appropriate range of phylogenetic distances. For example, a search for common regulatory motifs in humans via phylogenetic comparisons with orthologous sequences in mouse, rat, and dog identified 105 previously unknown promoter regulatory motifs and 106 motifs present in 3' untranslated re-

gions [126]. As more genome sequences become available, the 'phylogenetic footprinting' approach for identifying regulatory elements should increase in power. In a broad sense, the promise of all comparative genomics techniques as described here is dependent on more genome sequences than are currently available. This is a principle reason why genome sequencing will stay at the forefront of genetics research in the near future.

Summary

What have we learned about our genome using computational and comparative genomics? Without setting foot in the laboratory, we have successfully predicted many protein-coding genes, ncRNAs, and regulatory elements, via comparison with previously sequenced genomes. Furthermore, using sequence similarity tools such as BLAST, we could assign putative biochemical and biological functions to many genes by homology to genes of known function. However, despite this progress, the majority of genes in most newly sequenced species remain partially or completely uncharacterized at a functional level. It must be emphasized that computational predictions are dependent on the availability of data and thus can have limited power (for example, see [28] for review of computational prediction versus experimental validation of transcription factor binding sites). Conservation of DNA across an evolutionary distance does not necessarily mean that the sequences are functional [86] and conversely, the lack of sequence conservation does not necessarily imply a lack of function [34]. Therefore, computational predictions must be interpreted carefully. Finally, simply identifying the locations of putative functional elements on the chromosome provides little information on the dynamic functionality of the genome. Experimental functional genomic techniques aim to fill these voids. In the remaining sections we will describe analyzes at different levels with the goal of enriching this view of genome function.

Transcription

Cells regulate the transcription of RNA from genes to control biological processes. Therefore, a thorough understanding of transcriptional architecture is a chief requirement for the elucidation of organism physiology. Taken together, the transcription of a single gene is a highly complicated process involving genomic DNA, transcription factors, RNA polymerases, and epigenetic modifications such as chromatin dynamics. The functional genomics of transcription aims to analyze each of these phenomena in the context of the entire genome. Here we discuss whole-genome techniques to dissect transcription and their ap-

plications to transcript identification, RNA quantification, gene regulation by transcription factor binding localization, and epigenetics.

Before entering into the functional genomics of gene transcription, it is necessary to briefly introduce the primary technology used in the majority of these studies: the microarray. The microarray is a surface (typically glass) with thousands or even millions of unique immobilized molecules (most commonly DNA) systematically arrayed as 'probes'. DNA microarrays come in two primary types: 'spotted' arrays printed with PCR products (or long oligonucleotides) and 'oligo' arrays with short (usually 25 to 70 nucleotide) *in situ* synthesized probes. Regardless of their construction, one hybridizes a complex mixture of usually fluorescently labeled target nucleic acid to the microarray. The abundance of each nucleic acid species in the target sample is then monitored by the fluorescent intensity of the corresponding probe on the microarray.

Transcript Mapping Using Tiling Microarrays

The identification of transcripts is a logical starting point for any global study of transcription. Using unbiased whole-genome tiling arrays that interrogate every base of the genome, one can get a global view of the transcriptional architecture of an organism [7]. The application of this technique to various organisms including *Arabidopsis* [127], yeast [22], and human [30] has revealed a number of fundamental insights both at the global and single-gene level. At the single-gene level, genome tiling arrays provide a high-resolution view of the expression architecture of individual genes (see Fig. 2 for examples). In particular, the start/stop sites and exon/intron boundaries of individual transcripts can be localized to within a handful of base pairs. Importantly, these insights enable the reannotation of the original computational predictions of gene structure [22,132] and demonstrate how experimental methods can be used to systematically update original genome-wide predictions based on computation.

More importantly than the insights into single genes, tiling arrays are revolutionizing our global perspective on transcription [67]. First, there exists considerably more transcription than had previously been thought. Transcriptional elements (e.g. genes) are not arranged as beads-on-a-string along the genome. Strikingly, it appears that nearly every base in the yeast genome is transcribed, often from both strands [22]. In humans, 93% of genomic regions examined in detail (encompassing 1% of the total human genome) appear to be transcribed in an least one condition or tissue [30]. Second, in all organisms examined with tiling arrays to date, the vast major-
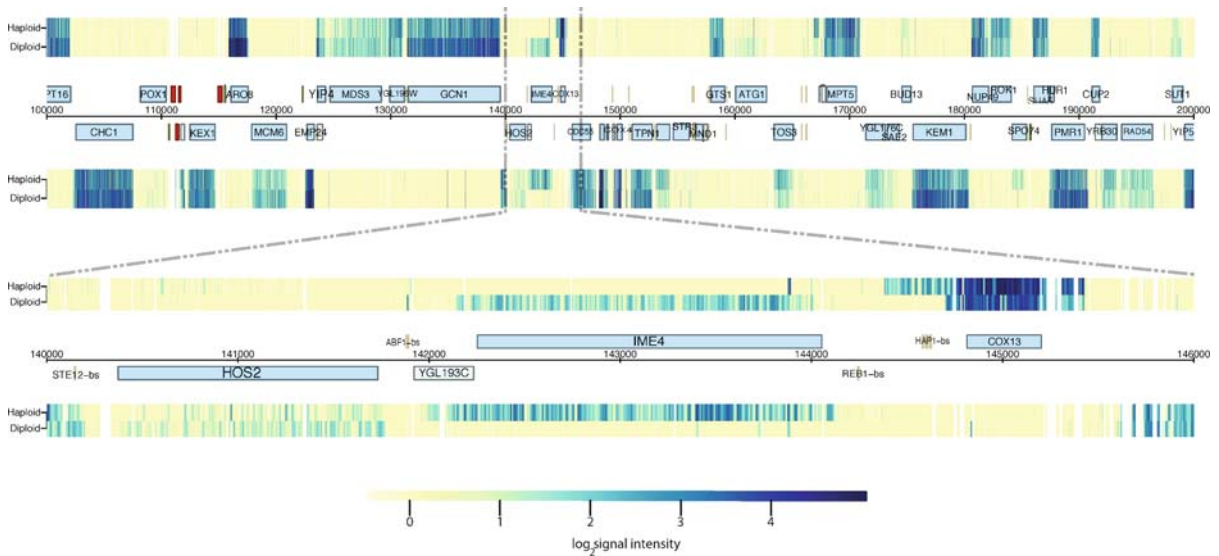
Functional Genomics for Characterization of Genome Sequences, Figure 2
Transcript mapping using tiling microarrays. The top panel shows data for a 100 kb region from a diploid and a haploid yeast strain. Probes on the array are 25 bases long and are tiled every 8 bases on each strand of the genome, the forward strand shown on top and the reverse strand on the bottom. Gene features, such as open reading frames, as annotated in the yeast genome database are shown by boxes. Expression signal is plotted with a color gradient for probes across the region and allows expressed transcripts to be detected. The bottom panel shows a close-up of the region around *IME4*. The Ime4 protein is expressed only in diploids cells. In haploids, an *IME4* antisense transcript inhibits the expression of Ime4 protein by transcriptional interference [56]

ity of this novel transcription does not appear to encode protein, rather the transcription is likely to be ncRNA. Often this ncRNA is in the form of antisense transcription that either physically overlaps the transcription of genes on the opposite strand (the sense strand) forming *cis*-encoded sense/antisense transcript pairs, or matches to sequence transcribed elsewhere in the genome forming *trans*-encoded sense/antisense pairs . Lastly, in humans, primary transcripts appear to be extensively overlapping (or 'interleaved') which may provide a mechanism for increasing the number of possible human proteins [30]. Overall, these findings imply the potential inhibitory function of antisense transcription, the regulatory roles of small ncRNA, and challenge the view that each human gene is its own modular entity.

There are several clues that suggest that natural antisense transcription (in *cis* or in *trans*) is an important mode of gene regulation. First, genome-wide assessments demonstrate that antisense transcription is prevalent in most (if not all) organisms. Second, complementary sense and antisense transcripts are often conserved across evolution [16]. Given the appreciation for the extent of antisense transcription, a challenge today is to decipher the mechanism(s) through which antisense transcripts contribute to gene regulation [74]. One possibility is that that two

transcripts encoded in *cis* with opposite orientations collide during RNA polymerase-based elongation [94]. This model provides a mechanism to modulate gene expression from the sense strand by altering the relative strengths of the two promoters. Such a transcriptional interference mechanism has recently been proposed to contribute to the entry of yeast into sporulation [56] (See Fig. 2 for details). Alternatively, antisense transcription might exert its effect through double-stranded RNA pathways such as RNAi [9]. Often these short *trans* encoded antisense transcripts are found to be expressed in alternate tissues and developmental stages than their sense targets, suggesting that some miRNAs may help maintain and define cell types by dampening the expression of unwanted transcripts [13].

The unbiased view of transcription derived from tiling microarrays has profound implications. First, the notion that the non protein-coding portion of genomes is 'junk' requires reassessment. The extent of (and presumably regulation of) transcription in these regions suggests that they serve a functional role. Second, many mechanisms of gene regulation are likely to exist. In addition to the antisense transcription described above, other modes of regulation including complex intercalated transcription, chimeric transcripts, and the action of microRNAs on

chromatin for example underscore the complexity of transcriptional architecture. Lastly, it has been proposed that the aberrant lack of correlation between organism complexity and gene number can be partially reconciled with an increase in ncRNA regulation in more complex organisms [114]. This view contends that an expansion of *trans*-acting regulatory ncRNA (together with an increase in *cis*-acting regulatory DNA) provides a substantial increase in genome information content.

Gene Expression (mRNA Quantification)

The most common functional genomic experiment is the (relative) quantification of mRNA levels using microarrays, typically referred to as gene expression or 'transcriptome' profiling. The reasons for this are twofold. The first is that this type of experiment is relatively straightforward to perform. All one needs is a microarray (many of which can be purchased commercially), the infrastructure to process the microarray, and an RNA sample. The second reason why gene expression experiments are popular is that they can provide substantial insights into transcription and its regulation, particularly when a large number of experiments are performed (the rationale is identical to the increasing power of comparative genomics with more complete genome sequences).

What can gene expression experiments tell us about the function of the genome? First, gene expression experiments can be used to infer the functionality of unknown genes under the hypothesis that genes with similar expression patterns are more likely to have similar functions [11]. Likewise, mutants with similar expression pattern are likely mutant for genes of similar function [57]. Second, gene expression data are useful for identifying regulatory sites that control transcription. This is accomplished by a computational analysis of the genomic region surrounding a set of genes identified as co-regulated. Lastly, gene expression experiments can provide insight into the gene regulatory network of an organism. For example, an expression compendium of over 100 diverse conditions in *E. coli* was sufficient to identify 1079 regulatory interactions including a novel relationship between iron transport and central carbon metabolism [31].

Gene Regulation

Gene expression data and comparative genomic approaches are insufficient in isolation to identify all of the genomic regulatory elements that control transcription. To identify these elements, multiple experiment-based approaches have been developed. The in vivo ChIP-chip method involves the cross-linking of a transcription factor

of interest (often epitope-tagged) to the DNA. Following immunoprecipitation of the transcription factor-DNA hybrid, the DNA is released from the protein, labeled, and detected on a DNA microarray. The consensus sequence bound by the transcription factor is identified by a comparative sequence analysis of all 'hits' on the microarray. In one example, ChIP-chip was used to associate the binding of the Rap1 transcription factor to 294 loci in the yeast genome, the majority of which were intergenic regions with a high probability of containing promoters [78]. On a more global level, Harbison et al. analyzed the genomic binding of 203 yeast transcription factors in rich media and across a range of diverse conditions [50]. The results of this work provided the first comprehensive regulatory map for a eukaryotic genome and demonstrated the dynamic nature of transcriptional regulation in the response to environmental change.

There are a number of design and analysis considerations that go into a successful ChIP-chip experiment [12]. One parameter of note is the density of the microarray used in the experiment. In general, a tiling microarray is superior to a conventional microarray because all genome coordinates are covered, and multiple, overlapping probes can interrogate a single genomic region. The principle advantage is that an increase in microarray signal across a number of overlapping probes provides both increased statistical significance and an increased ability to identify the actual binding sequences. However, even with tiling microarrays it is often difficult to determine the cutoff significance value for which microarray probes are enriched in the experiment (and hence, which genomic locations the transcription factor binds to). Another disadvantage of ChIP-chip is that tiling microarrays for large genomes are currently expensive (because multiple microarrays are necessary to adequately tile large genomes). Therefore, other approaches for the detection of DNA from chromatin immunoprecipitation experiments have been explored. One promising avenue is the use of massively parallel sequencing technology (see Future Directions) for the detection of DNA (ChIP-seq). ChIP-seq requires no microarray and provides a direct count of DNA sequences contained in the immunoprecipitated sample. Results obtained using human STAT1 [96] and NRSF [64] indicate that ChIP-seq can detect weakly bound sites that are likely to be missed by microarray based methods.

A promoter array is an in vitro method for identifying transcription factor binding sites [5]. Here, an oligonucleotide microarray with double stranded probes is directly bound by a labeled target protein. A primary advantage of a promoter array over ChIP-chip or ChIP-seq is that all combinations of short sequences (for example,

10mers in [5]) can be assayed for binding associations of different affinities. Additionally, the actual identification of the binding site is more straightforward than the computational analysis required for ChIP-chip or ChIP-seq experiments. Conversely, as an entirely in vitro assay, some positive promoter array results may be physiologically irrelevant.

Epigenetics

In the context of transcription, the regulation of chromatin represents the primary mode of epigenetic inheritance. In eukaryotic genomes, DNA ins packaged into chromatin, which consists of DNA wound around protein (histone) complexes known as nucleosomes. The positioning and regulation (via the modification of their histone subunits) of nucleosomes along the chromosome play a crucial role in gene expression, by 'exposing' regions of the genome for transcriptional activity. Given its importance to gene regulation, the localization and dynamics of nucleosome positioning are an active area of investigation. An early genome-level investigation of nucleosome positions showed that active promoters and their associated transcription factor binding sites are largely devoid of nucleosomes [129]. A second study demonstrated that, during the cell cycle, the occupancy of nucleosomes is reduced at the specific time when the gene is transcribed [54]. These results indicate that chromosome accessibility via nucleosome positioning is an important mechanism for regulating the expression state of genes. Using nucleosome localization data, Segal and colleagues developed a nucleosome-DNA interaction model suggesting that nucleosome positioning is an intrinsic feature of eukaryotic genomes [106]. Interestingly, this model takes advantage of the tendency of DNA to bend sharply around the nucleosome at periodic intervals. More recently, tiling micoarrays were used to construct a comprehensive, high-resolution "atlas" of over 70,000 nucleosome sites in the yeast genome [77].

In addition to the physical localization of nucleosomes, the modification state of the individual histone subunits is thought to be linked to the expression state of the cell. To test the hypothesis that histone methylation marks the expression state of differentiated cells in mammals, Mikkelsen et al. analyzed the genomic pattern of histone methylation in mouse pluripotent cells (capable of differentiating into multiple lineages) and terminal differentiated cells using ChIP-seq [83]. The authors demonstrate that specific histone methylations, such as those on lysine positions 4 and 27 of histone subunit H3, discriminate the expression state of genes and hence provide insight

into the lineage potential of different cell types. It is anticipated that future genome-wide epigenetic investigations of the kind described above will provide additional insight into the importance of epigenetics in gene regulation and disease.

Summary

Given that microarray design and construction requires only DNA sequence data, we can learn a lot about the architecture and regulation of transcription of our genome using the approaches described above. However, there are limitations to using transcription data alone to dissect the functionality of an entire genome. Most importantly, neither the presence of transcription, the expression level of genes, nor the identification of regulatory elements tell us much about what biological processes genes are actually involved in. For instance, the majority of genes that are differentially expressed under stress conditions in yeast display no obvious phenotype under the same condition when disrupted [42]. Taking into consideration these limitations, additional genome-wide approaches are necessary to more directly elucidate gene function.

Genetics Analysis

The field of genetics is largely the analysis of mutations. By studying the phenotypic effects of mutations, one can gain insight into the biological processes to which a gene contributes. Classically, in the absence of complete genome sequences, geneticists had to map interesting mutations to the causative gene. While this 'forward' genetic approach is still relevant, today more often mutagenesis approaches are being applied that take advantage of the complete genome sequence of an organism. In this article, we will primarily focus on these 'reverse' genetic techniques.

The mutations discussed below come in two main flavors: heritable and transient. Heritable mutations are permanent changes to the DNA sequence and are induced using techniques such as targeted gene 'knockouts' by homologous recombination and random mutagenesis by a mobile element (such as a transposon). Transient techniques are largely based on the ability of certain molecules to inhibit the expression of specific genes and consequently 'turning them down' for as long as the molecules are active. The most important of these techniques is RNA interference (RNAi). In RNAi, a short double strand RNA molecule degrades its specific complementary transcript. Therefore, RNAi mimics a loss-of-function mutation. The term 'transient' here can be misleading as the effects of the mutagenesis can last for multiple cell divisions. However, because the genomic DNA sequence is left untouched, the

effects of RNAi are not permanent. Additionally, transient perturbations can be introduced via small molecule perturbations, a field referred to as "chemical genomics". In yeast, chemical genomics has proved successful in identifying the targets of small molecules and the mode of action of small molecules[43,90].

Systematic Analysis of Single-Gene Mutations

Essential genes are a special class whose disruption results in a loss of organismal viability. In a number of microorganisms, comprehensive surveys have been undertaken to determine the catalog of essential genes in order to determine which core functions the cell needs to survive [44] and to identify potential drug targets [98]. Approximately 300 genes are essential in the genomes of the bacteria *E. coli* [3] and *B. subtilis* [69] (less than 10% of the total gene complement in each organism). Of these ∼300 genes, about one-half are orthologous (derived from a common ancestor) between these distantly related bacteria demonstrating the fundamental importance of these genes across evolution. In yeast, ∼18% of the genome is essential [42]. Reflecting the critical role of these genes in fundamental eukaryotic cellular processes, the essential gene class is more likely to have homologs in other organisms than the nonessential gene class [42].

As described above, only a small fraction of the total genome complement is essential for viability. For the nonessential gene class, the systematic determination of mutant phenotypes provides insight into all of the biological processes intrinsic to the organism, not just viability. For a large number of microorganisms, genome-wide approaches have been applied to determine which genes are 'conditionally essential' (required and presumably functional under specific conditions). In bacteria, techniques for the parallel analysis of mutant pools such as signature tagged mutagenesis [52] and transposon site hybridization with microarrays [100] are successful at identifying bacterial genes required for pathogenesis in a host. The conditionally essential gene sets identified from these studies represent attractive targets for novel antibiotics.

In 2002, an international consortium of yeast researchers completed a systematic knockout library of the entire yeast genome [42]. An important aspect to this project was the incorporation of unique molecular barcodes into each deletion strain that enabled pooling and analysis of the entire collection in a single microarray hybridization. Since its completion, the yeast knockout library has been profiled in hundreds of conditions both in pools with microarrays and as individual strains using miniaturized growth assays on plates and in liquid media [104]. The results of these studies have provided tantalizing clues into the putative function of hundreds of uncharacterized yeast genes. For example, an unbiased survey of genes required for sporulation identified virtually all known regulators of the process in addition to hundreds of genes previously unknown to be involved in sporulation [25]. In addition, the evolutionary conservation of many genes permits the reinterpretation of the yeast results across all eukaryotes, including human. For instance, a comprehensive screen for yeast mitochondrial genes was used to identify candidate human disease genes based on sequence homology and disease linkage map intervals [111].

The systematic examination of mutant phenotypes in multicellular eukaryotes is more difficult and costly than that in microorganisms. Using RNAi, however, an increasing number of genome-wide mutagenesis studies have demonstrated the utility and power of reverse genetics in worm [65], fly [37], and human [6,88]. The worm, where the capability of dsRNA to silence complementary transcripts was first demonstrated in 1998 [33], led the genome-wide RNAi initiative due to the early completion of its genome sequence and the ease with which dsRNA can be delivered to the whole organism (by feeding). By targeting 86% of the ∼19,000 worm genes by RNAi, Kamath and colleagues identified mutant phenotypes for 1722 genes, two-thirds of which had not been previously associated with a phenotype [65].

In contrast to worm RNAi which is performed against the entire organism, genome-wide RNAi efforts in Drosophila and human have been applied to cultured cells, primarily with well known signaling pathways as the focus. A number of lessons emerge from these studies. First, the number of unknown regulators of 'well' characterized pathways is large. For example, Friedman and Perrimon identified hundreds of potential regulators of the receptor tyrosine kinase/extracellular regulated kinase (RTK/ERK) pathway using a quantitative RNAi screen in Drosophila [37]. Second, the limitations of RNAi targeting efficiency in humans can be largely reduced by using multiple, unique inhibitory RNA molecules against a single transcript [6]. Lastly, molecular barcodes incorporated into the human RNAi libraries enable the pooling and parallel analysis of thousands of cultured cell lines using a barcode microarray [6,88]. Importantly, the use of molecular barcodes will facilitate the systematic profiling of the human RNAi collection across a wide range of cell types, cellular states, and drug conditions using methods analogous to the yeast deletion system.

Given the relatively recent awareness of the extent and importance of ncRNA, the studies described above all fo-

cus on protein-coding genes. However, a future challenge is the functional dissection of complex transcription including ncRNA across a number of species. At this point, it is unclear how to systematically dissect the function of ncRNA and what the expected outcomes are. Similar to its application to protein-coding genes, mutagenesis is an attractive method to assess the functional role of ncRNA. However, only 2 of 50 ncRNA transcripts in yeast exhibited a mutant phenotype when disrupted under rich media conditions [22]. In contrast, over one-third of protein-coding yeast genes are either essential or required for optimal growth under the same condition [24]. One interpretation of these results is that ncRNA is not as important to the cell as protein-coding genes. An alternative explanation is that this ncRNA plays an important cellular role but that more complex arrangements of mutations are

necessary to observe phenotypes for these transcripts (see below).

Genetic Interactions

Single gene mutant studies of the kind described above are limited in that many genes will not display a phenotype under any tested condition. Therefore, despite extensive profiling of a single-gene mutant library across a wide number of conditions, a sizeable fraction of the genome will escape functional annotation. This limitation of single gene mutant studies can be partially reconciled by the fact that the genome encodes a network of interacting proteins, RNA, metabolites, and other macromolecules that control complex physiological processes. One implication of this cellular complexity is genetic redundancy at the gene or
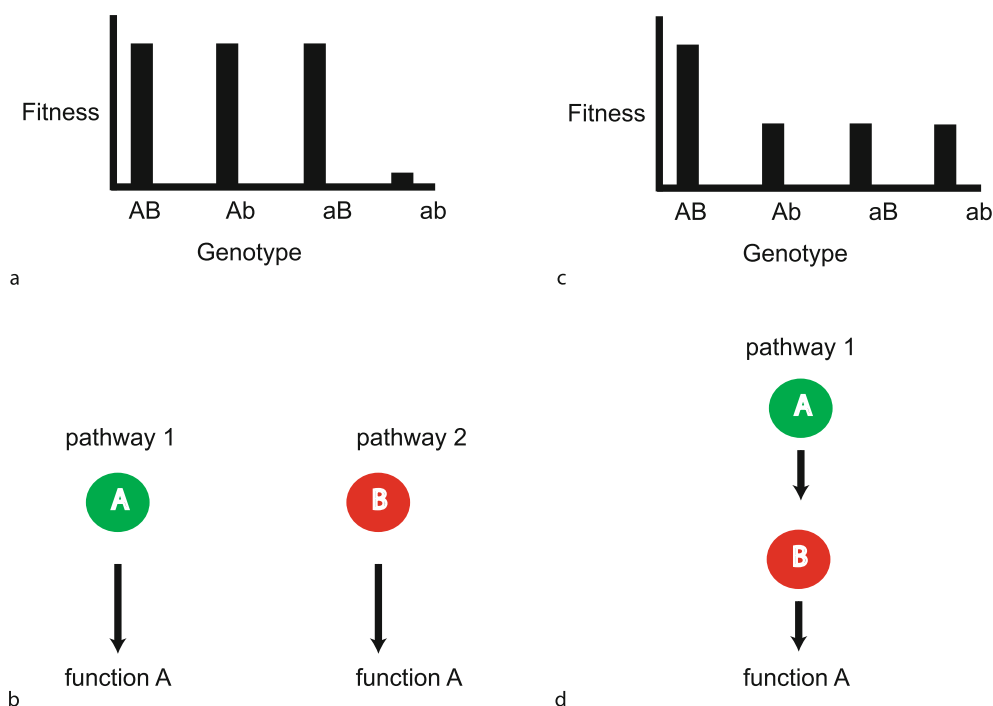
Functional Genomics for Characterization of Genome Sequences, Figure 3
Physical interpretation of genetic interactions. a Relative fitness of two genes, A and B, in combination. Uppercase letter refers to the functional (wildtype) copy of the gene. Lowercase letter refers to a complete loss of function (null) mutation in that gene. Single mutations of both A (genotype aB) and B (Ab) have no effect on fitness. A strain harboring mutations in both A and B (ab) exhibits a severe fitness defect. This deviation is referred to as synergistic epistasis. Synthetic lethality is a special instance of synergistic epistasis in which the double mutant is lethal. **b** The most common physical explanation for synergistic epistasis is that genes A and B act in separate pathways with redundant functions (function A). The loss of either A and B is compensated by the other, intact gene. **c** Using the same terminology in **a**, single gene mutants in both A and B have moderate effects on fitness. The theoretical expectation for genes acting in independent pathways is that the fitness of the double mutant (ab) should be the product of the individual fitness effects. Therefore, the fitness of the double mutant should be less than the fitness of the single mutants. In this example, however, the fitness of the double mutant is equal to the fitness of the single mutants. This deviation is referred to as antagonistic (or alleviating) epistasis. **d** Antagonistic epistasis occurs when two genes act in the same pathway. The loss of either gene singly abolishes the function of the pathway; therefore the loss of both genes would have an identical effect on fitness

pathway level. A mutated gene with an intact, functional copy elsewhere in the genome will not exhibit a mutant phenotype under most conditions. A second implication of cellular complexity is that the network structure itself has evolved robustness against mutations in many of its constituent components. Robustness against mutation is evident in the metabolic network of *E. coli*, where disruptions in many genes are managed by alternative routes through metabolism [61].

One attractive option for unraveling the complexity of these cellular networks and for assigning putative functions to genes is to analyze strains harboring multiple mutations. In instances where the phenotype of the mutant combination deviates from the expected phenotype of combining the single mutants, a genetic interaction is said to exist between the two genes (Fig. 3). Given the potential of genetic interactions to reveal novel functional linkages between genes, researchers have started to systematically identify these interactions at the genome-level. In yeast, techniques have been developed to probe a query mutation against all systematically deleted genes using both automated analysis of individual double mutants [115] and using molecular barcodes, mutant pools, and microarrays [89]. These findings indicate that the average yeast gene has a genetic interaction with ∼34 other nonessential genes. Furthermore, gene pairs that exhibit a genetic interaction are more likely to be functionally related providing evidence that these interactions are biologically meaningful and can be used to assign putative gene functions [115].

Summary

At the current stage, our genome is computationally annotated, transcripts have been mapped using tiling arrays, we have profiled a number of conditions using gene expression microarrays, and we have started to undertake systematic loss-of-function screens to uncover phenotypes for genes. This pattern of experimentation describes an increasingly common scenario for how new genome sequences are analyzed today. What did the mutant profiling screens add? Most importantly, at both the single and double mutant level, mutant phenotypes seems to be more direct indicators of gene function than gene expression data. Therefore, despite the cost and difficulty, it is imperative that systematic mutant collections are tested for a number of model organisms. Additionally, given the effort to generate these collections, novel phenotyping technology needs to be developed such that the value of each mutant collection is maximized. Nevertheless, the experiments described thus far have focused on elucidating genome function in the context of a single representative individual of

a species. Natural populations within species, however, exhibit substantial phenotypic diversity due to the presence of genetic variation in the population (polymorphisms) or environmental effects. Consequently, the precise functionality of a gene in a species can be dependent on the individual that is sampled. In the next section, we therefore describe the nature of intraspecific variation and how the tools of functional genomics are being used to tackle this problem.

Functional Genomics and Complex Traits

Most intraspecific phenotypic variation is conditioned by multiple genes and the environment. Furthermore, these genes can interact with each other and the environment both additively and epistatically to determine phenotype. For these reasons, this class of traits is often referred to as complex (referring to the underlying genetics) or quantitative (referring to the effect of each contributing gene or quantitative trait locus to the overall phenotype). Finding the genetic variation that contributes to quantitative phenotypic variation is a major challenge in contemporary genetics. The susceptibility to many common human diseases including type II diabetes, bipolar disorder, and heart disease is a genetically complex, quantitative trait. Here we present the challenge of complex genetics and discuss how the techniques of transcript profiling and mutant analysis are being applied to address this issue both in human and in model organisms.

Genotyping, Linkage, and Association

A primary goal of quantitative genetics is to identify the genetic variation that contributes to differences in observable traits among the population. Most genetic variation within species is single base pair differences commonly referred to as single nucleotide polymorphisms (SNPs). Therefore, it is necessary that one is able to genotype SNPs within a population and relate this information to phenotypic measurements. SNPs that correlate with a particular phenotypic value (for instance, tall height in humans) are potentially causative. In practice, there are two fundamental approaches for mapping SNPs that contribute to phenotypic variation: mapping based approaches such as linkage analysis and association studies and candidate gene approaches. Following a discussion of modern genotyping techniques, we will first describe the experimental design and application of linkage and association.

There are numerous techniques for genotyping individuals within a population. One can directly sequence a known SNP of interest using terminator chemistry or pyrosequencing. Additionally, single nucleotide extension

assays can be used to genotype known SNPs. The advantages of these techniques are the accuracy of the SNP calls and the simplicity of doing the experiment. The primary disadvantage is the cost and labor required to genotype thousands of SNPs from a single individual. Given the availability and necessity of dense SNP (marker) maps in humans and other model organisms, it is imperative that genotyping techniques be both rapid (so that numerous individuals can be genotyped) and massively parallel (so that thousands of SNPs can be genotyped at once). Microarray-based approaches for genotyping fit these criteria and are available from a number of commercial suppliers [32].

Linkage methods rely on known family history (in human) or controlled laboratory crosses (in model organisms) to map causative genetic variation. The basic premise behind linkage is that the causative allele will segregate preferentially among members of the family sharing a certain phenotypic value and not in those without this phenotype. In model organisms, the advantages of linkage techniques include experimental control and the capability to detect genetic loci with major effects on the phenotype. In one classic example, linkage mapping led to the discovery of fw2.2, a major effect QTL that contributes significantly to the size of tomatoes [36]. In humans, the primary disadvantages of linkage mapping are small family sizes and the absence of a detailed genealogy for most populations. The population of Iceland, in which an isolated population kept detailed records of family history, is an exception; linkage methodology applied to this population has led to the discovery of allelic variants that contribute to a number of diseases including type II diabetes [46].

Given the lack of family history for most human populations, the majority of human mapping efforts are currently focused on association. In association studies, a case group (with the phenotype of interest) and an appropriately chosen control are genotyped for a large number of SNPs. In association studies, the case and control groups rarely contain family members, rather the participants are chosen from broader populations. The goal of association studies is to identify SNPs that are statistically overrepresented in the case group compared to the control group. These SNPs represent candidate chromosomal regions that may contain causative genetic variation. Technological advances underlie the recent success of human association studies. First, a haplotype map of the human genome [60] has reduced the number of SNPs that need to genotyped (due to linkage disequilibrium) to achieve genome-wide coverage. Second, hundreds of thousands of SNPs can be genotyped in parallel using microarray-based approaches. With these resources in hand, researchers

have used association-based methods to identify genetic variants associated with a number of multigenic diseases including type II diabetes [101,105,131] and bipolar disorder [124].

The successful application of linkage and association studies described above emphasizes the complexity of quantitative traits. As such, a number of fundamental questions remain to be answered: how many genes condition a quantitative trait, what are the effect sizes of each of the genes, is the causative variation more likely to be at the regulatory or protein level, and finally, how does genetic variation interact at the molecular level to determine phenotype? In a recurrent theme during the history of molecular biology, the use of model organisms is necessary to uncover fundamental principles of biology. In addition to novel insights into the genetic architecture of quantitative traits, it is anticipated that the techniques used to dissect quantitative variation in model organisms will be applied to the same class of traits in humans. Below we briefly discuss some research developments in model organism quantitative genetics (by primarily candidate gene approaches) using the functional genomic techniques described in this article as a framework.

Application of Transcription

Differences in transcript levels contribute to observable phenotypic variation (for review, [97]). Consequently, researchers have looked into the genetics of gene expression variation in a number of species including yeast [10], mouse, and human [102]. Because each gene and its transcript level can be considered an individual trait, the genetic analysis of gene expression is a powerful model for QTL studies as thousands of traits can be considered simultaneously. A number of lessons emerge from these studies. Early work in yeast suggests that, for the majority of genes, the underlying genetics controlling gene expression is multigenic and complex [10]. Additionally, the majority of loci that affect gene expression are distant acting factors, not local regulatory sequences linked to the gene under investigation. Finally, the majority of causative distant acting factors do not appear to be transcription factors [130]. This argues that the genetic network that contributes to gene expression variation is affected by many global biological processes, not just those directly related to gene transcription.

In mouse and human, the genetic analysis of gene expression is aiding in the identification of QTLs (by proposing candidate genes) and providing insight into disease mechanisms. Schadt et al. examined the transcript levels of 111 mice derived from a cross between two laboratory

a

b
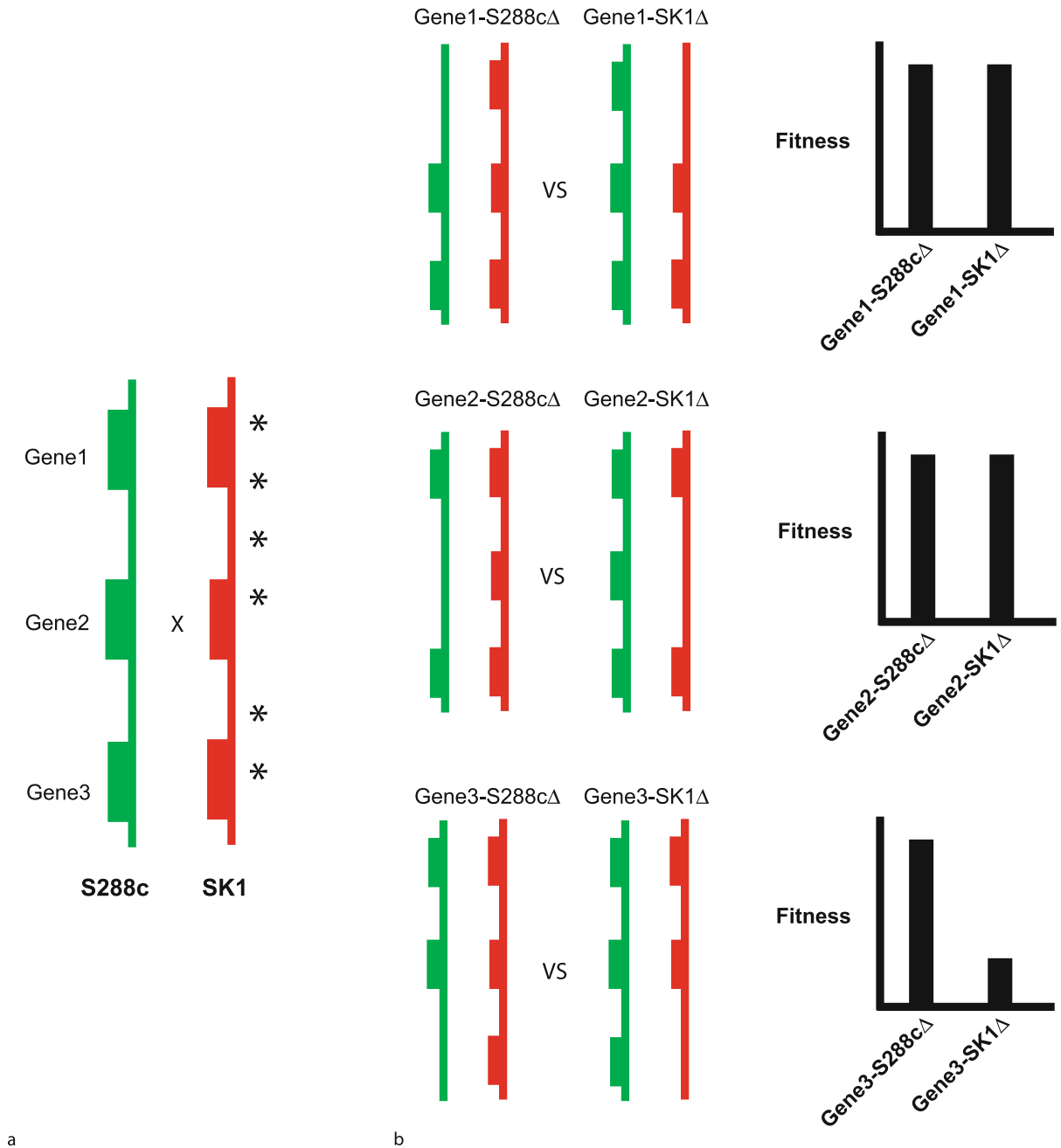
Functional Genomics for Characterization of Genome Sequences, Figure 4

Reciprocal hemizygosity analysis. **a** Reciprocal hemizygosity analysis (RHA) is the quantitative profiling of hemizygous (one allele is deleted), hybrid diploid strains. Illustrated is a hybrid diploid yeast strain derived from a cross between two divergent strains; shown here for S288c (green) and SK1 (red). The asterisks along the SK1 chromosome represent single nucleotide polymorphisms (SNPs) between the two strains. **b** In RHA, two reciprocal deletion strains are compared directly for their contribution to the phenotype (in the example here, fitness). The two strains are completely isogenic except each carries a single gene from a different genetic background. For Gene1 and Gene2, the reciprocal deletion strains harbor identical phenotypes. One can conclude that the SK1 and S288c versions of these alleles contribute equally to fitness in the S288c/SK1 hybrid background. For Gene3, the strain with a deletion in the SK1 allele has a significantly lower fitness than the S288c deletion strain. The interpretation is that the SK1 allele has a larger effect on fitness than the S288c allele in the hybrid background. The SNPs in and around Gene3 are candidates for the causative genetic variation

strains segregating for obesity [102]. Over 2000 genes with transcript level differences between the parents significantly linked to quantitative trait loci which may influence gene expression levels or contribute to obesity. Among the progeny two distinct obesity gene expression subtypes could be defined. Interestingly, the two subtypes are under the genetic control of different QTLs suggesting that multiple molecular mechanisms can contribute to what appears to be a single phenotype (obesity). These results demonstrate that the genetic dissection of gene expression in mammals can classify disease subtypes, identify novel interactions between loci, and identify potential therapeutic targets. However, it is important to note that not all phenotypic variation will be the result of transcript level differences. The importance of other molecular mechanisms, including structural mutations in proteins, implies that multiple, complementary techniques are required to fully understand intraspecific phenotypic variation.

Application of Genetics Analysis

Mutants can provide substantial insight into the genes underlying phenotypic variation because they can mirror the effects of certain types of allelic variation present in the natural population. The analysis of mutants can be used as a verification tool for the identification of the causative locus or as a technique for generating a list of candidate genes for future investigation. In mouse, large forward genetic approaches are being applied to isolate interesting mutants that have phenotypes relevant to human health including obesity, diabetes, and heart disease [85]. While this approach is promising, it is not amenable to high-throughput and the mapping of mutations is still time-consuming. Therefore, reverse-genetic efforts to systematically mutate every gene in the mouse genome, although a monumental task, are likely to provide better coverage for the entire genome in the long run [19].

Mutant screening for QTL detection has limitations similar to those of loss-of-function phenotype screens in model systems, in that redundant gene products are missed, more complex interactions not tested, and expectations are unclear when natural, causative SNPs do not elicit the same effects as the laboratory induced mutations. For these reasons, the power of mutant analysis is increased when it is applied so that a mutant is not compared to a wildtype, but instead a mutation enables a comparison between one allele and another. This technique has been established in yeast by reciprocal hemizygosity analysis (Fig. 4). After linkage mapping, reciprocal hemizygosity analysis was applied to identify the causative genes in each QTL region for the complex traits of high-temper-

ature growth and sporulation efficiency [4,23,112]. These studies highlight the complexity of quantitative traits including the presence of multiple causative genes in a single mapped interval and the non-intuitive genetic interactions between the causative genes. In addition, the yeast work suggests that regulatory and protein coding polymorphisms in both candidate and non-candidate genes contribute to quantitative phenotypic variation. One implication of these results is that new techniques for the high-resolution dissection of QTLs are required in multicellular model organisms to truly understand the genetic basis of phenotype variation.

Summary

We argue that a complete understanding of gene function requires an understanding of how gene function varies (often subtly) among individuals within a population, not just a single representative isolate. To fulfill this promise, two substantial challenges need to be addressed. The first challenge is the identification of the genetic variants that influence any particular phenotype. In model organisms, genome-wide techniques based on linkage/association combined with transcription and mutation analysis are proving successful. Further application of mutant approaches requires substantial infrastructure to generate libraries of strains and reagents for the systematic identification of interesting phenotypes but have promise for circumventing the need for linkage mapping (Fig. 4). The second challenge is to unravel the complexity of genetic and environmental interactions at a molecular level. This will first be accomplished in model systems although it is anticipated that the lessons learned from these studies will be applicable to humans. Functional genomic techniques (combined with non-nucleic based approaches, see **a**) provide the necessary global views into gene function and hence are perhaps most promising for dissecting the genetic interactions underlying phenotypic variation.

Future Directions

Technology and Implications

The relationship between technology and biological insight is circular: biology drives technology development and new technology enables new biology. For example, the desire to sequence human genomes has driven the efforts to develop novel, ultra high-throughput sequencing technology. In turn, functional genomic techniques such as automated DNA sequencing and the DNA microarray have revealed numerous unexpected insights into the biology and functioning of genomes. No doubt, the data de-

rived from novel technology, such as ultra high-throughput sequencing of millions of humans, will revolutionize our view of living systems (and so on). Given the importance of technology in functional genomics (and biology in general), we discuss here novel techniques and their implications for discovery.

Ultra high-throughput sequencing encompass methods capable of sequencing millions of molecules in parallel in reaction volumes far lower than used for conventional Sanger sequencing [80,107]. Given the central role of DNA sequence data in functional genomics, this new generation of technology has important implications for genome-level research and comparative genomics in particular. The increased throughput of DNA sequencing will allow researchers to sample more of the natural variation present on earth. It is anticipated that the genome sequences for most commonly known multicellular organisms will be completed in the next 20 years. Furthermore, novel sequencing technologies may have utility in the analysis of extinct organisms as well. In one example, partial Neanderthal genome sequences were obtained using template derived from a fossil bone [47,87]. Such analyzes of extinct organisms can provide insights into the evolutionary past that is not possible using the DNA sequence data from contemporary species. Lastly, the environmental sequences (metagenome) of largely uncultivatable microbial communities are reshaping the way we view ecosystems [120], biogeochemical cycles [118], and even human health [117].

Above all, novel DNA sequencing technologies have the potential to revolutionize human health by cost-effectively providing the genome sequence of many individuals within the human population. The realization of such promise would usher in a new era of personalized medicine based on one's genome sequence. The identification of all polymorphisms among millions of individuals would increase the power of association studies for finding common disease genes (particularly those with small effect). Additionally, the resequencing of tumor samples will aid in the identification of the responsible mutations and perhaps suggest the appropriate therapeutic treatment.

As described in this review, our ability to sequence genomes far exceeds our ability to systematically determine genome function. Therefore, the development of ultra high-throughput DNA sequencing emphasizes the need for novel functional genomic techniques to decipher gene function. For the analysis of gene expression, we believe two major trends will dominate. First, microarrays will trend towards higher density, multiplexing, and custom design. Higher density of probes on microarrays will enable higher-resolution measurements of transcrip-

tional architecture. The multiplexing of arrays will lower costs per experiment by reducing the amount of necessary reagents. Custom *in situ* synthesized microarrays will allow researchers to keep experimentation up to the pace of genome sequencing. The second major trend in gene expression promises to be the use of ultra high-throughput sequencing as a 'digital' readout of gene expression by directly counting the number of individual RNA molecules in a sample (even at the single cell level). The advantages of this approach are that no microarray is needed, data analysis is more straightforward, and most cellular RNA is assayed in an unbiased manner.

Despite the cost and effort, we anticipate that genome-wide resources (such as defined mutant libraries for all genes) will become available for a number of organisms. Such resources offer the potential to systematically determine gene function on a number of levels from genetic to biochemical. However, we believe the full potential of these resources can only be realized with concurrent advances in experimentation such as robotic automation, assay miniaturization, and labeling/detection systems for monitoring phenotypes and individual molecules (for reviews, see [55,81]). Most importantly, increasing accuracy and throughput of experimentation while simultaneously reducing cost will enable the application of novel technology in the individual laboratory and hence spread the impact of these approaches [110].

Data Integration and Systems Biology

In order to understand cellular function one also needs to examine the dynamics and interactions of biological data types beyond DNA, such as proteins, metabolites, and lipids. Like for DNA, high-throughput technologies have been developed to increase the speed and decrease the cost of obtaining this information, and methods have been developed to integrate diverse data types. These developments characterize the fields of proteomics, metabolomics, modeling, and systems biology, to name but a few. While each of these topics are subjects of other chapters in this book, we touch upon them briefly here to provide a context in which all genomic DNA data needs to be analyzed today. The key is to incorporate all of the available information together, to gain a picture of biological function greater than the sum of the individual parts.

Cells are enormously complex and each of the thousands of biochemical reactions carried out by genes is part of a large network of connected reactions – encoded by information in the genome but regulated by cues from the environment. Conceptually, disturbances in cells, like what happens during any type of disease, can be viewed

as a disturbance in the network, and in order to predict the effect of disturbances we need to identify all molecular components, know where they localize, with whom they interact, and how their activity is regulated.

Among biological entities other than DNA, proteins have been most amenable to high-throughput study (often employing genetic manipulation) and their global characterization is perhaps furthest along. High-throughput technologies have been developed to analyze the expression of proteins [41], their localization [58], phosphorylation patterns [21], protein-DNA interactions [50], and protein-protein interactions [39,40,53,62,71,119]. Any dataset that defines a relationship between proteins can be used for the purpose of making a network. Thus, interaction linkages have been defined based on physical protein-protein interactions [39,40,53,62,71,119], expression regulation [76,113], mutant phenotypes [111,115], phylogenetic profiles [91], literature mining [79], and orthology transfer of interaction evidence across species [128]. Not surprisingly, we now know at least some of the putative interrelationships among most of the proteome of yeast (and other organisms are following) [38]. While many of the interactions are suggestive and need to be confirmed, already they provide a context to functionally characterize proteins and enable an understanding of the whole that goes beyond knowledge from the components in isolation.

Because individual large scale datasets are often incomplete, more information is obtained by integrating datasets. Integrating heterogeneous but complementary interaction data types has improved the accuracy and the coverage in detecting protein associations [121] and has been implemented globally [20,49,59,63,75,84,116]. In one example, integrated protein networks have been constructed for mitochondria, where they have enabled functional predictions to be made for hundreds of previously uncharacterized components, providing a survey of systems properties and enabling the prediction of disease candidate genes [93,108].

Currently, we are still far away from being able to develop any complete models of eukaryotic cells. In order to identify and understand the function of biological networks we need to move from static to dynamic models, from defining the parts to understanding the biological processes in the context of the entire system. Knowledge about systems biology can then fuel the next series of applications, to enable the design and construction of biological systems that can (for example) process information, generate energy, fabricate materials, provide food, or maintain and enhance human health and our environment. For these reasons, as well as for the insights obtained from cracking the code itself, understanding the blueprint of the genome using functional genomics is a fascinating quest, and will remain so for some time to come.

Acknowledgments

Bibliography

1. Abby S, Daubin V (2007) Comparative genomics and the evolution of prokaryotes. Trends Microbiol 15:135–141
2. Adams MD, Celniker SE, Holt RA, Evans CA, Gocayne JD, Amanatides PG, Scherer SE, Li PW, Hoskins RA, Galle RF, George RA, Lewis SE, Richards S, Ashburner M, Henderson SN, Sutton GG, Wortman JR, Yandell MD, Zhang Q, Chen LX, Brandon RC, Rogers YH, Blazej RG, Champe M, Pfeiffer BD, Wan KH, Doyle C, Baxter EG, Helt G, Nelson CR, Gabor GL, Abril JF, Agbayani A, An HJ, Andrews-Pfannkoch C, Baldwin D, Ballew RM, Basu A, Baxendale J, Bayraktaroglu L, Beasley EM, Beeson KY, Benos PV, Berman BP, Bhandari D, Bolshakov S, Borkova D, Botchan MR, Bouck J, Brokstein P, Brottier P, Burtis KC, Busam DA, Butler H, Cadieu E, Center A, Chandra I, Cherry JM, Cawley S, Dahlke C, Davenport LB, Davies P, de Pablos B, Delcher A, Deng Z, Mays AD, Dew I, Dietz SM, Dodson K, Doup LE, Downes M, Dugan-Rocha S, Dunkov BC, Dunn P, Durbin KJ, Evangelista CC, Ferraz C, Ferriera S, Fleischmann W, Fosler C, Gabrielian AE, Garg NS, Gelbart WM, Glasser K, Glodek A, Gong F, Gorrell JH, Gu Z, Guan P, Harris M, Harris NL, Harvey D, Heiman TJ, Hernandez JR, Houck J, Hostin D, Houston KA, Howland TJ, Wei MH, Ibegwam C, et al. (2000) The genome sequence of Drosophila melanogaster. Science 287:2185–2195
3. Baba T, Ara T, Hasegawa M, Takai Y, Okumura Y, Baba M, Datsenko KA, Tomita M, Wanner BL, Mori H (2006) Construction of Escherichia coli K-12 in-frame, single-gene knockout mutants: the Keio collection. Mol Syst Biol 2:2006 0008
4. Ben-Ari G, Zenvirth D, Sherman A, David L, Klutstein M, Lavi U, Hillel J, Simchen G (2006) Four linked genes participate in controlling sporulation efficiency in budding yeast. PLoS Genet 2:e195
5. Berger MF, Philippakis AA, Qureshi AM, He FS, Estep PW, Bulyk ML (2006) Compact, universal DNA microarrays to comprehensively determine transcription-factor binding site specificities. Nat Biotechnol 24:1429–1435
6. Berns K, Hijmans EM, Mullenders J, Brummelkamp TR, Velds A, Heimerikx M, Kerkhoven RM, Madiredjo M, Nijkamp W, Weigelt B, Agami R, Ge W, Cavet G, Linsley PS, Beijersbergen RL, Bernards R (2004) A large-scale RNAi screen in human cells identifies new components of the p53 pathway. Nature 428:431–437
7. Bertone P, Gerstein M, Snyder M (2005) Applications of DNA tiling arrays to experimental genome annotation and regulatory pathway discovery. Chromosome Res 13:259–274
8. Bertone P, Stolc V, Royce TE, Rozowsky JS, Urban AE, Zhu X, Rinn JL, Tongprasit W, Samanta M, Weissman S, Gerstein M, Snyder M (2004) Global identification of human transcribed sequences with genome tiling arrays. Science 306:2242–2246

9. Borsani O, Zhu J, Verslues PE, Sunkar R, Zhu JK (2005) Endogenous siRNAs derived from a pair of natural cis-antisense transcripts regulate salt tolerance in Arabidopsis. Cell 123:1279–1291

10. Brem RB, Yvert G, Clinton R, Kruglyak L (2002) Genetic dissection of transcriptional regulation in budding yeast. Science 296:752–755

11. Brown MP, Grundy WN, Lin D, Cristianini N, Sugnet CW, Furey TS, Ares M Jr, Haussler D (2000) Knowledge-based analysis of microarray gene expression data by using support vector machines. Proc Natl Acad Sci USA 97:262–267

12. Buck MJ, Lieb JD (2004) ChIP-chip: considerations for the design, analysis, and application of genome-wide chromatin immunoprecipitation experiments. Genomics 83:349–360

13. Bushati N, Cohen SM (2007) microRNA Functions. Annu Rev Cell Dev Biol 23:175–205

14. Bustamante CD, Fledel-Alon A, Williamson S, Nielsen R, Hubisz MT, Glanowski S, Tanenbaum DM, White TJ, Sninsky JJ, Hernandez RD, Civello D, Adams MD, Cargill M, Clark AG (2005) Natural selection on protein-coding genes in the human genome. Nature 437:1153–1157

15. C. elegans Sequencing Consortium (1998) Genome sequence of the nematode C. elegans: a platform for investigating biology. Science 282:2012–2018

16. Chen J, Sun M, Hurst LD, Carmichael GG, Rowley JD (2005) Genome-wide analysis of coordinate expression and evolution of human cis-encoded sense-antisense transcripts. Trends Genet 21:326–329

17. Clark AG, Eisen MB, Smith DR, Bergman CM, Oliver B, Markow TA, Kaufman TC, Kellis M, Gelbart W, Iyer VN, Pollard DA, Sackton TB, Larracuente AM, Singh ND, Abad JP, Abt DN, Adryan B, Aguade M, Akashi H, Anderson WW, Aquadro CF, Ardell DH, Arguello R, Artieri CG, Barbash DA, Barker D, Barsanti P, Batterham P, Batzoglou S, Begun D, Bhutkar A, Blanco E, Bosak SA, Bradley RK, Brand AD, Brent MR, Brooks AN, Brown RH, Butlin RK, Caggese C, Calvi BR, Bernardo de Carvalho A, Caspi A, Castrezana S, Celniker SE, Chang JL, Chapple C, Chatterji S, Chinwalla A, Civetta A, Clifton SW, Comeron JM, Costello JC, Coyne JA, Daub J, David RG, Delcher AL, Delehaunty K, Do CB, Ebling H, Edwards K, Eickbush T, Evans JD, Filipski A, Findeiss S, Freyhult E, Fulton L, Fulton R, Garcia AC, Gardiner A, Garfield DA, Garvin BE, Gibson G, Gilbert D, Gnerre S, Godfrey J, Good R, Gotea V, Gravely B, Greenberg AJ, Griffiths-Jones S, Gross S, Guigo R, Gustafson EA, Haerty W, Hahn MW, Halligan DL, Halpern AL, Halter GM, Han MV, Heger A, Hillier L, Hinrichs AS, Holmes I, Hoskins RA, Hubisz MJ, Hultmark D, Huntley MA, Jaffe DB, Jagadeeshan S, et al. (2007) Evolution of genes and genomes on the Drosophila phylogeny. Nature 450:203–218

18. Cliften P, Sudarsanam P, Desikan A, Fulton L, Fulton B, Majors J, Waterston R, Cohen BA, Johnston M (2003) Finding functional features in Saccharomyces genomes by phylogenetic footprinting. Science 301:71–76

19. Collins FS, Rossant J, Wurst W (2007) A mouse for all reasons. Cell 128:9–13

20. Covert MW, Knight EM, Reed JL, Herrgard MJ, Palsson BO (2004) Integrating high-throughput and computational data elucidates bacterial networks. Nature 429:92–96

21. Cox J, Mann M (2007) Is proteomics the new genomics? Cell 130:395–398

22. David L, Huber W, Granovskaia M, Toedling J, Palm CJ, Bofkin L, Jones T, Davis RW, Steinmetz LM (2006) A high-resolution map of transcription in the yeast genome. Proc Natl Acad Sci USA 103:5320–5325

23. Deutschbauer AM, Davis RW (2005) Quantitative trait loci mapped to single-nucleotide resolution in yeast. Nat Genet 37:1333–1340

24. Deutschbauer AM, Jaramillo DF, Proctor M, Kumm J, Hillenmeyer ME, Davis RW, Nislow C, Giaever G (2005) Mechanisms of haploinsufficiency revealed by genome-wide profiling in yeast. Genetics 169:1915–1925

25. Deutschbauer AM, Williams RM, Chu AM, Davis RW (2002) Parallel phenotypic analysis of sporulation and postgermination growth in Saccharomyces cerevisiae. Proc Natl Acad Sci USA 99:15530–15535

26. Dujon B, Sherman D, Fischer G, Durrens P, Casaregola S, Lafontaine I, De Montigny J, Marck C, Neuveglise C, Talla E, Goffard N, Frangeul L, Aigle M, Anthouard V, Babour A, Barbe V, Barnay S, Blanchin S, Beckerich JM, Beyne E, Bleykasten C, Boisrame A, Boyer J, Cattolico L, Confanioleri F, De Daruvar A, Despons L, Fabre E, Fairhead C, Ferry-Dumazet H, Groppi A, Hantraye F, Hennequin C, Jauniaux N, Joyet P, Kachouri R, Kerrest A, Koszul R, Lemaire M, Lesur I, Ma L, Muller H, Nicaud JM, Nikolski M, Oztas S, Ozier-Kalogeropoulos O, Pellenz S, Potier S, Richard GF, Straub ML, Suleau A, Swennen D, Tekaia F, Wesolowski-Louvel M, Westhof E, Wirth B, Zeniou-Meyer M, Zivanovic I, Bolotin-Fukuhara M, Thierry A, Bouchier C, Caudron B, Scarpelli C, Gaillardin C, Weissenbach J, Wincker P, Souciet JL (2004) Genome evolution in yeasts. Nature 430:35–44

27. Eddy SR (2002) Computational genomics of noncoding RNA genes. Cell 109:137–140

28. Elnitski L, Jin VX, Farnham PJ, Jones SJ (2006) Locating mammalian transcription factor binding sites: a survey of computational and experimental techniques. Genome Res 16:1455–1464

29. Enard W, Przeworski M, Fisher SE, Lai CS, Wiebe V, Kitano T, Monaco AP, Paabo S (2002) Molecular evolution of FOXP2, a gene involved in speech and language. Nature 418:869–872

30. ENCODE Consortium (2007a) Identification and analysis of functional elements in 1% of the human genome by the ENCODE pilot project. Nature 447:799–816

31. Faith JJ, Hayete B, Thaden JT, Mogno I, Wierzbowski J, Cottarel G, Kasif S, Collins JJ, Gardner TS (2007) Large-scale mapping and validation of Escherichia coli transcriptional regulation from a compendium of expression profiles. PLoS Biol 5:e8

32. Fan JB, Chee MS, Gunderson KL (2006) Highly parallel genomic assays. Nat Rev Genet 7:632–644

33. Fire A, Xu S, Montgomery MK, Kostas SA, Driver SE, Mello CC (1998) Potent and specific genetic interference by double-stranded RNA in Caenorhabditis elegans. Nature 391:806–811

34. Fisher S, Grice EA, Vinton RM, Bessling SL, McCallion AS (2006) Conservation of RET regulatory function from human to zebrafish without sequence similarity. Science 312:276–279

35. Fleischmann RD, Adams MD, White O, Clayton RA, Kirkness EF, Kerlavage AR, Bult CJ, Tomb JF, Dougherty BA, Merrick JM, et al. (1995) Whole-genome random sequencing and assembly of Haemophilus influenzae Rd. Science 269:496–512

36. Frary A, Nesbitt TC, Grandillo S, Knaap E, Cong B, Liu J, Meller J, Elber R, Alpert KB, Tanksley SD (2000) fw2.2: a quantitative

trait locus key to the evolution of tomato fruit size. Science 289:85–88

37. Friedman A, Perrimon N (2006) A functional RNAi screen for regulators of receptor tyrosine kinase and ERK signalling. Nature 444:230–234

38. Gagneur J, David L, Steinmetz LM (2006) Capturing cellular machines by systematic screens of protein complexes. Trends Microbiol 14:336–339

39. Gavin AC, Aloy P, Grandi P, Krause R, Boesche M, Marzioch M, Rau C, Jensen LJ, Bastuck S, Dumpelfeld B, Edelmann A, Heurtier MA, Hoffman V, Hoefert C, Klein K, Hudak M, Michon AM, Schelder M, Schirle M, Remor M, Rudi T, Hooper S, Bauer A, Bouwmeester T, Casari G, Drewes G, Neubauer G, Rick JM, Kuster B, Bork P, Russell RB, Superti-Furga G (2006) Proteome survey reveals modularity of the yeast cell machinery. Nature 440:631–636

40. Gavin AC, Bosche M, Krause R, Grandi P, Marzioch M, Bauer A, Schultz J, Rick JM, Michon AM, Cruciat CM, Remor M, Hofert C, Schelder M, Brajenovic M, Ruffner H, Merino A, Klein K, Hudak M, Dickson D, Rudi T, Gnau V, Bauch A, Bastuck S, Huhse B, Leutwein C, Heurtier MA, Copley RR, Edelmann A, Querfurth E, Rybin V, Drewes G, Raida M, Bouwmeester T, Bork P, Seraphin B, Kuster B, Neubauer G, Superti-Furga G (2002) Functional organization of the yeast proteome by systematic analysis of protein complexes. Nature 415:141–147

41. Ghaemmaghami S, Huh WK, Bower K, Howson RW, Belle A, Dephoure N, O'Shea EK, Weissman JS (2003) Global analysis of protein expression in yeast. Nature 425:737–741

42. Giaever G, Chu AM, Ni L, Connelly C, Riles L, Veronneau S, Dow S, Lucau-Danila A, Anderson K, Andre B, Arkin AP, Astromoff A, El-Bakkoury M, Bangham R, Benito R, Brachat S, Campanaro S, Curtiss M, Davis K, Deutschbauer A, Entian KD, Flaherty P, Foury F, Garfinkel DJ, Gerstein M, Gotte D, Guldener U, Hegemann JH, Hempel S, Herman Z, Jaramillo DF, Kelly DE, Kelly SL, Kotter P, LaBonte D, Lamb DC, Lan N, Liang H, Liao H, Liu L, Luo C, Lussier M, Mao R, Menard P, Ooi SL, Revuelta JL, Roberts CJ, Rose M, Ross-Macdonald P, Scherens B, Schimmack G, Shafer B, Shoemaker DD, Sookhai-Mahadeo S, Storms RK, Strathern JN, Valle G, Voet M, Volckaert G, Wang CY, Ward TR, Wilhelmy J, Winzeler EA, Yang Y, Yen G, Youngman E, Yu K, Bussey H, Boeke JD, Snyder M, Philippsen P, Davis RW, Johnston M (2002) Functional profiling of the Saccharomyces cerevisiae genome. Nature 418:387–391

43. Giaever G, Flaherty P, Kumm J, Proctor M, Nislow C, Jaramillo DF, Chu AM, Jordan MI, Arkin AP, Davis RW (2004) Chemogenomic profiling: identifying the functional interactions of small molecules in yeast. Proc Natl Acad Sci USA 101:793–798

44. Glass JI, Assad-Garcia N, Alperovich N, Yooseph S, Lewis MR, Maruf M, Hutchison CA, Smith HO, Venter JC (2006) Essential genes of a minimal bacterium. Proc Natl Acad Sci USA 103:425–430

45. Goffeau A, Barrell BG, Bussey H, Davis RW, Dujon B, Feldmann H, Galibert F, Hoheisel JD, Jacq C, Johnston M, Louis EJ, Mewes HW, Murakami Y, Philippsen P, Tettelin H, Oliver SG (1996) Life with 6000 genes. Science 274:546, 563–567

46. Grant SF, Thorleifsson G, Reynisdottir I, Benediktsson R, Manolescu A, Sainz J, Helgason A, Stefansson H, Emilsson V, Helgadottir A, Styrkarsdottir U, Magnusson KP, Walters GB, Palsdottir E, Jonsdottir T, Gudmundsdottir T, Gylfason A, Saemundsdottir J, Wilensky RL, Reilly MP, Rader DJ, Bagger Y, Christiansen C, Gudnason V, Sigurdsson G, Thorsteinsdottir U,

Gulcher JR, Kong A, Stefansson K (2006) Variant of transcription factor 7-like 2 (TCF7L2) gene confers risk of type 2 diabetes. Nat Genet 38:320–323

47. Green RE, Krause J, Ptak SE, Briggs AW, Ronan MT, Simons JF, Du L, Egholm M, Rothberg JM, Paunovic M, Paabo S (2006) Analysis of one million base pairs of Neanderthal DNA. Nature 444:330–336

48. Guigo R, Dermitzakis ET, Agarwal P, Ponting CP, Parra G, Reymond A, Abril JF, Keibler E, Lyle R, Ucla C, Antonarakis SE, Brent MR (2003) Comparison of mouse and human genomes followed by experimental verification yields an estimated 1,019 additional genes. Proc Natl Acad Sci USA 100: 1140–1145

49. Gunsalus KC, Ge H, Schetter AJ, Goldberg DS, Han JD, Hao T, Berriz GF, Bertin N, Huang J, Chuang LS, Li N, Mani R, Hyman AA, Sonnichsen B, Echeverri CJ, Roth FP, Vidal M, Piano F (2005) Predictive models of molecular machines involved in Caenorhabditis elegans early embryogenesis. Nature 436:861–865

50. Harbison CT, Gordon DB, Lee TI, Rinaldi NJ, Macisaac KD, Danford TW, Hannett NM, Tagne JB, Reynolds DB, Yoo J, Jennings EG, Zeitlinger J, Pokholok DK, Kellis M, Rolfe PA, Takusagawa KT, Lander ES, Gifford DK, Fraenkel E, Young RA (2004) Transcriptional regulatory code of a eukaryotic genome. Nature 431:99–104

51. Hardison RC (2003) Comparative genomics. PLoS Biol 1:e58

52. Hensel M, Shea JE, Gleeson C, Jones MD, Dalton E, Holden DW (1995) Simultaneous identification of bacterial virulence genes by negative selection. Science 269:400–403

53. Ho Y, Gruhler A, Heilbut A, Bader GD, Moore L, Adams SL, Millar A, Taylor P, Bennett K, Boutilier K, Yang L, Wolting C, Donaldson I, Schandorff S, Shewnarane J, Vo M, Taggart J, Goudreault M, Muskat B, Alfarano C, Dewar D, Lin Z, Michalickova K, Willems AR, Sassi H, Nielsen PA, Rasmussen KJ, Andersen JR, Johansen LE, Hansen LH, Jespersen H, Podtelejnikov A, Nielsen E, Crawford J, Poulsen V, Sorensen BD, Matthiesen J, Hendrickson RC, Gleeson F, Pawson T, Moran MF, Durocher D, Mann M, Hogue CW, Figeys D, Tyers M (2002) Systematic identification of protein complexes in Saccharomyces cerevisiae by mass spectrometry. Nature 415:180–183

54. Hogan GJ, Lee CK, Lieb JD (2006) Cell cycle-specified fluctuation of nucleosome occupancy at gene promoters. PLoS Genet 2:e158

55. Hong JW, Quake SR (2003) Integrated nanoliter systems. Nat Biotechnol 21:1179–1183

56. Hongay CF, Grisafi PL, Galitski T, Fink GR (2006) Antisense transcription controls cell fate in Saccharomyces cerevisiae. Cell 127:735–745

57. Hughes TR, Marton MJ, Jones AR, Roberts CJ, Stoughton R, Armour CD, Bennett HA, Coffey E, Dai H, He YD, Kidd MJ, King AM, Meyer MR, Slade D, Lum PY, Stepaniants SB, Shoemaker DD, Gachotte D, Chakraburtty K, Simon J, Bard M, Friend SH (2000) Functional discovery via a compendium of expression profiles. Cell 102:109–126

58. Huh WK, Falvo JV, Gerke LC, Carroll AS, Howson RW, Weissman JS, O'Shea EK (2003) Global analysis of protein localization in budding yeast. Nature 425:686–691

59. Ideker T, Thorsson V, Ranish JA, Christmas R, Buhler J, Eng JK, Bumgarner R, Goodlett DR, Aebersold R, Hood L (2001) Inte-

grated genomic and proteomic analyses of a systematically perturbed metabolic network. Science 292:929–934

60. International HapMap Consortium (2005) A haplotype map of the human genome. Nature 437:1299–1320

61. Ishii N, Nakahigashi K, Baba T, Robert M, Soga T, Kanai A, Hirasawa T, Naba M, Hirai K, Hoque A, Ho PY, Kakazu Y, Sugawara K, Igarashi S, Harada S, Masuda T, Sugiyama N, Togashi T, Hasegawa M, Takai Y, Yugi K, Arakawa K, Iwata N, Toya Y, Nakayama Y, Nishioka T, Shimizu K, Mori H, Tomita M (2007) Multiple high-throughput analyses monitor the response of E. coli to perturbations. Science 316:593–597

62. Ito T, Chiba T, Ozawa R, Yoshida M, Hattori M, Sakaki Y (2001) A comprehensive two-hybrid analysis to explore the yeast protein interactome. Proc Natl Acad Sci USA 98:4569–4574

63. Jansen R, Yu H, Greenbaum D, Kluger Y, Krogan NJ, Chung S, Emili A, Snyder M, Greenblatt JF, Gerstein M (2003) A Bayesian networks approach for predicting protein-protein interactions from genomic data. Science 302:449–453

64. Johnson DS, Mortazavi A, Myers RM, Wold B (2007) Genome-wide mapping of in vivo protein-DNA interactions. Science 316:1497–1502

65. Kamath RS, Fraser AG, Dong Y, Poulin G, Durbin R, Gotta M, Kanapin A, Le Bot N, Moreno S, Sohrmann M, Welchman DP, Zipperlen P, Ahringer J (2003) Systematic functional analysis of the Caenorhabditis elegans genome using RNAi. Nature 421:231–237

66. Kapranov P, Cheng J, Dike S, Nix DA, Duttagupta R, Willingham AT, Stadler PF, Hertel J, Hackermuller J, Hofacker IL, Bell I, Cheung E, Drenkow J, Dumais E, Patel S, Helt G, Ganesh M, Ghosh S, Piccolboni A, Sementchenko V, Tammana H, Gingeras TR (2007a) RNA maps reveal new RNA classes and a possible function for pervasive transcription. Science 316:1484–1488

67. Kapranov P, Willingham AT, Gingeras TR (2007b) Genome-wide transcription and the implications for genomic organization. Nat Rev Genet 8:413–423

68. Kellis M, Patterson N, Endrizzi M, Birren B, Lander ES (2003) Sequencing and comparison of yeast species to identify genes and regulatory elements. Nature 423:241–254

69. Kobayashi K, Ehrlich SD, Albertini A, Amati G, Andersen KK, Arnaud M, Asai K, Ashikaga S, Aymerich S, Bessieres P, Boland F, Brignell SC, Bron S, Bunai K, Chapuis J, Christiansen LC, Danchin A, Debarbouille M, Dervyn E, Deuerling E, Devine K, Devine SK, Dreesen O, Errington J, Fillinger S, Foster SJ, Fujita Y, Galizzi A, Gardan R, Eschevins C, Fukushima T, Haga K, Harwood CR, Hecker M, Hosoya D, Hullo MF, Kakeshita H, Karamata D, Kasahara Y, Kawamura F, Koga K, Koski P, Kuwana R, Imamura D, Ishimaru M, Ishikawa S, Ishio I, Le Coq D, Masson A, Mauel C, Meima R, Mellado RP, Moir A, Moriya S, Nagakawa E, Nanamiya H, Nakai S, Nygaard P, Ogura M, Ohanan T, O'Reilly M, O'Rourke M, Pragai Z, Pooley HM, Rapoport G, Rawlins JP, Rivas LA, Rivolta C, Sadaie A, Sadaie Y, Sarvas M, Sato T, Saxild HH, Scanlan E, Schumann W, Seegers JF, Sekiguchi J, Sekowska A, Seror SJ, Simon M, Stragier P, Studer R, Takamatsu H, Tanaka T, Takeuchi M, Thomaides HB, Vagner V, van Dijl JM, Watabe K, Wipat A, Yamamoto H, Yamamoto M, Yamamoto Y, Yamane K, Yata K, Yoshida K, Yoshikawa H, Zuber U, Ogasawara N (2003) Essential Bacillus subtilis genes. Proc Natl Acad Sci USA 100:4678–4683

70. Korf I, Flicek P, Duan D, Brent MR (2001) Integrating genomic homology into gene structure prediction. Bioinformatics 17 Suppl 1:S140–S148

71. Krogan NJ, Cagney G, Yu H, Zhong G, Guo X, Ignatchenko A, Li J, Pu S, Datta N, Tikuisis AP, Punna T, Peregrin-Alvarez JM, Shales M, Zhang X, Davey M, Robinson MD, Paccanaro A, Bray JE, Sheung A, Beattie B, Richards DP, Canadien V, Lalev A, Mena F, Wong P, Starostine A, Canete MM, Vlasblom J, Wu S, Orsi C, Collins SR, Chandran S, Haw R, Rilstone JJ, Gandi K, Thompson NJ, Musso G, St Onge P, Ghanny S, Lam MH, Butland G, Altaf-Ul AM, Kanaya S, Shilatifard A, O'Shea E, Weissman JS, Ingles CJ, Hughes TR, Parkinson J, Gerstein M, Wodak SJ, Emili A, Greenblatt JF (2006) Global landscape of protein complexes in the yeast Saccharomyces cerevisiae. Nature 440:637–643

72. Kumar S, Subramanian S (2002) Mutation rates in mammalian genomes. Proc Natl Acad Sci USA 99:803–808

73. Lander ES, Linton LM, Birren B, Nusbaum C, Zody MC, Baldwin J, Devon K, Dewar K, Doyle M, FitzHugh W, Funke R, Gage D, Harris K, Heaford A, Howland J, Kann L, Lehoczky J, LeVine R, McEwan P, McKernan K, Meldrim J, Mesirov JP, Miranda C, Morris W, Naylor J, Raymond C, Rosetti M, Santos R, Sheridan A, Sougnez C, Stange-Thomann N, Stojanovic N, Subramanian A, Wyman D, Rogers J, Sulston J, Ainscough R, Beck S, Bentley D, Burton J, Clee C, Carter N, Coulson A, Deadman R, Deloukas P, Dunham A, Dunham I, Durbin R, French L, Grafham D, Gregory S, Hubbard T, Humphray S, Hunt A, Jones M, Lloyd C, McMurray A, Matthews L, Mercer S, Milne S, Mullikin JC, Mungall A, Plumb R, Ross M, Shownkeen R, Sims S, Waterston RH, Wilson RK, Hillier LW, McPherson JD, Marra MA, Mardis ER, Fulton LA, Chinwalla AT, Pepin KH, Gish WR, Chissoe SL, Wendl MC, Delehaunty KD, Miner TL, Delehaunty A, Kramer JB, Cook LL, Fulton RS, Johnson DL, Minx PJ, Clifton SW, Hawkins T, Branscomb E, Predki P, Richardson P, Wenning S, Slezak T, Doggett N, Cheng JF, Olsen A, Lucas S, Elkin C, Uberbacher E, Frazier M, et al. (2001) Initial sequencing and analysis of the human genome. Nature 409:860–921

74. Lapidot M, Pilpel Y (2006) Genome-wide natural antisense transcription: coupling its regulation to its different regulatory mechanisms. EMBO Rep 7:1216–1222

75. Lee I, Date SV, Adai AT, Marcotte EM (2004) A probabilistic functional network of yeast genes. Science 306:1555–1558

76. Lee TI, Rinaldi NJ, Robert F, Odom DT, Bar-Joseph Z, Gerber GK, Hannett NM, Harbison CT, Thompson CM, Simon I, Zeitlinger J, Jennings EG, Murray HL, Gordon DB, Ren B, Wyrick JJ, Tagne JB, Volkert TL, Fraenkel E, Gifford DK, Young RA (2002) Transcriptional regulatory networks in Saccharomyces cerevisiae. Science 298:799–804

77. Lee W, Tillo D, Bray N, Morse RH, Davis RW, Hughes TR, Nislow C (2007) A high-resolution atlas of nucleosome occupancy in yeast. Nat Genet 39:1235–1244

78. Lieb JD, Liu X, Botstein D, Brown PO (2001) Promoter-specific binding of Rap1 revealed by genome-wide maps of protein-DNA association. Nat Genet 28:327–334

79. Marcotte EM, Xenarios I, Eisenberg D (2001) Mining literature for protein-protein interactions. Bioinformatics 17:359–363

80. Margulies M, Egholm M, Altman WE, Attiya S, Bader JS, Bemben LA, Berka J, Braverman MS, Chen YJ, Chen Z, Dewell SB, Du L, Fierro JM, Gomes XV, Godwin BC, He W, Helgesen S, Ho CH, Irzyk GP, Jando SC, Alenquer ML, Jarvie TP, Jirage KB, Kim JB, Knight JR, Lanza JR, Leamon JH, Lefkowitz SM, Lei M, Li J, Lohman KL, Lu H, Makhijani VB, McDade KE, McKenna MP,

Myers EW, Nickerson E, Nobile JR, Plant R, Puc BP, Ronan MT, Roth GT, Sarkis GJ, Simons JF, Simpson JW, Srinivasan M, Tartaro KR, Tomasz A, Vogt KA, Volkmer GA, Wang SH, Wang Y, Weiner MP, Yu P, Begley RF, Rothberg JM (2005) Genome sequencing in microfabricated high-density picolitre reactors. Nature 437:376–380

81. Melin J, Quake SR (2007) Microfluidic large-scale integration: the evolution of design rules for biological automation. Annu Rev Biophys Biomol Struct 36:213–231

82. Meyer IM, Durbin R (2004) Gene structure conservation aids similarity based gene prediction. Nucleic Acids Res 32:776–783

83. Mikkelsen TS, Ku M, Jaffe DB, Issac B, Lieberman E, Giannoukos G, Alvarez P, Brockman W, Kim TK, Koche RP, Lee W, Mendenhall E, O'Donovan A, Presser A, Russ C, Xie X, Meissner A, Wernig M, Jaenisch R, Nusbaum C, Lander ES, Bernstein BE (2007) Genome-wide maps of chromatin state in pluripotent and lineage-committed cells. Nature 448:553–560

84. Myers CL, Robson D, Wible A, Hibbs MA, Chiriac C, Theesfeld CL, Dolinski K, Troyanskaya OG (2005) Discovery of biological networks from diverse functional genomic data. Genome Biol 6:R114

85. Nadeau JH, Balling R, Barsh G, Beier D, Brown SD, Bucan M, Camper S, Carlson G, Copeland N, Eppig J, Fletcher C, Frankel WN, Ganten D, Goldowitz D, Goodnow C, Guenet JL, Hicks G, Hrabe de Angelis M, Jackson I, Jacob HJ, Jenkins N, Johnson D, Justice M, Kay S, Kingsley D, Lehrach H, Magnuson T, Meisler M, Poustka A, Rinchik EM, Rossant J, Russell LB, Schimenti J, Shiroishi T, Skarnes WC, Soriano P, Stanford W, Takahashi JS, Wurst W, Zimmer A (2001) Sequence interpretation. Functional annotation of mouse genome sequences. Science 291:1251–1255

86. Nobrega MA, Zhu Y, Plajzer-Frick I, Afzal V, Rubin EM (2004) Megabase deletions of gene deserts result in viable mice. Nature 431:988–993

87. Noonan JP, Coop G, Kudaravalli S, Smith D, Krause J, Alessi J, Chen F, Platt D, Paabo S, Pritchard JK, Rubin EM (2006) Sequencing and analysis of Neanderthal genomic DNA. Science 314:1113–1118

88. Paddison PJ, Silva JM, Conklin DS, Schlabach M, Li M, Aruleba S, Balija V, O'Shaughnessy A, Gnoj L, Scobie K, Chang K, Westbrook T, Cleary M, Sachidanandam R, McCombie WR, Elledge SJ, Hannon GJ (2004) A resource for large-scale RNA-interference-based screens in mammals. Nature 428:427–431

89. Pan X, Yuan DS, Xiang D, Wang X, Sookhai-Mahadeo S, Bader JS, Hieter P, Spencer F, Boeke JD (2004) A robust toolkit for functional profiling of the yeast genome. Mol Cell 16:487–496

90. Parsons AB, Lopez A, Givoni IE, Williams DE, Gray CA, Porter J, Chua G, Sopko R, Brost RL, Ho CH, Wang J, Ketela T, Brenner C, Brill JA, Fernandez GE, Lorenz TC, Payne GS, Ishihara S, Ohya Y, Andrews B, Hughes TR, Frey BJ, Graham TR, Andersen RJ, Boone C (2006) Exploring the mode-of-action of bioactive compounds by chemical-genetic profiling in yeast. Cell 126:611–625

91. Pellegrini M, Marcotte EM, Thompson MJ, Eisenberg D, Yeates TO (1999) Assigning protein functions by comparative genome analysis: protein phylogenetic profiles. Proc Natl Acad Sci USA 96:4285–4288

92. Pena-Castillo L, Hughes TR (2007) Why are there still over 1000 uncharacterized yeast genes? Genetics 176:7–14

93. Perocchi F, Jensen LJ, Gagneur J, Ahting U, von Mering C, Bork P, Prokisch H, Steinmetz LM (2006) Assessing systems properties of yeast mitochondria through an interaction map of the organelle. PLoS Genet 2:e170

94. Prescott EM, Proudfoot NJ (2002) Transcriptional collision between convergent genes in budding yeast. Proc Natl Acad Sci USA 99:8796–8801

95. Rivas E, Klein RJ, Jones TA, Eddy SR (2001) Computational identification of noncoding RNAs in E. coli by comparative genomics. Curr Biol 11:1369–1373

96. Robertson G, Hirst M, Bainbridge M, Bilenky M, Zhao Y, Zeng T, Euskirchen G, Bernier B, Varhol R, Delaney A, Thiessen N, Griffith OL, He A, Marra M, Snyder M, Jones S (2007) Genome-wide profiles of STAT1 DNA association using chromatin immunoprecipitation and massively parallel sequencing. Nat Methods 4:651–657

97. Rockman MV, Kruglyak L (2006) Genetics of global gene expression. Nat Rev Genet 7:862–872

98. Roemer T, Jiang B, Davison J, Ketela T, Veillette K, Breton A, Tandia F, Linteau A, Sillaots S, Marta C, Martel N, Veronneau S, Lemieux S, Kauffman S, Becker J, Storms R, Boone C, Bussey H (2003) Large-scale essential gene identification in Candida albicans and applications to antifungal drug discovery. Mol Microbiol 50:167–181

99. Sanger F, Nicklen S, Coulson AR (1977) DNA sequencing with chain-terminating inhibitors. Proc Natl Acad Sci USA 74:5463–5467

100. Sassetti CM, Boyd DH, Rubin EJ (2001) Comprehensive identification of conditionally essential genes in mycobacteria. Proc Natl Acad Sci USA 98:12712–12717

101. Saxena R, Voight BF, Lyssenko V, Burtt NP, de Bakker PI, Chen H, Roix JJ, Kathiresan S, Hirschhorn JN, Daly MJ, Hughes TE, Groop L, Altshuler D, Almgren P, Florez JC, Meyer J, Ardlie K, Bengtsson Bostrom K, Isomaa B, Lettre G, Lindblad U, Lyon HN, Melander O, Newton-Cheh C, Nilsson P, Orho-Melander M, Rastam L, Speliotes EK, Taskinen MR, Tuomi T, Guiducci C, Berglund A, Carlson J, Gianniny L, Hackett R, Hall L, Holmkvist J, Laurila E, Sjogren M, Sterner M, Surti A, Svensson M, Svensson M, Tewhey R, Blumenstiel B, Parkin M, Defelice M, Barry R, Brodeur W, Camarata J, Chia N, Fava M, Gibbons J, Handsaker B, Healy C, Nguyen K, Gates C, Sougnez C, Gage D, Nizzari M, Gabriel SB, Chirn GW, Ma Q, Parikh H, Richardson D, Ricke D, Purcell S (2007) Genome-wide association analysis identifies loci for type 2 diabetes and triglyceride levels. Science 316:1331–1336

102. Schadt EE, Monks SA, Drake TA, Lusis AJ, Che N, Colinayo V, Ruff TG, Milligan SB, Lamb JR, Cavet G, Linsley PS, Mao M, Stoughton RB, Friend SH (2003) Genetics of gene expression surveyed in maize, mouse and man. Nature 422:297–302

103. Schena M, Shalon D, Davis RW, Brown PO (1995) Quantitative monitoring of gene expression patterns with a complementary DNA microarray. Science 270:467–470

104. Scherens B, Goffeau A (2004) The uses of genome-wide yeast mutant collections. Genome Biol 5:229

105. Scott LJ, Mohlke KL, Bonnycastle LL, Willer CJ, Li Y, Duren WL, Erdos MR, Stringham HM, Chines PS, Jackson AU, Prokunina-Olsson L, Ding CJ, Swift AJ, Narisu N, Hu T, Pruim R, Xiao R, Li XY, Conneely KN, Riebow NL, Sprau AG, Tong M, White PP, Hetrick KN, Barnhart MW, Bark CW, Goldstein JL, Watkins L, Xiang F, Saramies J, Buchanan TA, Watanabe RM, Valle TT, Kinnunen L, Abecasis GR, Pugh EW, Doheny KF, Bergman RN, Tuomilehto J, Collins FS, Boehnke M (2007) A genome-wide

association study of type 2 diabetes in Finns detects multiple susceptibility variants. Science 316:1341–1345

106. Segal E, Fondufe-Mittendorf Y, Chen L, Thastrom A, Field Y, Moore IK, Wang JP, Widom J (2006) A genomic code for nucleosome positioning. Nature 442:772–778

107. Shendure J, Porreca GJ, Reppas NB, Lin X, McCutcheon JP, Rosenbaum AM, Wang MD, Zhang K, Mitra RD, Church GM (2005) Accurate multiplex polony sequencing of an evolved bacterial genome. Science 309:1728–1732

108. Shutt TE, Shadel GS (2007) Expanding the mitochondrial interactome. Genome Biol 8:203

109. Smith LM, Sanders JZ, Kaiser RJ, Hughes P, Dodd C, Connell CR, Heiner C, Kent SB, Hood LE (1986) Fluorescence detection in automated DNA sequence analysis. Nature 321:674–679

110. Steinmetz LM, Davis RW (2004) Maximizing the potential of functional genomics. Nat Rev Genet 5:190–201

111. Steinmetz LM, Scharfe C, Deutschbauer AM, Mokranjac D, Herman ZS, Jones T, Chu AM, Giaever G, Prokisch H, Oefner PJ, Davis RW (2002a) Systematic screen for human disease genes in yeast. Nat Genet 31:400–404

112. Steinmetz LM, Sinha H, Richards DR, Spiegelman JI, Oefner PJ, McCusker JH, Davis RW (2002b) Dissecting the architecture of a quantitative trait locus in yeast. Nature 416:326–330

113. Stuart JM, Segal E, Koller D, Kim SK (2003) A gene-coexpression network for global discovery of conserved genetic modules. Science 302:249–255

114. Taft RJ, Pheasant M, Mattick JS (2007) The relationship between non-protein-coding DNA and eukaryotic complexity. Bioessays 29:288–299

115. Tong AH, Lesage G, Bader GD, Ding H, Xu H, Xin X, Young J, Berriz GF, Brost RL, Chang M, Chen Y, Cheng X, Chua G, Friesen H, Goldberg DS, Haynes J, Humphries C, He G, Hussein S, Ke L, Krogan N, Li Z, Levinson JN, Lu H, Menard P, Munyana C, Parsons AB, Ryan O, Tonikian R, Roberts T, Sdicu AM, Shapiro J, Sheikh B, Suter B, Wong SL, Zhang LV, Zhu H, Burd CG, Munro S, Sander C, Rine J, Greenblatt J, Peter M, Bretscher A, Bell G, Roth FP, Brown GW, Andrews B, Bussey H, Boone C (2004) Global mapping of the yeast genetic interaction network. Science 303:808–813

116. Troyanskaya OG, Dolinski K, Owen AB, Altman RB, Botstein D (2003) A Bayesian framework for combining heterogeneous data sources for gene function prediction (in Saccharomyces cerevisiae). Proc Natl Acad Sci USA 100:8348–8353

117. Turnbaugh PJ, Ley RE, Mahowald MA, Magrini V, Mardis ER, Gordon JI (2006) An obesity-associated gut microbiome with increased capacity for energy harvest. Nature 444:1027–1031

118. Tyson GW, Chapman J, Hugenholtz P, Allen EE, Ram RJ, Richardson PM, Solovyev VV, Rubin EM, Rokhsar DS, Banfield JF (2004) Community structure and metabolism through reconstruction of microbial genomes from the environment. Nature 428:37–443

119. Uetz P, Giot L, Cagney G, Mansfield TA, Judson RS, Knight JR, Lockshon D, Narayan V, Srinivasan M, Pochart P, Qureshi-Emili A, Li Y, Godwin B, Conover D, Kalbfleisch T, Vijayadamodar G, Yang M, Johnston M, Fields S, Rothberg JM (2000) A comprehensive analysis of protein-protein interactions in Saccharomyces cerevisiae. Nature 403:623–627

120. Venter JC, Remington K, Heidelberg JF, Halpern AL, Rusch D, Eisen JA, Wu D, Paulsen I, Nelson KE, Nelson W, Fouts DE, Levy S, Knap AH, Lomas MW, Nealson K, White O, Peterson J, Hoffman J, Parsons R, Baden-Tillson H, Pfannkoch C, Rogers YH,

Smith HO (2004) Environmental genome shotgun sequencing of the Sargasso Sea. Science 304:66–74

121. von Mering C, Jensen LJ, Snel B, Hooper SD, Krupp M, Foglierini M, Jouffre N, Huynen MA, Bork P (2005) STRING: known and predicted protein-protein associations, integrated and transferred across organisms. Nucleic Acids Res 33:D433–D437

122. Washietl S, Hofacker IL, Stadler PF (2005) Fast and reliable prediction of noncoding RNAs. Proc Natl Acad Sci USA 102:2454–2459

123. Waterston RH, Lindblad-Toh K, Birney E, Rogers J, Abril JF, Agarwal P, Agarwala R, Ainscough R, Alexandersson M, An P, Antonarakis SE, Attwood J, Baertsch R, Bailey J, Barlow K, Beck S, Berry E, Birren B, Bloom T, Bork P, Botcherby M, Bray N, Brent MR, Brown DG, Brown SD, Bult C, Burton J, Butler J, Campbell RD, Carninci P, Cawley S, Chiaromonte F, Chinwalla AT, Church DM, Clamp M, Clee C, Collins FS, Cook LL, Copley RR, Coulson A, Couronne O, Cuff J, Curwen V, Cutts T, Daly M, David R, Davies J, Delehaunty KD, Deri J, Dermitzakis ET, Dewey C, Dickens NJ, Diekhans M, Dodge S, Dubchak I, Dunn DM, Eddy SR, Elnitski L, Emes RD, Eswara P, Eyras E, Felsenfeld A, Fewell GA, Flicek P, Foley K, Frankel WN, Fulton LA, Fulton RS, Furey TS, Gage D, Gibbs RA, Glusman G, Gnerre S, Goldman N, Goodstadt L, Grafham D, Graves TA, Green ED, Gregory S, Guigo R, Guyer M, Hardison RC, Haussler D, Hayashizaki Y, Hillier LW, Hinrichs A, Hlavina W, Holzer T, Hsu F, Hua A, Hubbard T, Hunt A, Jackson I, Jaffe DB, Johnson LS, Jones M, Jones TA, Joy A, Kamal M, Karlsson EK, et al. (2002) Initial sequencing and comparative analysis of the mouse genome. Nature 420:520–562

124. Wellcome Trust Case Control Consortium (2007) Genomewide association study of 14,000 cases of seven common diseases and 3,000 shared controls. Nature 447:661–678

125. Wolfe KH, Sharp PM, Li WH (1989) Mutation rates differ among regions of the mammalian genome. Nature 337:283–285

126. Xie X, Lu J, Kulbokas EJ, Golub TR, Mootha V, Lindblad-Toh K, Lander ES, Kellis M (2005) Systematic discovery of regulatory motifs in human promoters and 3′ UTRs by comparison of several mammals. Nature 434:338–345

127. Yamada K, Lim J, Dale JM, Chen H, Shinn P, Palm CJ, Southwick AM, Wu HC, Kim C, Nguyen M, Pham P, Cheuk R, Karlin-Newmann G, Liu SX, Lam B, Sakano H, Wu T, Yu G, Miranda M, Quach HL, Tripp M, Chang CH, Lee JM, Toriumi M, Chan MM, Tang CC, Onodera CS, Deng JM, Akiyama K, Ansari Y, Arakawa T, Banh J, Banno F, Bowser L, Brooks S, Carninci P, Chao Q, Choy N, Enju A, Goldsmith AD, Gurjal M, Hansen NF, Hayashizaki Y, Johnson-Hopson C, Hsuan VW, Iida K, Karnes M, Khan S, Koesema E, Ishida J, Jiang PX, Jones T, Kawai J, Kamiya A, Meyers C, Nakajima M, Narusaka M, Seki M, Sakurai T, Satou M, Tamse R, Vaysberg M, Wallender EK, Wong C, Yamamura Y, Yuan S, Shinozaki K, Davis RW, Theologis A, Ecker JR (2003) Empirical analysis of transcriptional activity in the Arabidopsis genome. Science 302:842–846

128. Yu H, Luscombe NM, Lu HX, Zhu X, Xia Y, Han JD, Bertin N, Chung S, Vidal M, Gerstein M (2004) Annotation transfer between genomes: protein-protein interologs and protein-DNA regulogs. Genome Res 14:1107–1118

129. Yuan GC, Liu YJ, Dion MF, Slack MD, Wu LF, Altschuler SJ, Rando OJ (2005) Genome-scale identification of nucleosome positions in S. cerevisiae. Science 309:626–630

130. Yvert G, Brem RB, Whittle J, Akey JM, Foss E, Smith EN, Mackelprang R, Kruglyak L (2003) Trans-acting regulatory variation in Saccharomyces cerevisiae and the role of transcription factors. Nat Genet 35:57–64

131. Zeggini E, Weedon MN, Lindgren CM, Frayling TM, Elliott KS, Lango H, Timpson NJ, Perry JR, Rayner NW, Freathy RM, Barrett JC, Shields B, Morris AP, Ellard S, Groves CJ, Harries LW, Marchini JL, Owen KR, Knight B, Cardon LR, Walker M, Hitman GA, Morris AD, Doney AS, Burton PR, Clayton DG, Craddock N, Deloukas P, Duncanson A, Kwiatkowski DP, Ouwehand WH, Samani NJ, Todd JA, Donnelly P, Davison D, Easton D, Evans D, Leung HT, Spencer CC, Tobin MD, Attwood AP, Boorman JP, Cant B, Everson U, Hussey JM, Jolley JD, Knight AS, Koch K, Meech E, Nutland S, Prowse CV, Stevens HE, Taylor NC, Walters GR, Walker NM, Watkins NA, Winzer T, Jones RW, McArdle WL, Ring SM, Strachan DP, Pembrey M, Breen G, St Clair D, Caesar S, Gordon-Smith K, Jones L, Fraser C, Green EK, Grozeva D, Hamshere ML, Holmans PA, Jones IR, Kirov G, Moskvina V, Nikolov I, O'Donovan MC, Owen MJ, Collier DA, Elkin A, Farmer A, Williamson R, McGuffin P, Young AH, Ferrier IN, Ball SG, Balmforth AJ, Barrett JH, Bishop DT, Iles MM, Maqbool A, Yuldasheva N, Hall AS, Braund PS, Dixon RJ, Mangino M, Stevens S, Thompson JR, Bredin F, Tremelling M, et al. (2007) Replication of genome-wide association signals in UK samples reveals risk loci for type 2 diabetes. Science 316:1336–1341

132. Zhang Z, Hesselberth JR, Fields S (2007) Genome-wide identification of spliced introns using a tiling microarray. Genome Res 17:503–509

Fuzzy Logic

Lotfi A. Zadeh
Department of EECS, University of California,
Berkeley, USA

Article Outline

Glossary

Cointension A qualitative measure of proximity of meanings/input-output relations.

Extension principle A principle which relates to propagation of generalized constraints.

f-validity fuzzy validity.

Fuzzy if-then rule A rule of the form: if X is A then Y is B. In general, A and B are fuzzy sets.

Fuzzy logic (FL) A precise logic of imprecision, uncertainty and approximate reasoning.

Fuzzy logic gambit Exploitation of tolerance for imprecision through deliberate m-imprecisiation followed by mm-precisiation.

Fuzzy set A class with a fuzzy boundary.

Generalized constraint A constraint of the form X isr R, where X is the constrained variable, R is the constraining relation and r is an indexical variable which defines the modality of the constraint, that is, its semantics. In general, generalized constraints have elasticity.

Generalized constraint language A language generated by combination and propagation of generalized constraints.

Graduation Association of a scale of degrees with a fuzzy set.

Granuland Result of granulation.

Granular variable A variable which takes granules as variables.

Granulation Partitioning of an object/set into granules.

Granule A clump of attribute values drawn together by indistinguishability, equivalence, similarity, proximity or functionality.

Linguistic variable A granular variable with linguistic labels of granular values.

m-precision Precision of meaning.

mh-precisiand m-precisiand which is described in a natural language (human-oriented).

mm-precisiand m-precisiand which is described in a mathematical language (machine-oriented).

p-validity provable validity.

Precisiand Result of precisiation.

Precisiend Object of precisiation.

v-precision Precision of value.

Definition of the Subject

Viewed in a historical perspective, fuzzy logic is closely related to its precursor – fuzzy set theory [70]. Conceptually, fuzzy logic has a much broader scope and a much higher level of generality than traditional logical systems, among them the classical bivalent logic, multivalued logics, model logics, probabilistic logics, etc. The principal objective of fuzzy logic is formalization – and eventual mechanization – of two remarkable human capabilities. First, the capability to converse, communicate, reason and make ra-

tional decisions in an environment of imprecision, uncertainty, incompleteness of information, partiality of truth and partiality of possibility. And second, the capability to perform a wide variety of physical and mental tasks – such as driving a car in city traffic and summarizing a book – without any measurement and any computations.

A concept which has a position of centrality in fuzzy logic is that of a fuzzy set. Informally, a fuzzy set is a class with a fuzzy boundary, implying a gradual transition from membership to nonmembership. A fuzzy set is precisiated through graduation, that is, through association with a scale of grades of membership. Thus, membership in a fuzzy set is a matter of degree. Importantly, in fuzzy logic everything is or is allowed to be graduated, that is, be a matter of degree. Furthermore, in fuzzy logic everything is or is allowed to be granulated, with a granule being a clump of attribute-values drawn together by indistinguishability, equivalence, similarity, proximity or functionality. Graduation and granulation form the core of fuzzy logic. Graduated granulation is the basis for the concept of a linguistic variable – a variable whose values are words rather than numbers [73]. The concept of a linguistic variable is employed in almost all applications of fuzzy logic.

During much of its early history, fuzzy logic was an object of controversy stemming in part from the pejorative connotation of the term "fuzzy". In reality, fuzzy logic is not fuzzy. Basically, fuzzy logic is a precise logic of imprecision and uncertainty.

An important milestone in the evolution of fuzzy logic was the development of the concept of a linguistic variable and the machinery of fuzzy if-then rules [73,90]. Another important milestone was the conception of possibility theory [79]. Possibility theory and probability theory are complimentary. A further important milestone was the development of the formalism of computing with words (CW) [93]. Computing with words opens the door to a wide-ranging enlargement of the role of natural languages in scientific theories.

In the following, fuzzy logic is viewed in a nontraditional perspective. In this perspective, the cornerstones of fuzzy logic are graduation, granulation, precisisation and the concept of a generalized constraint. The concept of a generalized constraint serves to precisiate the concept of granular information. Granular information is the basis for granular computing (GrC) [2,29,37,81,91,92]. In granular computing the objects of computation are granular variables, with a granular value of a granular variable representing an imprecise and/or uncertain information about the value of the variable. In effect, granular computing is the computational facet of fuzzy logic. GrC and

CW are closely related. In coming years, GrC and CW are likely to play increasingly important roles in the evolution of fuzzy logic and its applications.

Introduction

Science deals not with reality but with models of reality. In large measure, scientific progress is driven by a quest for better models of reality.

In the real world, imprecision, uncertainty and complexity have a pervasive presence. In this setting, construction of better models of reality requires a better understanding of how to deal effectively with imprecision, uncertainty and complexity. To a significant degree, development of fuzzy logic has been, and continues to be, motivated by this need.

In essence, logic is concerned with formalization of reasoning. Correspondently, fuzzy logic is concerned with formalization of fuzzy reasoning, with the understanding that precise reasoning is a special case of fuzzy reasoning.

Humans have many remarkable capabilities. Among them there are two that stand out in importance. First, the capability to converse, communicate, reason and make rational decisions in an environment of imprecision, uncertainty, incompleteness of information, partiality of truth and partiality of possibility. And second, the capability to perform a wide variety of physical and mental tasks – such as driving a car in heavy city traffic and summarizing a book – without any measurements and any computations. In large measure, fuzzy logic is aimed at formalization, and eventual mechanization, of these capabilities. In this perspective, fuzzy logic plays the role of a bridge from natural to machine intelligence.

There are many misconceptions about fuzzy logic. A common misconception is that fuzzy logic is fuzzy. In reality, fuzzy logic is not fuzzy. Fuzzy logic deals precisely with imprecision and uncertainty. In fuzzy logic, the objects of deduction are, or are allowed to be fuzzy, but the rules governing deduction are precise. In summary, fuzzy logic is a precise system of reasoning, deduction and computation in which the objects of discourse and analysis are associated with information which is, or is allowed to be, imprecise, uncertain, incomplete, unreliable, partially true or partially possible. For illustration, here are a few simple examples of reasoning in which the objects of reasoning are fuzzy.

First, consider the familiar example of deduction in Aristotelian, bivalent logic.

> all men are mortal
> Socrates is a man
> _____
> Socrates is mortal

In this example, there is no imprecision and no uncertainty. In an environment of imprecision and uncertainty, an analogous example – an example drawn from fuzzy logic – is

> most Swedes are tall
> Magnus is a Swede
> _____
> it is likely that Magnus is tall

with the understanding that Magnus is an individual picked at random from a population of Swedes. To deduce the answer from the premises, it is necessary to precisiate the meaning of "most" and "tall," with "likely" interpreted as a fuzzy probability which, as a fuzzy number, is equal to "most". This simple example points to a basic characteristic of fuzzy logic, namely, in fuzzy logic precisiation of meaning is a prerequisite to deduction. In the example under consideration, deduction is contingent on precisiation of "most", "tall" and "likely". The issue of precisiation has a position of centrality in fuzzy logic.

In fuzzy logic, deduction is viewed as an instance of question-answering. Let I be an information set consisting of a system of propositions $p_1, \ldots, p_n$, $I = S(p_1, \ldots, p_n)$. Usually, I is a conjunction of $p_1, \ldots, p_n$. Let q be a question. A question-answering schema may be represented as

> I
> q
> _____
> $\text{ans}(q/I)$

where $\text{ans}(q/I)$ denotes the answer to q given I. The following examples are instances of the deduction schema

$I:$ most Swedes are tall
Magnus is a Swede
$q:$ what is the probability that Magnus is tall? (1)
_____
$\text{ans}(q/I)$ is likely, likely=most

$I:$ most Swedes are tall
$q:$ what fraction of Swedes are not tall? (2)
_____
$\text{ans}(q/I)$ is $(1 - \text{most})$

$I:$ most Swedes are tall
$q:$ what fraction of Swedes are short? (3)
_____
$\text{ans}(q/I)$

$I:$ most Swedes are tall
$q:$ what is the average height of Swedes? (4)
_____
$\text{ans}(q/I)$

$I:$ a box contains balls of various sizes
most are small
there are many more small balls than large balls
$q:$ what is the probability that a ball
drawn at random is neither large nor small?
_____
$\text{ans}(q/I)$.

(5)

In these examples, rules of deduction in fuzzy logic must be employed to compute $\text{ans}(q/I)$. For (1) and (2) deduction is simple. For (3)–(5) deduction requires the use of what is referred to as the extension principle [70,75]. This principle is discussed in Sect. "The Concept of a Generalized Constraint".

A less simple example of deduction involves interpolation of an imprecisely specified function. Interpolation of imprecisely specified functions, or interpolative deduction for short, plays a pivotal role in many applications of fuzzy logic, especially in the realm of control.

For simplicity, assume that f is a function from reals to reals, $Y = f(X)$. Assume that what is known about f is a collection of input-output pairs of the form

$$^*f = ((^*a_1, ^*b_1), \ldots, (^*a_n, ^*b_n)) ,$$

where $^*a$ is an abbreviation of "approximately a". Such a collection is referred to as a fuzzy graph of f [74]. A fuzzy graph of f may be interpreted as a summary of f. In many applications, a fuzzy graph is described as a collection of fuzzy if-then rules of the form

if X is $^*a_i$ then Y is $^*b_i$, $i = 1, \ldots, n$.

Let a be a value of X. Viewing $^*f$ as an information set, I, interpolation of f may be expressed as a question-answering schema

> $I: ^*f$
> $q: ^*f(^*a)$
> _____
> $\text{ans}(q/(^*f, ^*a))$

A very simple example of interpolative deduction is the following. Assume that $^*a_1, ^*a_2, ^*a_3$ are labeled small, medium and large, respectively. A fuzzy graph of f may be expressed as a calculus of fuzzy if-then rules.

$^*f:$ if X is small then Y is small
if X is medium than Y is large
if X is large then Y is small.

Given a value of X, $X = a$, the question is: What is $^*f(^*a)$? The so-called Mamdani rule [39,74,78] provides an answer to this question.

More generally, in an environment of imprecision and uncertainty, fuzzy if-then rules may be of the form

$$\text{if} \quad X \text{ is } {}^*a_i, \quad \text{then usually } (Y \text{ is } {}^*b_i), \quad i = 1, \dots, n.$$

Rules of this form endow fuzzy logic with the capability to model complex causal dependencies, especially in the realms of economics, social systems, forecasting and medicine.

What is not widely recognized is that fuzzy logic is more than an addition to the methods of dealing with imprecision, uncertainty and complexity. In effect, fuzzy logic represents a paradigm shift. More specifically, it is traditional to associate scientific progress with progression from perceptions to numbers. What fuzzy logic adds to this capability are four basic capabilities.

(a) Nontraditional. Progression from perceptions to precisiated words
(b) Nontraditional. Progression from unprecisiated words to precisiated words
(c) Countertraditional. Progression from numbers to precisiated words
(d) Nontraditional. Computing with words (CW)/NL-computation.

These capabilities open the door to a wide-ranging enlargement of the role of natural languages in scientific theories.

Our brief discussion of deduction in the context of fuzzy logic is intended to clarify the nature of problems which fuzzy logic is designed to address. The principal concepts and techniques which form the core of fuzzy logic are discussed in the following. Our discussion draws on the concepts and ideas introduced in [102].

Conceptual Structure of Fuzzy Logic

There are many logical systems, among them the classical, Aristotelian, bivalent logic, multivalued logics, model logics, probabilistic logic, logic, dynamic logic, etc. What differentiates fuzzy logic from such logical systems is that fuzzy logic is much more than a logical system.

The point of departure in fuzzy logic is the concept of a fuzzy set. Informally, a fuzzy set is a class with a fuzzy boundary, implying that, in general, transition from membership to nonmembership in a fuzzy set is gradual rather than abrupt. A set is a class with a crisp boundary (Fig. 1). A set, A, in a space U, $U = \{u\}$, is precisiated through association with a characteristic function which maps U into $\{0, 1\}$. More generally, a fuzzy set, A, is precisiated through graduation, that is, through association with A of a membership function, μ_A – a mapping from U to a grade of
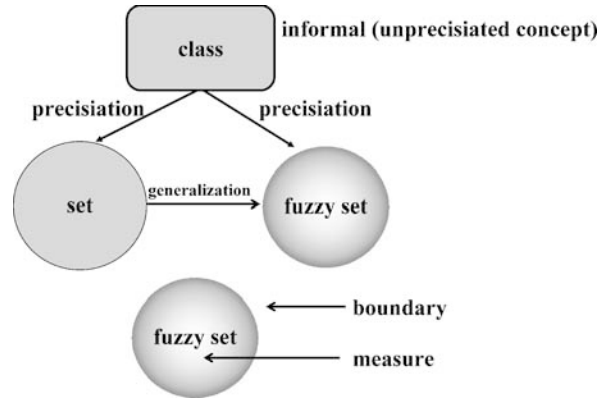
Fuzzy Logic, Figure 1
The concepts of a set and a fuzzy set are derived from the concept of a class through precisiation. A fuzzy set has a fuzzy boundary. A fuzzy set is precisiated through graduation

membership space, G, with $\mu_A(u)$ representing the grade of membership of u in A. In other words, membership in a fuzzy set is a matter of degree. A familiar example of graduation is the association of Richter scale with the class of earthquakes. A fuzzy set is basic if G is the unit interval. More generally, G may be a partially ordered set. L-fuzzy sets [23] fall into this category. A basic fuzzy set is of Type 1. A fuzzy set, A, is of Type 2 if $\mu_A(u)$ is a fuzzy set of Type 1. Recursively, a fuzzy set, A, is of Type n if $\mu_A(u)$ is a fuzzy set of Type $n - 1$, $n = 2, 3, \dots$ [75]. Fuzzy sets of Type 2 have become an object of growing attention in the literature of fuzzy logic [40]. Unless stated to the contrary, a fuzzy set is assumed to be of Type 1 (basic).

Note. A clarification is in order. Consider a concatenation of two words, A and B, with A modifying B, e. g. A is an adjective and B is a noun. Usually, A plays the role of an s-modifier, that is, a modifier which specializes B in the sense that AB is a subset of B, as in convex set. In some instances, however, A plays the role of a g-modifier, that is, a modifier which generalizes B. In this sense, fuzzy in fuzzy set, fuzzy logic and, more generally, in fuzzy B, is a g-modifier. Examples: fuzzy topology, fuzzy measure, fuzzy arithmetic, fuzzy stability, etc. Many misconceptions about fuzzy logic are rooted in incorrect interpretation of fuzzy as an s-modifier.

What is important to note is that (a) $\mu_A(u)$ is name-based (extensional) if u is the name of an object in U, e. g., $\mu_{\text{middle-aged}}(\text{Vera})$; (b) $\mu_A(u)$ is attribute-based (intensional) if the grade of membership is a function of an attribute of u, e. g., age; and (c) $\mu_A(u)$ is perception-based if u is a perception of an object, e. g., Vera's grade of membership in the class of middle-aged women, based on her appearance, is 0.8.

It should be observed that a class of objects, A, has two basic attributes: (a) the boundary of A; and (b) the cardinality (count) or, more generally, the measure of A (Fig. 1). In this perspective, fuzzy set theory is, in the main, boundary-oriented, while probability theory is, in the main, measure-oriented. Fuzzy logic is, in the main, both boundary- and measure-oriented. The concept of a membership function is the centerpiece of fuzzy set theory.

With the concept of a fuzzy set as the point of departure, different directions may be pursued, leading to various facets of fuzzy logic. More specifically, the following are the principal facets of fuzzy logic: the logical facet, FLl; the fuzzy-set-theoretic facet, FLs, the epistemic facet, Fle; and the relational facet, FLr (Fig. 2).

The logical facet of FL, FLl, is fuzzy logic in its narrow sense. FLl may be viewed as a generalization of multivalued logic. The agenda of FLl is similar in spirit to the agenda of classical logic [17,22,25,44,45].

The fuzzy-set-theoretic facet, FLs, is focused on fuzzy sets. The theory of fuzzy sets is central to fuzzy logic. Historically, the theory of fuzzy sets [70] preceded fuzzy logic [77]. The theory of fuzzy sets may be viewed as an entry to generalizations of various branches of mathematics, among them fuzzy topology, fuzzy measure theory, fuzzy graph theory, fuzzy algebra and fuzzy differential equations. Note that fuzzy X is a fuzzy-set-theory-based or, more generally, fuzzy-logic-based generalization of X.

The epistemic facet of FL, FLe, is concerned with knowledge representation, semantics of natural languages and information analysis. In FLe, a natural language is viewed as a system for describing perceptions. An important branch of FLe is possibility theory [15,79,83]. Another important branch of FLe is the computational theory of perceptions [93,94,95].

The relational facet, FLr, is focused on fuzzy relations and, more generally, on fuzzy dependencies. The concept of a linguistic variable – and the associated calculi of fuzzy

Fuzzy Logic, Figure 2
Principal facets of fuzzy logic (FL). The nucleus of fuzzy logic is the concept of a fuzzy set
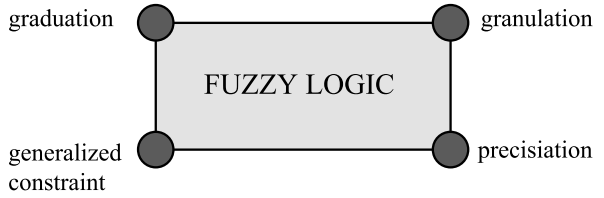
Fuzzy Logic, Figure 3
The cornerstones of a nontraditional view of fuzzy logic

if-then rules – play pivotal roles in almost all applications of fuzzy logic [1,3,6,8,12,13,18,26,28,32,36,40,48,52,65,66, 67,69].

The cornerstones of fuzzy logic are the concepts of graduation, granulation, precisiation and generalized constraints (Fig. 3). These concepts are discussed in the following.

The Basics of Fuzzy Set Theory

The grade of membership, $\mu_A(u)$, of u in A may be interpreted in various ways. It is convenient to employ the proposition p: Vera is middle-aged, as an example, with middle-age represented as a fuzzy set shown in Fig. 4. Among the various interpretations of p are the following.

Assume that q is the proposition: Vera is 43 years old; and r is the proposition: the grade of membership of Vera in the fuzzy set of middle-aged women is 0.8.

(a) The truth value of p given r is 0.8.
(b) The possibility of q given p and r is 0.8.
(c) The degree to which the concept of middle-age has to be stretched to apply to Vera is (1–0.8).
(d) 80% of voters in a voting group vote for p given q and r.
(e) The probability of p given q and r is 0.8.

Of these interpretations, (a)–(c) are most compatible with intuition.

If A and B are fuzzy sets in U, then their intersection (conjunction), $A \cap B$, and union (disjunction), $A \cup B$, are defined as

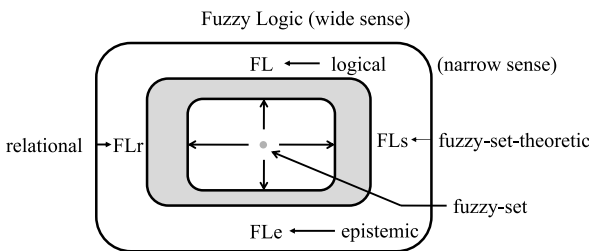$$\mu_{A \cap B}(u) = \mu_A(u) \wedge \mu_B(u), \quad u \in U$$
$$\mu_{A \cup B}(u) = \mu_A(u) \vee \mu_B(u), \quad u \in U$$

where $\wedge = \min$ and $\vee = \max$. More generally, conjunction and disjunction are defined through the concepts of t-norm and t-conorm, respectively [48].

When U is a finite set, $U = \{u_1, \ldots, u_n\}$, it is convenient to represent A as a union of fuzzy singletons,

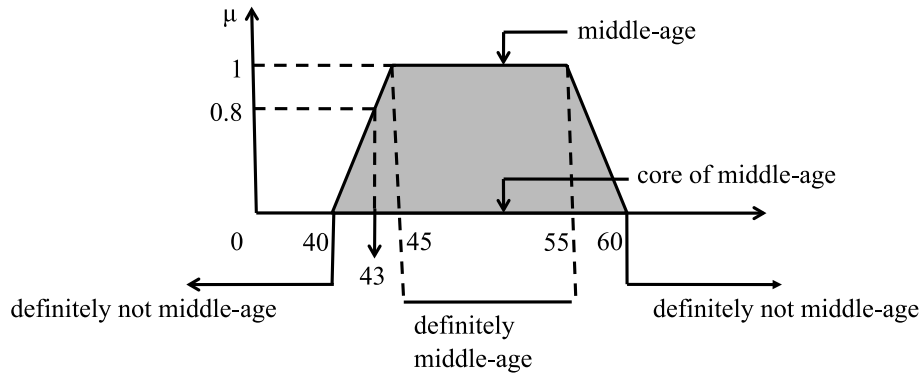$$\mu_A(u_i)/u_i, \quad i = 1, \ldots, n.$$

Specifically, $A = \mu_A(u_1)/u_1 + \cdots + \mu_A(u_n)/u_n$

- **Imprecision of meaning = elasticity of meaning**

- **Elasticity of meaning = fuzziness of meaning**

Example: middle-age

Fuzzy Logic, Figure 4
Precisiation of middle-age through graduation

in which $+$ denotes disjunction. More compactly,

$$A = \Sigma_i \mu_A(u_i)/u_i \,, \quad i = 1, \ldots, n \,.$$
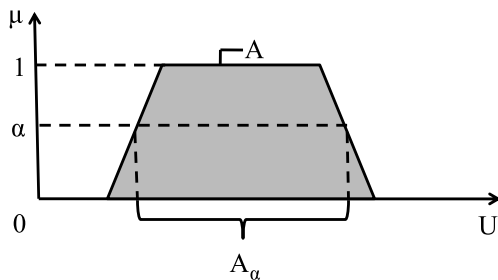
When U is a continuum, A may be expressed as

$$\int_U \mu_A(u)/u \,.$$

A basic concept in fuzzy set theory is that of a level set [70], commonly referred to as an α-cut [48]. Specifically, if A is a fuzzy set in U, then an α-cut, A_α, is defined as (Fig. 5)

$$A_\alpha = \{u \mid \mu_A(u) \geqslant \alpha\} \,, \quad 0 < \alpha \leqslant 1 \,.$$

The core of A is the α-cut with $\alpha = 1$.

A fuzzy set, A, is a fuzzy interval if all of its α-cuts are intervals. A fuzzy set, A, is convex if all of its α-cuts are convex [70].

In most practical applications, a membership function is assumed to have a simple form, e. g. a triangle or a trapezoid (Fig. 6). For convenience, triangular and trapezoidal fuzzy sets are abbreviated as tr-sets and tp-sets, respectively. It should be noted that a singleton is associated with one parameter, an interval with two parameters, a tr-set with three parameters and a tp-set with four parameters. The number of parameters may be interpreted as the number of degrees of freedom. In general, the number of degrees of freedom is covariant with closeness of approximation.

A concept which plays an important role in fuzzy set theory is that of cardinality, that is, the count of elements in a fuzzy set [25,51,64,86]. Basically, there are two ways in which cardinality can be defined: (a) crisp (scaler) cardinality and (b) fuzzy (stratified) cardinality. In the case of (a), the count of elements in a fuzzy set is a crisp number; in the case of (b) it is a fuzzy set.

More specifically, consider a fuzzy set, A, defined in $U = \{u_1, \ldots, u_n\}$ through its membership function,

Fuzzy Logic, Figure 5
Definition of α-cut
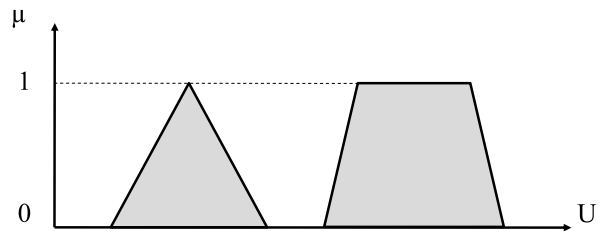
Fuzzy Logic, Figure 6
Triangular and trapezoidal fuzzy sets. tr-sets and tp-sets

$\mu_A(u)$. The sigma-count of A is defined as

$$\Sigma\,\mathrm{Count}(A) = \sum_{i=1}^{n} \mu_A(u_i)\,.$$

If A and B are fuzzy sets defined in U, then the relative sigma-count, $\Sigma\,\mathrm{Count}(A/B)$, is defined as

$$\Sigma\,\mathrm{Count}(A/B) = \frac{\Sigma\,\mathrm{Count}(A \cap B)}{\Sigma\,\mathrm{Count}(B)}$$

$$= \frac{\sum\limits_{i=1}^{n} \mu_A(u_i) \wedge \mu_B(u_i)}{\sum\limits_{i=1}^{n} \mu_B(u_i)}\,,$$

where $\wedge = \min$, and summations are arithmetic.

A stratified count, $\mathrm{S\,Count}(A)$, is a fuzzy set. As such, it is more informative then $\Sigma\,\mathrm{Count}(A)$ but has the disadvantage of greater complexity. A stratified count is defined in terms of α-cuts of A [86]. Specifically, let $A_{\alpha 1}, \ldots, A_{\alpha n}$ be α-cuts of A, with $\alpha_1 < \alpha_2 < \cdots < \alpha_n \leq 1$. Then

$$\mathrm{S\,Count}(A) = \alpha_1/\mathrm{Count}(A_{\alpha 2}) + \cdots + \alpha_n/\mathrm{Count}(A_{\alpha n})$$

or equivalently,

$$\mathrm{S\,Count}(A) = \{(\alpha_i, A_{\alpha i})\}\,, \quad i = 1, \ldots, n\,.$$

As a simple illustration, consider the fuzzy set

$$A = 0.4/u_1 + 0.8/u_2 + 1/u_3 + 0.9/u_4 + 0.3/u_5\,.$$

In this case,

$$A_{0.3} = \{u_1, u_2, u_3, u_4, u_5\}\,, \quad \mathrm{Count}(A_{0.3}) = 5$$
$$A_{0.4} = \{u_1, u_2, u_3, u_4\}\,, \quad \mathrm{Count}(A_{0.4}) = 4$$
$$A_{0.8} = \{u_2, u_3, u_4\}\,, \quad \mathrm{Count}(A_{0.8}) = 3$$
$$A_{0.9} = \{u_3, u_4\}\,, \quad \mathrm{Count}(A_{0.9}) = 2$$
$$A_1 = \{u_3\}\,, \quad \mathrm{Count}(A_1) = 1$$

$$\mathrm{S\,Count}(A) = 1/1 + 0.9/2 + 0.8/3 + 0.4/4 + 0.3/5$$

$$\Sigma\,\mathrm{Count}(A) = 3.4\,.$$

The concept of a membership function has a position of centrality in fuzzy logic. In this context, a natural question is: How can a membership function be constructed? There are three basic approaches: graduation through declaration; graduation through composition/deduction; and graduation through exemplification/ostention/elicitation.

Graduation through declaration is employed in many applications of fuzzy logic, especially in the realm of control systems. In this mode of graduation, a membership function is declared (specified) by the designer of a system. An example is the Honda fuzzy logic transmission (Fig. 7).

Declarative graduation has the potential for important applications in the realm of precisiation/standardization of many concepts and terms in various fields of science, especially in the realms of medicine and economics. A very simple example is standardization of the meaning of normal temperature, mild fever, high fever, etc. It is convenient to standardize the meaning of such terms through


Fuzzy Set
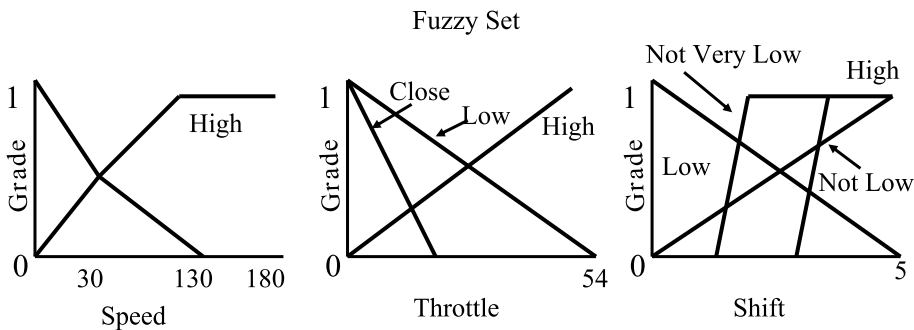
Control fuzzy if-then rules:

1. If (speed is low) and (shift is high) then (-3)
2. If (speed is high) and (shift is low) then (+3)
3. If (throt is low) and (speed is high) then (+3)
4. If (throt is low) and (speed is low) then (+1)
5. If (throt is high) and (speed is high) then (-1)
6. If (throt is high) and (speed is low) then (-3)

Fuzzy Logic, Figure 7
Declared membership functions and associated fuzzy if-then rules in Honda's fuzzy logic transmission
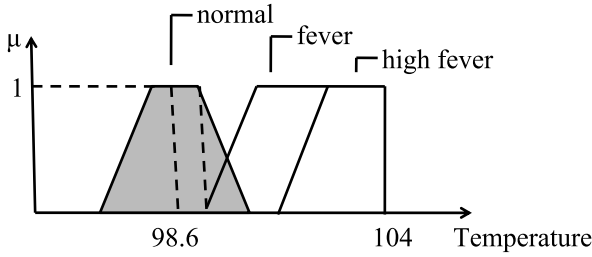
Fuzzy Logic, Figure 8
Declarative graduation/standardization of normal temperature, fever and high fever

the use of trapezoidal membership functions, that is, tp-sets (Fig. 8).

In many cases, a membership function is associated with a number of adjustable parameters, e. g. the four parameters which define a tp-set. The parameters are adjusted through experimentation or self-learning. For this purpose, techniques drawn from neurocomputing and evolutionary computing are commonly employed [54,55].

In construction through composition/deduction, a membership function is composed from other membership functions. Composition/deduction may be simple e. g. a combination of conjunctive or disjunctive operations. More generally, composition/deduction is associated with a chain of deductions which involve generalized constraint propagation. Propagation of generalized constraints is discussed in Sect. "The Concept of a Generalized Constraint".

A special case of composition which plays an important role in fuzzy set theory involves the concept of a convex combination [70]. More specifically, let $A_1, \ldots, A_n$ be a collection of crisp or fuzzy sets in U. Let $a_1, \ldots, a_n$ be numbers in [0,1] which add up to unity. A fuzzy set, A, is a convex combination of the A_i, $i = 1, \ldots, n$, if

$$A = a_1 A_1 + \cdots + a_n A_n \, ,$$

with the understanding that

$$\mu_A(u) = a_1 \mu_{A_1}(u) + \cdots + a_n \mu_{A_n}(u) \, .$$

What is important to note is that a convex combination of crisp sets is a fuzzy set. If the coefficients $a_1, \ldots, a_n$ are interpreted as probabilities, then A may be interpreted as the expected value of a random set. This connection between fuzzy sets and random sets has been an object of considerable attention in the literature of fuzzy logic [24,46,47].

Underlying graduation through exemplification is the remarkable human capability to rank-order perceptions. It

should be noted that humans learn the meaning of terms and concepts mostly through exemplification.

It is convenient to employ an example to describe graduation through exemplification. Assume that A tells B that Vera is middle-aged. B can elicit A's meaning of middle-aged by asking A to mark on the scale [0, 1] the degree to which a particular age, say 43, fits A's meaning of middle-aged. The process is repeated for various values of age. Eventually, the collected data are employed to approximate to A's meaning of middle-aged by a trapezoidal membership function. In the case of fuzzy sets of Type 2, the mark is a fuzzy point.
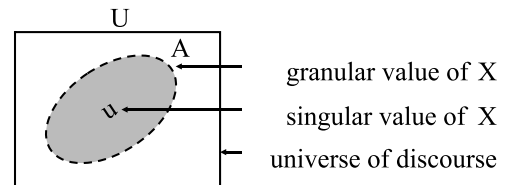
A basic feature of fuzzy logic is: In fuzzy logic everything is or is allowed to be graduated.

The Concept of Granulation

The concept of granulation is unique to fuzzy logic and plays a pivotal role in its applications [81,91]. The concept of granulation is inspired by the way in which humans deal with imprecision, uncertainty and complexity.

Granulation is rooted in the concept of a granule. Informally, a granule, G, in a universe of discourse, U, is a clump of elements of U which are drawn together by indistinguishability, equivalence, similarity, proximity or functionality (Fig. 9). For example, an interval is a granule. So is a Gaussian distribution, and so is a fuzzy interval. A granule, G, is precisiated through association of G with a generalized constraint – a concept which will be defined in Sect. "The Concept of a Generalized Constraint". The concept of a generalized constraint is more general than that of a membership function.

A singular variable takes singletons in U as values. A granular variable, X, is a variable which takes granules as values. For example, age is a granular variable if its val-

	singular	granular
unemployment	7.3%	high
temperature	102.5	very high
blood pressure	160/80	high

Fuzzy Logic, Figure 9
Singular and granular values

ues are young, middle-aged and old. A linguistic variable is a granular variable whose granular values carry linguistic labels (Fig. 10). The concept of a linguistic variable [73,75] is employed in almost all applications of fuzzy logic.

Granulation is an operation which maps singletons in U into granules. Granulation applied to a singular variable, X, transforms X into a granular variable, $^*X$. For example, if age is a real-valued variable taking values in the interval [0, 120], granulation of X transforms X into a linguistic variable, $^*X$, with values young, middle-aged and old. When convenient, the result of granulation is referred to as the granuland.

Granulation of a set, A, results in a partition of A into granules. For example, granulation applied to the interval [0, l20] results in the granules young, middle-aged and old.

The membership function of a fuzzy set takes values in the interval [0, 1]. Not infrequently, the grade of membership is not known precisely. In this case, it may be expedi-

ent to granulate the interval [0, 1], with the granular values of membership being zero, low, medium, high and 1. Such granular values of membership are particularly useful in dealing with fuzzy sets of Type 2.

Granulation may be applied to arbitrarily complex objects. Application of granulation to an expression involves replacement of one or more singular variables in the expression with granular variables. For example, if

$$Z = X + Y$$

is an arithmetic expression, then its granuland may be expressed as
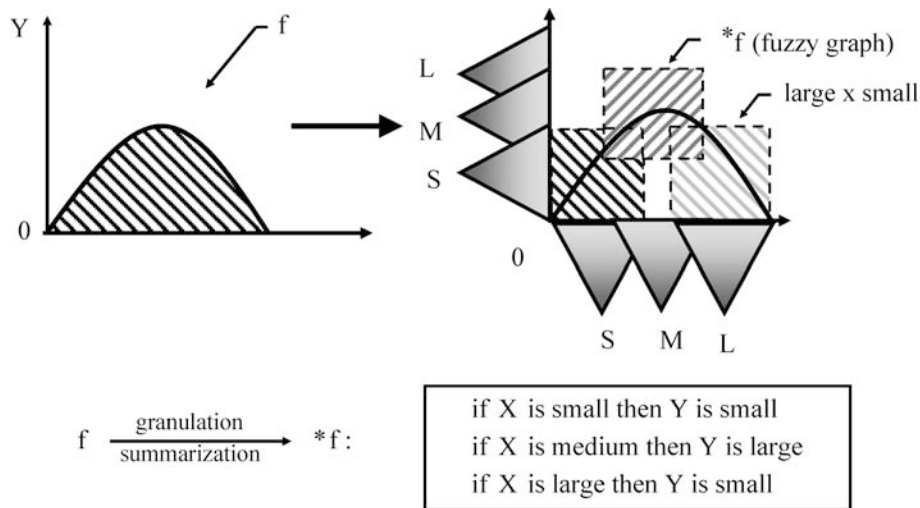
$$^*Z = \,^*X + \,^*Y$$

in which the starred variables are granular variables. In this sense, interval arithmetic may be viewed as the result of granulation of numerical arithmetic. Figure 11 shows an application of granulation to a function, f. The result of
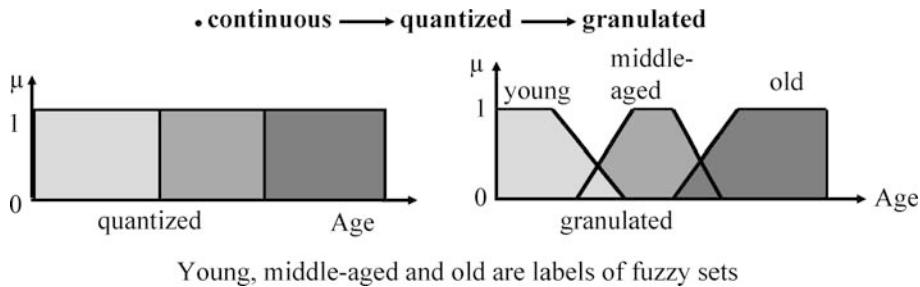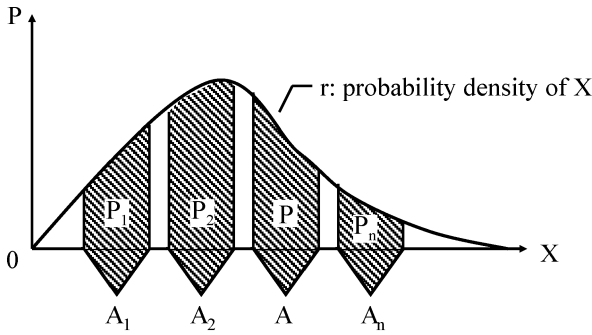

Fuzzy Logic, Figure 10
Granulation of age; young, middle-aged and old are linguistic (granular) values of age

Fuzzy Logic, Figure 11
Granulation of a function. *S* (small), *M* (medium) and *L* (large) are fuzzy sets. The granuland of f, $^*f$, may be viewed as a summary of f

p_i is P_i : granular value of p_i , i=1, …, n
(P_i , A_i) , i=1, …, n   are given
A is given
$(?P, A)$

Fuzzy Logic, Figure 12
Granulation and interpolation of a probability distribution

granulation, $^*f$, is the fuzzy graph of f [74,78]. The fuzzy graph of f may be described as a collection of fuzzy if-then rules; $^*f$ may be viewed as a summary of f. Describing a function as a collection of fuzzy if-then rules may be regarded as a form of information compression.

Application of granulation to probability distributions is illustrated in Fig. 12. The granules of X play the role of fuzzy events and the P_i are their granular probabilities [99].

It should be pointed out that in the case of probability distributions it is necessary to differentiate between granular probability distributions and granule-valued probabil-
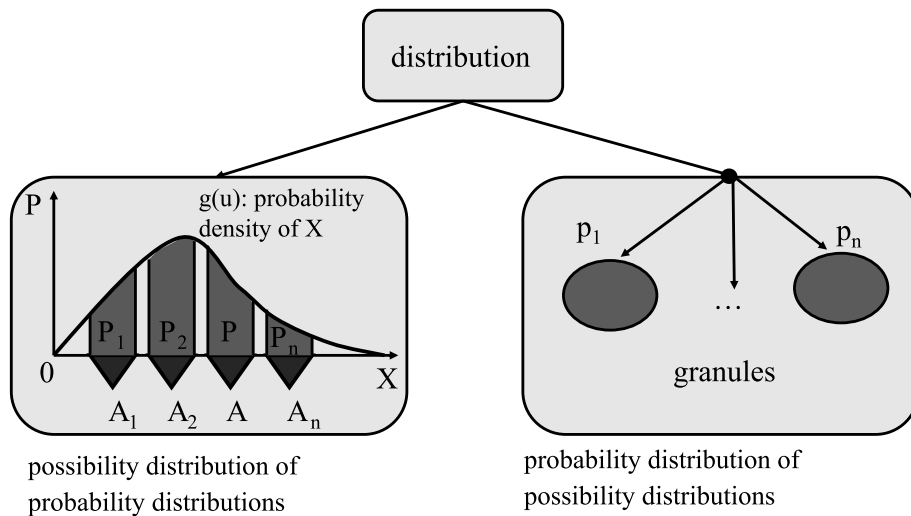
ity distributions [96,101] (Fig. 13). It should be noted that there is a connection between the concept of a granular probability distribution and the notion of Perfilieva transform [49].

An instance of a granule-valued distribution is a random set. There is a close connection between granule-valued distributions and the Dempster–Shafer theory of evidence [11,56,57]. More about granular and granule-valued distributions will be said in Sect. "The Concept of a Generalized Constraint".

In addition to the concepts of graduation and granulation there are two basic concepts which play important roles in fuzzy logic. These are the concepts of precisiation [87,97] and generalized constraint [89,93]. These concepts along with the concepts of graduation and granulation form the cornerstones of fuzzy logic. The concepts of precisiation and cointensive precisiation are discussed in the following section. The concept of a generalized constraint is discussed in Sect. "The Concept of a Generalized Constraint".

The Concepts of Precisiation and Cointensive Precisiation

There are not many concepts in science that are as pervasive as the concept of precision. There is an enormous literature. And yet, there are some important facets of the concept of precision which have received little if any attention. In particular, an issue that appears to have been overlooked relates to the need for differentiation between two forms of precision: precision of value, v-precision,

possibility distribution of
probability distributions

probability distribution of
possibility distributions

Fuzzy Logic, Figure 13
Granular vs. granule-valued distributions

and precision of meaning, m-precision [101]. Specifically, consider a variable, X, whose value is not known precisely. In this event, the proposition $a \leqslant X \leqslant b$, where a and b are precisely specified numbers, is v-imprecise and m-precise. Similarly, the proposition: X is a normally distributed random variable with mean a and variance b, is v-imprecise and m-precise. On the other hand, the proposition: X is small, is both v-imprecise and m-imprecise if small is not defined precisely. If small is defined precisely as a fuzzy set, then the proposition in question is v-imprecise and m-precise.

Informally, precisiation is an operation which transforms an object, p, into another object, p^*, which is more precisely defined, in some specified sense, than p. The reverse applies to imprecisiation. In the realm of our discourse p is a proposition, predicate, question, command or, more generally, a linguistic expression which has a semantic identity. Unless stated to the contrary, p will be assumed to be a proposition. For convenience, the object and the result of precisiation are referred to as precisiend and precisiand, respectively (Fig. 14).
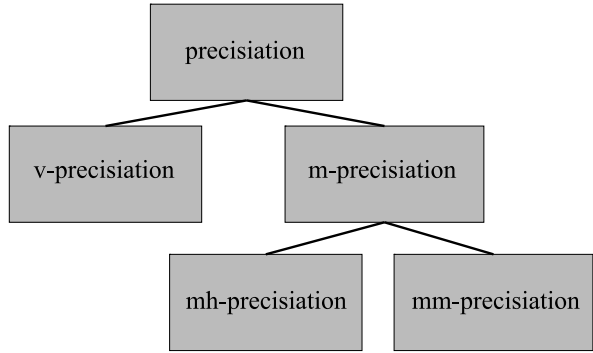
As in the case of precision/imprecision, there is a need for differentiation between v-precisiation/imprecisiation and m-precisiation/imprecisiation. Example:

$$X = 5 \xrightarrow[\text{m-imprecisiation}]{\text{v-imprecisiation}} X \text{ is small}$$

$$X \text{ is small} \xrightarrow{\text{m-precisiation}} X \text{ is small}$$

$$\text{(small is defined as a fuzzy set)} \,.$$

It should be noted that data compression and summarization are forms of v-imprecisiation.

Fuzzy Logic, Figure 15
Modalities of precisiation. mh-precisiation = nonmathematical precisiation of meaning; mm-precisiation = mathematical precisiation of meaning

In the case of m-precisiation, there is a need for additional differentiation between m-precisiation which is human-oriented (non-mathematical), or mh-precisiation for short, and m-precisiation which is machine-oriented (mathematical), or mm-precisiation for short (Fig. 15). In this sense, a dictionary definition is a form of mh-precisiation, with the definiens and the definiendum playing the roles of precisiend and precisiand, respectively. A mathematical definition of a concept, say stability, is an instance of mm-precisiation.

A convenient illustration of mh-precisiation and mm-precisiation is provided by the concept of "bear market".

mh-precisiand declining stock market with expectation of further decline

**mm-precisiand** 30 percent decline after 50 days, or a 13 percent decline after 145 days (Robert Shuster).
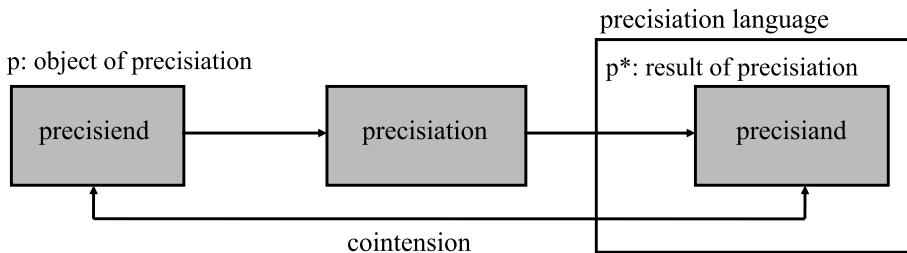
precisiand = model of meaning
cointension = measure of closeness of meanings of p and p*
A precisiend may have many precisiands.

precisiation = translation into precisiation language

Fuzzy Logic, Figure 14
Basic concepts relating to precisiation and cointension

P: most tall Swedes   **P(A) is?**
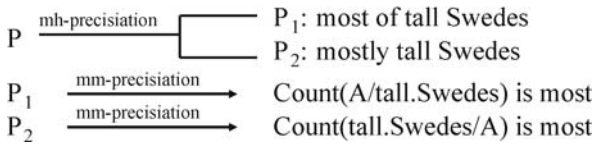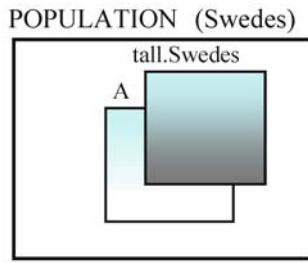
Fuzzy Logic, Figure 16
Disambiguation of P: most tall Swedes

It is of interest to note that disambiguation is a form of m-precisiation. As an illustration, alternative interpretations of the predicate, P: Most tall Swedes, are shown in Fig. 16.

So far as v-imprecisiation is concerned, there is a need for differentiation between (a) v-imprecisiation which is forced; and (b) v-imprecisiation which is deliberate. To illustrate, if the value of X is described imprecisely because the precise value of X is not known, then what is involved is forced v-imprecisiation. On the other hand, if the value of X is described imprecisely even though the precise value of X is known, then v-imprecisiation is deliberate.

What is the point of deliberate v-imprecisiation? The rationale is that in many cases precision carries a cost. In such cases, deliberate v-imprecisiation serves a useful purpose if it provides a way of reducing cost. Familiar examples of deliberate v-imprecisiation are data compression and summarization. As will be seen at a later point, deliberate v-imprecisiation underlies the fuzzy logic gambit (Sect. "The Concept of a Generalized Constraint"). The fuzzy logic gambit plays an important role in many applications of fuzzy logic, especially in the realm of consumer products – a realm in which cost is an important parameter.

A preciend may be precisiated in a large, perhaps unbounded, number of ways. As an illustration, Fig. 17a and b shows some of the simpler mm-precisiands of the predicate "approximately a," or $^*a$ for short. Note that the simplest mm-precisiand is a. This very simple mode of mm-precisiation is widely employed in many fields of science. Probability theory is a case in point. Most real-world probabilities are based on perceptions rather than on measurements. Perceptions are intrinsically imprecise. As a consequence, so are most real-world probabilities. And yet, in most computations involving probabilities, probabilities are treated as exact numbers. For example, *0.7 is treated as 0.7000.

The Concept of Cointension
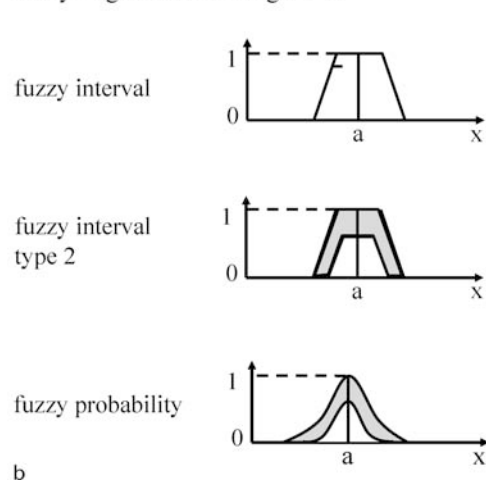
Let p^* be a precisiand of p. It is expedient to associate with p^* two basic metrics: (a) cointension – a qualitative mea-
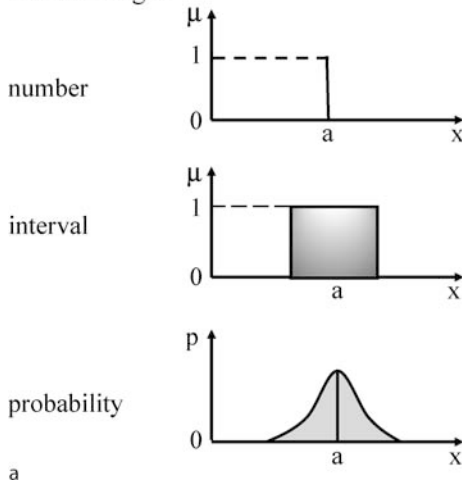
Fuzzy Logic, Figure 17
a Alternative modes of mm-precisiation of "approximately a," $^*a$, within the framework of bivalent logic; **b** Alternative modes of mm-precisiation of "approximately a," $^*a$, within the framework of fuzzy logic

sure of the proximity of the meanings of p and p^*; and (b) the computational complexity of p^*. In general, cointension and computational complexity are covariant in the sense that an increase in the cointension of p^* is associated with an increase in the computational complexity of p^*.

Cointension is a new term which is in need of clarification [101]. In logic, intension and extension are defined, respectively, as attribute-based meaning, or i-meaning for short, and name-based meaning, or e-meaning for short [5,10,35]. For our purposes, it is helpful to interpret attribute-based as measurement-based and, name-based as perception-based. What this means is that a precisiend, p, is viewed as a perception of a concept, while its mm-precisiand, p^*, is viewed as a measurement-based definition of p. For example, in the case of the concept of bear market, we have

mm-precisiand 30 percent decline after 50 days, or a 13 percent decline after 145 days (Robert Shuster).

More concretely, if p^* is an mm-precisiand of p, then the cointension of p^* in relation to p, $C(p^*, p)$, is a qualitative measure of the degree of proximity of the i-meanings of p^* and p, or the proximity of the i-meaning of p^* and the e-meaning of p. p^* is cointensive if the degree of proximity is high. In the case of the bear market example, cointension is a measure of the degree to which the mm-precisiand fits our perception of "bear market". Although this degree cannot be assessed precisely, it is evident that the degree is not high.

It should be noted that there is a close analogy between the concept of mm-precisiation and mathematical modeling. More specifically, the analog of mm-precisiation is mathematical modeling; the analog of precisiend is the object of modeling; the analog of precisiand is the model; and the analog of meaning is the input/output relation.

In the context of modeling, cointension is a measure of proximity of the input/output relations of the object of modeling and the model. A model is cointensive if its proximity is high.

The concept of cointension highlights an important issue. Specifically, what should be noted is that, in general, mm-precisiation of p is not the final objective. What matters is the cointension of an mm-precisiand of p. In general, what are sought are mm-precisiands which have high cointension. In other words, the desideratum in not merely mm-precisiation but, more importantly, cointensive mm-precisiation. As will be seen at a later point, a striking feature of fuzzy logic is its high power of cointensive precisiation.

The concept of cointensive mm-precisiation has an important implication for scientific theories. In large measure, scientific theories are based on bivalent logic. As a consequence, definitions of concepts are, generally, bivalent, in the sense that no degrees of truth are allowed. An example is the definition of bear market. The same applies to the definitions of recession, stability, independence, causality, stationarity, etc. The problem is that most of the concepts which are associated with bivalent definitions are in fact fuzzy, that is, are a matter of degree. For instance, the reason why the mm-precisiand of bear market is not cointensive is rooted in the fact that bear market is a fuzzy concept. What can be said in a general way is that bivalent-logic-based definitions of fuzzy concepts cannot be expected to be cointensive, just as linear models of nonlinear systems cannot be expected to be good models. In summary, an important conclusion relating to the concept of cointension may be stated as the Cointension Principle: To Achieve high cointesion it is necessary, in general, to associate a fuzzy precisiend with a fuzzy precisiand.

Note. In the sequel, unless stated to the contrary, precisiation should be understood as cointensive mm-precisiation.

In the foregoing discussion, we talked about mm-precisiation but have not addressed a basic question: How can a proposition, p, be mm-precisiated? In fuzzy logic, a concept which plays a pivotal role in mm-precisiation is that of a generalized constraint. A brief discussion of this concept is presented in the following section.

The Concept of a Generalized Constraint

The concept of a constraint is high on the list of basic concepts in science. There is an extensive literature in the realms of mathematical programming and optimal control. Particularly worthy of note is the rapid growth of interest in constraint programming within computer science and related fields [53].

A basic assumption which is commonly made in the literature is that constraints are hard (inelastic) and are precisely defined. This assumption is not a good fit to reality. In most realistic settings, constraints have some elasticity and are not precisely defined. Familiar examples are:

- Check-out time is l pm. A constraint on check-out time.
- Speed limit is l00 km/h. A constraint on speed.
- Vera has a son in mid-twenties and a daughter in mid-thirties. A constraint on Vera's age.

A first step toward precisiation of such constraints was taken in [4]. A further step was taken in [76]. The concept of a generalized constraint was introduced in [89]. A more detailed description was given in [93] and in [96,98,101].

r: $\varepsilon\,|=|\le|\ge|\subset|\dots|$ blank $|\,p\,|\,v\,|\,u\,|$ rs $|$ fg $|$ g$|\dots$

bivalent

primary
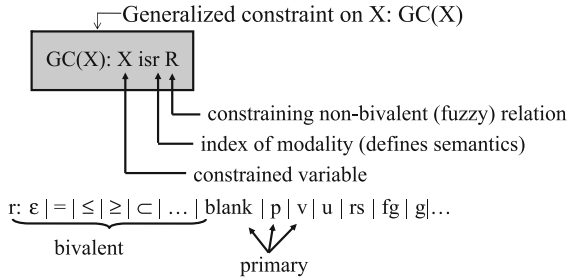
- open GC(X): X is free (GC(X) is a predicate)
- closed GC(X): X is instantiated (GC(X) is a proposition)

Fuzzy Logic, Figure 18
Principal features of a generalized constraint

The principal features of the concept of generalized constraint are summarized in Fig. 18 and et seq.

Constrained Variable

The constraint variable, X, can assume a variety of forms. Among the principal forms are the following.

X is an n-ary variable, $X = (X_1, \dots, X_n)$

X is a proposition, e. g., Leslie is tall

X is a relation

X is a function of another variable: $X = f(Y)$

X is conditioned on another variable, X/Y

X is conditioned on another generalized constraint, X isr if Y iss S

X has a structure, e. g., X = Location (Residence(Carol))

X is a generalized constraint, $X : Y$ isr R

X is a group variable. In this case, there is a group, $G : (Name_1, \dots, Name_n)$, with each member of the group, $Name_i, i = 1, \dots, n$, associated with an attribute-value, h_i, of attribute H. h_i may be vector-valued. Symbolically
$G = (Name_1, \dots, Name_n)$,
$G[H] = (Name_1/h_1, \dots, Name_n/h_n)$,
$G[H$ is $A] = (\mu_A(h_i)/Name_1, \dots, \mu_A(h_n)/Name_n)$.
Basically, $G[H]$ is a relation and $G[H$ is $A]$ is a fuzzy restriction of $G[H]$ [73,75,76]
The concept of a group variable is closely related to the concept of a fuzzy relation.

Modalities of Generalized Constraints

The indexical variable, r, defines the modality of a generalized constraint, that is, its semantics. The principal modalities are listed below.

$r :=$ equality constraint: $X = R$ is an abbreviation of X is $= R$

$r :\leqslant$ inequality constraint: $X \leqslant R$

$r :\subset$ subsethood constraint: $X \subset R$

$r :$ blank   possibilistic constraint; X is R; R is the possibility distribution of X

$r : p$ probabilistic constraint; X isp R; R is the probability distribution of X

$r : v$ veristic constraint; X isv R; R is the verity (truth) distribution of X

Standard constraints   bivalent possibilistic, bivalent veristic and probabilistic

$r : rs$ random set constraint; X isrs R; R is the set-valued probability distribution of X

$r : fg$ fuzzy graph constraint; X isfg R; X is a function and R is its fuzzy graph

$r : u$ usuality constraint; X isu R means usually (X is R)

$r : g$ group constraint; X isg R means that R constrains the attribute-values of the group

$r : i$ iterated constraint; X isi R means that X iss S and S ist T.

To define the semantics of various modalities it is convenient to assume that the constrained variable, X, takes values in a finite set $U = (u_1, \dots, u_n)$. With this assumption, the semantics of various constraints may be defined as follows.

Possibilistic Constraint

Consider the possibilistic constraint

X is A ,

where A is a fuzzy set in U, defined as [73,75].

$$A = \mu_1/u_1 + \cdots + \mu_n/u_n ,$$

with the understanding that $+$ is a separator and μ_i is the grade of membership of u_i in A, $i = 1, \dots, n$. The meaning of the possibilistic constraint, X is A, is defined as

$$X \text{ is } A \xrightarrow{\text{definition}} \text{Poss}(X = u_i) = \mu_i , \quad i = 1, \dots, n .$$

Probabilistic Constraint

Let P be a probability distribution defined on U. P may be expressed as [96,101]

$$P = p_1 \backslash u_1 + \cdots + p_n \backslash u_n .$$

In this case, X isp $P \xrightarrow{\text{definition}} \text{Prob}(X = u_i) = p_i$, $i = 1, \dots, n$.

It should be noted that p_i and u_i are allowed to take granular values, P_i and U_i, respectively, meaning that

$$p_i \text{ is } P_i \text{ and } u_i \text{ is } U_i , \quad i = 1, \dots, n .$$

In this case, the probability distribution

$$P = P_1 \backslash U_1 + \cdots + P_n \backslash U_n$$

is a granule-valued probability distribution in the sense defined in Sect. "The Concept of Granulation". Alternatively, a granule-valued distribution may be viewed as the result of granulation of the expression

$$P = p_1 \backslash u_1 + \cdots + p_n \backslash u_n \,.$$

A granular probability distribution may be defined as an iterated generalized constraint. More specifically,

X isp P

$P :$ Prob$(X$ is $A_i)$ is $P_i \,, \quad i = 1, \ldots, n \,,$

where the A_i are granules of X and the P_i are granular probabilities (Fig. 5). In this case, as in the case of granule-valued distributions, P is expressed as

$$P = P_1 \backslash U_1 + \cdots + P_n \backslash U_n \,.$$

Note. Two examples will help to clarify the distinction between granular probability distributions and granule-valued probability distributions.

Example (a): Granular probability distribution. X is a real-valued random variable with probability distribution P. What is known about P is: Prob$(X$ is small) is low; Prob$(X$ is medium) is high; Prob$(X$ is large) is low. Example (b): Granule-valued probability distribution. X is a random variable taking the values small, medium and large with respective granular probabilities low, high and low. Question: What are the expected values of these probability distributions?

Note. The concepts of granular and granule-valued probability distributions are closely related to the concepts of granular and granule-valued possibility distributions.
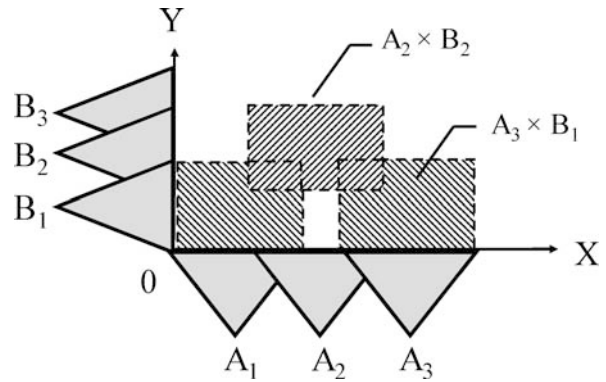
Veristic Constraint [93,98]

In this case, the semantics of X isv A is defined by

$$X \text{ isv } A \xrightarrow{\text{definition}} \text{Ver}(X = u_i) = \mu_i \,, \quad i = 1, \ldots, n \,.$$

where Ver$(X = u_i)$ is the verity (truth) of the proposition $X = u_i$.

Fuzzy Graph Constraint

In this case, X is a function, f, from U to V. Assume that U and V are granulated, with the granules of U and V being $A_1, \ldots, A_m$ and $B_1, \ldots, B_n$, respectively. The fuzzy graph,

Fuzzy Logic, Figure 19
Fuzzy graph

R, of f is defined as the disjunction of Cartesian products of the A_i and the $B_{j(i)}$ [73,74,78,90] (Fig. 19).

$$R = A_1 \times B_{j(1)} + \cdots + A_m \times B_{j(m)} \,.$$

The fuzzy graph constraint is defined as the possibilistic constraint

$$f \text{ isfg } R \xrightarrow{\text{definition}} f \text{ is } (A_1 \times B_{j(1)} + \cdots + A_m \times B_{j(m)}) \,.$$

Usuality Constraint [93,96]

The constraint X isu A is defined by

$$X \text{ isu } A \xrightarrow{\text{definition}} \text{Prob}(X \text{ is } A) \text{ is usually} \,,$$

where usually is a fuzzy set in $[0, 1]$ which represents a fuzzy probability.

Primary Constraints

Every conceivable constraint can be viewed as an instance of a generalized constraint. In practice, such generality is rarely needed. What is sufficient for most practical purposes is a subset of generalized constraints which can be generated from so-called primary constraints through combination, projection, qualification, propagation and counterpropagation. The primary constraints are: (a) possibilistic; (b) probabilistic; and (c) veristic. The primary constraints are somewhat analogous to the primary colors: red, green and blue.

Standard Constraints

In large measure, scientific theories are based on what may be called standard constraints – constraints which are

F

a subset of primary constraints. The standard constraints are: (a) bivalent possibilistic; (b) probabilistic; and (c) bivalent veristic.

What is important to note is that generality of generalized constraints goes far beyond the generality of standard constraints. What this points to is that the concept of a generalized constraint opens the door to a wide-ranging generalization of scientific theories.

Generalized Constraint Language (GCL) [93,98]

The concept of a generalized constraint serves as a basis for construction of what is referred to as the Generalized Constraint Language (GCL). More specifically, GCL is generated by combination, projection, qualification, propagation and counterpropagation of generalized constraints. For example, combination of the probabilistic constraint

$$X \text{ isp } R \,,$$

where X is a variable which takes values in a finite set, and the possibilistic constraint

$$(X, Y) \text{ is } S \,,$$

generates the fuzzy random set constraint [96,98,101]

$$Y \text{ isfrs } T \,.$$

Fuzzy random sets are closely related to fuzzy-set-valued random variables. There is an extensive literature on fuzzy-set-valued random variables [9,24,43,46,47,50,60]. Random sets and set-valued random variables are closely related to the Dempster–Shafer theory of evidence [11,56,57]. An extension of the Dempster–Shafer theory to fuzzy sets and fuzzy probabilities is sketched in [81,83].

GCL is an open language in the sense that generalized constraints may be added to GCL at will. Simple examples of generalized constraints in GCL are:

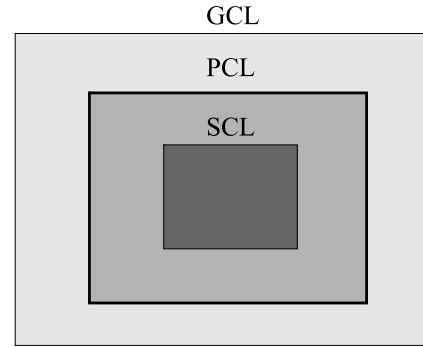$$(X \text{ is } R) \text{ and } ((X, Y) \text{ isp } S)$$
$$(X \text{ isu } R) \text{ and } (Y \text{ isu } S)$$
$$(X \text{ is } R) \text{ isp } S$$
$$(Y \text{ isu } B) \text{ if } (X \text{ is } A) \,.$$

In relation to GCL, PCL (Primary Constraint Language) and SCL (Standard Constraint Language) are subsets of GCL which are generated, respectively, by primary and standard constraints (Fig. 20).

Note. GCL is more than a language – it is a language system. A language has descriptive capability. A language

Fuzzy Logic, Figure 20
GCL, PCL and SCL

system has deductive capability in addition to descriptive capability. GCL has both capabilities.

The concept of a generalized constraint plays a pivotal role in fuzzy logic. In particular, it serves to precisiate the concepts of information and meaning. More specifically, the fundamental thesis of fuzzy logic is that information may be represented as a generalized constraint.

$$\text{information} = \text{generalized constraint} \,,$$

with the understanding that information relates to the value of a variable, X, to which the generalized constraint applies. For example, if the information about X is that X is small, then this information may be represented as a possibilistic constraint

$$X \text{ is small} \,,$$

with small being a granular value of X. It should be noted that the traditional view that information is statistical in nature is a special, albeit important case, of viewing information as a generalized constraint. Another point which should be noted is that in information theory the primary concern is not with the substance of information but with its measure. The fundamental thesis relates to substance rather than measure [16,33].

An important corollary of the fundamental thesis is the meaning postulate of fuzzy logic. More specifically, let p be a proposition. A proposition is a carrier of information. Consequently,

$$\text{meaning of } p = \text{generalized constraint} \,.$$

More concretely,

$$\text{meaning of proposition} = \text{closed generalized constraint}$$
$$\text{meaning of predicate} = \text{open generalized constraint} \,.$$

The meaning postulate leads to an important connection between the concept of a generalized constraint and the concept of mm-precisiation. More specifically, what can be concluded is the equality

mm-precisiand = generalized constraint .

Equivalently,

mm-precisiation = translation into GCL .

This equality may be viewed as a more concrete statement of the meaning postulate. Transparency of translation may be enhanced through annotation. Details may be found in [101].

The meaning postulate points to an important aspect of translation into GCL.

Let SCL denote the subset of GCL which is associated with standard constraints, that is, bivalent possibilistic, probabilistic and bivalent veristic constraints. Let p be a proposition. An mm-precisiand of p, p^*, is an element of GCL.

Let $C(p^*, p)$ be the cointension of p^* in relation to p, and let $\sup_{GCL} C(p^*, p)$ and $\sup_{SCL} C(p^*, p)$ be the suprema of $C(p^*, p)$ over GCL and SCL, respectively. Since SCL is a subset of GCL, we have

$$\sup_{GCL} C(p^*, p) \geqslant \sup_{SCL} C(p^*, p) .$$

This obvious inequality has an important implication. Specifically, as a meaning representation language, fuzzy logic dominates bivalent logic. As a very simple example consider the proposition p: Speed limit is 65 mph. Realistically, what is the meaning of p? The inequality implies that employment of fuzzy logic for precisiation of p would lead to a precisiand whose cointension is at least as high – and generally significantly higher – than the cointension which is achievable through the use of bivalent logic.

More concretely, assume that A tells B that the speed limit is 65 mph, with the understanding that 65 mph should be interpreted as "approximately 65 mph". B asks A to precisiate what is meant by "approximately 65 mph", and stipulates that no imprecise numbers and no probabilities should be used in precisiation. With this restriction, A is not capable of formulating a realistic meaning of "approximately 65 mph". Next, B allows A to use imprecise numbers but no probabilities. B is still unable to formulate a realistic definition. Next, B allows A to employ imprecise numbers but no imprecise probabilities. Still, A is unable to formulate a realistic definition. Finally, B allows A to use imprecise numbers and imprecise probabilities. This allows A to formulate a realistic definition of "approximately

65 mph". This simple example is intended to demonstrate the need for the vocabulary of fuzzy logic to precisiate the meaning of terms and concepts which involve imprecise probabilities.

In addition to serving as a basis for precisiation of meaning, GCL serves another important function – the function of a deductive question-answering system [100]. In this role, what matters are the rules of deduction. In GCL, the rules of deduction coincide with the rules which govern constraint propagation and counterpropagation. Basically, these are the rules which govern generation of a generalized constraint from other generalized constraints [100,101].

The principal rule of deduction in fuzzy logic is the so-called extension principle [70,75]. The extension principle can assume a variety of forms, depending on the generalized constraints to which it applies. A basic form which involves possibilistic constraints is the following. An analogous principle applies to probabilistic constraints.

Let X be a variable which takes values in U, and let f be a function from U to V. The point of departure is a possibilistic constraint on $f(X)$ expressed as

$$f(X) \text{ is } A ,$$

where A is a fuzzy relation in V which is defined by its membership function $\mu_A(v)$, $v \in V$.

Let g be a function from U to W. The possibilistic constraint on $f(X)$ induces a possibilistic constraint on $g(X)$ which may be expressed as

$$g(X) \text{ is } B ,$$

where B is a fuzzy relation. The question is: What is B?

The extension principle reduces the problem of computation of B to the solution of a variational problem. Specifically,

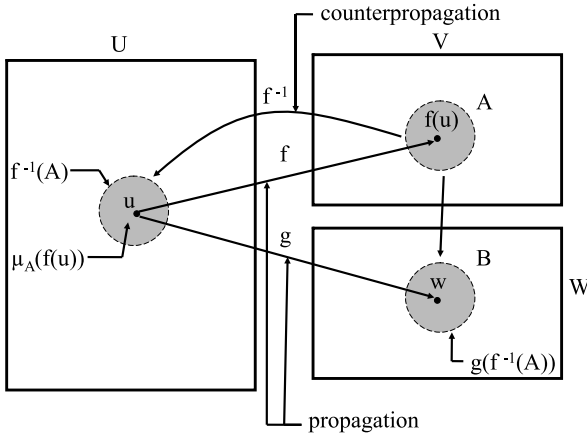$$\frac{f(X) \text{ is } A}{g(X) \text{ is } B}$$

where

$$\mu_B(w) = \sup_u \mu_A(f(u))$$

subject to

$$w = g(u) .$$

The structure of the solution is depicted in Fig. 21. Basically, the possibilistic constraint on $f(X)$ counterpropagates to a possibilistic constraint on X. Then, the possibilistic constraint on X propagates to a possibilistic constraint on $g(X)$.
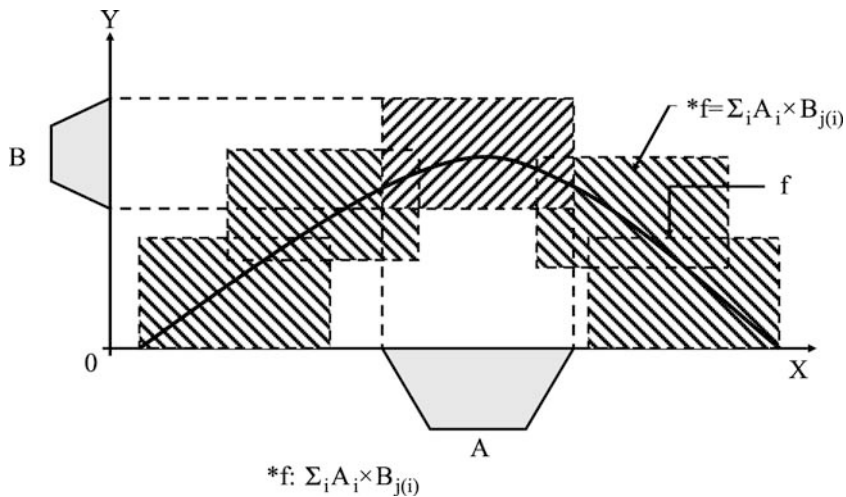
Fuzzy Logic, Figure 21
Structure of the extension principle

There is a version of the extension principle – referred to as the fuzzy-graph extension principle – which plays an important role in control and systems analysis [74,78,90]. More specifically, let f be a function from reals to reals, $Y = f(X)$. Let $*f$ and $*X$ be the granulands of f and X, respectively, with $*f$ having the form of a fuzzy graph (Sect. "The Concept of Granulation").

$$*f = A_1 \times B_{j(1)} + \cdots + A_m \times B_{j(m)} \,,$$

where the A_i, $i = 1, \ldots, m$ and the B_j, $j = 1, \ldots, n$, are granules of X and Y, respectively; $\times$ denotes Cartesian product; and $+$ denotes disjunction (Fig. 22).

In this instance, the extension principle may be expressed as follows.

$$\frac{X \text{ is } A}{Y \text{ is } (m_1 \wedge B_{j(1)} + \cdots + m_m \wedge B_{j(m)})} \\ f \text{ is } (A_1 \times B_{j(1)} + \cdots + A_m \times B_{j(m)})$$

where the m_i are matching coefficients, defined as [78]

$$m_i = \sup(A \cap A_i) \,, \quad i = 1, \ldots, m$$

and $\wedge$ denotes conjunction (min). In the special case where X is a number, a, the possibilistic constraint on Y may be expressed as

$$Y \text{ is } (\mu_{A_1}(a) \wedge B_{j(1)} + \cdots + \mu_{A_m}(a) \wedge B_{j(m)}) \,.$$

In this form, the extension principle plays a key role in the Mamdani–Assilian fuzzy logic controller [39].

Deduction

Assume that we are given an information set, I, which consists of a system of propositions $(p_1, \ldots, p_n)$. I will be referred to as the initial information set. The canonical problem of deductive question-answering is that of computing an answer to q, ans$(q|I)$, given I [100,101].

The first step is to ask the question: What information is needed to answer q? Suppose that the needed information consists of the values of the variables $X_1, \ldots, X_n$. Thus,

$$\text{ans}(q|I) = g(X_1, \ldots, X_n) \,,$$

Fuzzy Logic, Figure 22
Fuzzy-graph extension principle. $B = {}^*f(A)$

where g is a known function.

Using GCL as a meaning precisiation language, one can express the initial information set as a generalized constraint on $X_1, \ldots, X_n$. In the special case of possibilistic constraints, the constraint on the X_i may be expressed as

$$f(X_1, \ldots, X_n) \text{ is } A \,,$$

where A is a fuzzy relation.

At this point, what we have is (a) a possibilistic constraint induced by the initial information set, and (b) an answer to q expressed as

$$\text{ans}(q|I) = g(X_1, \ldots, X_n) \,,$$

with the understanding that the possibilistic constraint on f propagates to a possibilistic constraint on g. To compute the induced constraint on g what is needed is the extension principle of fuzzy logic [70,75].

As a simple illustration of deduction, it is convenient to use an example which was considered earlier.

Initial information set,	p : most Swedes are tall
Question,	q : what is the average height of Swedes?

What information is needed to compute the answer to q? Let P be a population of n Swedes, Swede$_1, \ldots,$ Swede$_n$. Let h_i be the height of Swede$_i$, $i = 1, \ldots, n$. Knowing the h_i, one can express the answer to q as

$$\text{av}(h) : \text{ans}(q|p) = \frac{1}{n}(h_1 + \cdots + h_n) \,.$$

Turning to the constraint induced by p, we note that the mm-precisiand of p may be expressed as the possibilistic constraint

$$p \xrightarrow{\text{mm-precisiation}} \frac{1}{n} \sum \text{Count(tall.Swedes) is most} \,,$$

where $\Sigma \text{Count(tall.Swedes)}$ is the number of tall.Swedes in P, with the understanding that tall.Swedes is a fuzzy subset of P. Using this definition of ΣCount [86], one can write the expression for the constraint on the h_i as

$$\frac{1}{n}(\mu_{\text{tall}}(h_1) + \cdots + \mu_{\text{tall}}(h_n)) \text{ is most} \,.$$

At this point, application of the extension principle leads to a solution which may be expressed as

$$\mu_{\text{av}(h)}(v) = \sup_h \left(\frac{1}{n} \mu_{\text{most}}(\mu_{\text{tall}}(h_1) + \cdots + \mu_{\text{tall}}(h_n)) \right) \,,$$
$$h = (h_1, \ldots, h_n)$$

subject to

$$v = \frac{1}{n}(h_1 + \cdots + h_n) \,.$$

In summary, the Generalized Constraint Language is, by construction, maximally expressive. Importantly, what this implies is that, in realistic settings, fuzzy logic, viewed as a modeling language, has a significantly higher level of power and generality than modeling languages based on standard constraints or, equivalently, on bivalent logic and bivalent-logic-based probability theory.

Principal Contributions of Fuzzy Logic

As was stated earlier, fuzzy logic is much more than an addition to existing methods for dealing with imprecision, uncertainty and complexity. In effect, fuzzy logic represents a paradigm shift. The structure of the shift is shown in Fig. 23.

Contributions of fuzzy logic range from contributions to basic sciences to applications involving various types of systems and products. The principal contributions are summarized in the following.

Fuzzy Logic as the Basis for Generalization of Scientific Theories

One of the principal contributions of fuzzy logic to basic sciences relates to what is referred to as FL-generalization.

By construction, fuzzy logic has a much more general conceptual structure than bivalent logic. A key element in the transition from bivalent logic to fuzzy logic is the generalization of the concept of a set to a fuzzy set. This generalization is the point of departure for FL-generalization.

More specifically, FL-generalization of any theory, T, involves an addition to T of concepts drawn from fuzzy
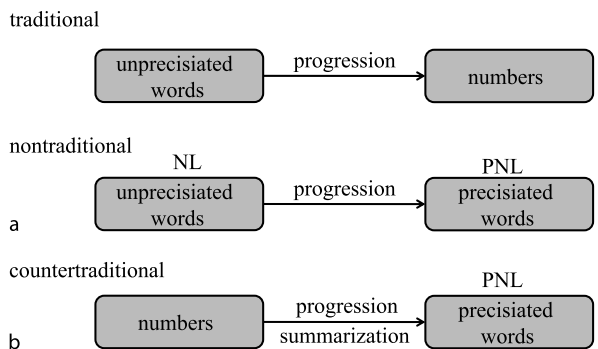
Fuzzy Logic, Figure 23
Fuzzy logic as a paradigm shift

logic. In the limit, as more and more concepts which are drawn from fuzzy logic are added to T, the foundation of T is shifted from bivalent logic to fuzzy logic. By construction, FL-generalization results in an upgraded theory, T^+, which is at least as rich and, in general, significantly richer than T.

As an illustration, consider probability theory, PT – a theory which is bivalent-logic-based. Among the basic concepts drawn from fuzzy logic which may be added to PT are the following [96].

> set + fuzzy set
> event + fuzzy event
> relation + fuzzy relation
> probability + fuzzy probability
> random set + fuzzy random set
> independence + fuzzy independence
> stationarity + fuzzy stationarity
> random variable + fuzzy random variable
> etc.

As a theory, PT^+ is much richer than PT. In particular, it provides a basis for construction of models which are much closer to reality than those that can be constructed through the use of PT. This applies, in particular, to computation with imprecise probabilities.

A number of scientific theories have already been FL-generalized to some degree, and many more are likely to be FL-generalized in coming years. Particularly worthy of note are the following FL-generalizations.

> control → fuzzy control
>   [12,18,20,69,72]
> linear programming → fuzzy linear programming
>   [19,103]
> probability theory → fuzzy probability theory
>   [96,98,101]
> measure theory → fuzzy measure theory
>   [14,62]
> topology → fuzzy topology
>   [38,68]
> graph theory → fuzzy graph theory
>   [34,41]
> cluster analysis → fuzzy cluster analysis
>   [7,27]
> Prolog → fuzzy Prolog
>   [21,42]
> etc.

FL-generalization is a basis for an important rationale for the use of fuzzy logic. It is conceivable that eventually the foundations of many scientific theories may be shifted from bivalent logic to fuzzy logic.

Linguistic Variables and Fuzzy If-Then Rules   The most visible, the best understood and the most widely used contribution of fuzzy logic is the concept of a linguistic variable and the associated machinery of fuzzy if-then rules [90].

The machinery of linguistic variables and fuzzy if-then rules is unique to fuzzy logic. This machinery has played and is continuing to play a pivotal role in the conception and design of control systems and consumer products. However, its applicability is much broader. A key idea which underlies the machinery of linguistic variables and fuzzy if-then rules is centered on the use of information compression. In fuzzy logic, information compression is achieved through the use of graduated (fuzzy) granulation.

The Concepts of Precisiation and Cointension   The concepts of precisiation and cointension play pivotal roles in fuzzy logic [101]. In fuzzy logic, differentiation is made between two concepts of precision: precision of value, v-precision; and precision of meaning, m-precision. Furthermore, differentiation is made between precisiation of meaning which is (a) human-oriented, or mh-precisiation for short; and (b) machine-oriented, or mm-precisiation for short. It is understood that mm-precisiation is mathematically well defined. The object of precisiation, p, and the result of precisiation, p^*, are referred to as precisiend and precisiand, respectively. Informally, cointension is defined as a measure of closeness of the meanings of p and p^*. Precisiation is cointensive if the meaning of p^* is close to the meaning of p. One of the important features of fuzzy logic is its high power of cointensive precisiation. What this implies is that better models of reality can be achieved through the use of fuzzy logic.

Cointensive precisiation has an important implication for science. In large measure, science is bivalent-logic-based. In consequence, in science it is traditional to define concepts in a bivalent framework, with no degrees of truth allowed. The problem is that, in reality, many concepts in science are fuzzy, that is, are a matter of degree. For this reason, bivalent-logic-based definitions of scientific concepts are, in many cases, not cointensive. To formulate cointensive definitions of fuzzy concepts it is necessary to employ fuzzy logic.

As was noted earlier, one of the principal contributions of fuzzy logic is its high power of cointensive precisiation. The significance of this capability of fuzzy logic is underscored by the fact that it has always been, and continues

to be, a basic objective of science to precisiate and clarify what is imprecise and unclear.

Computing with Words (CW), NL-Computation and Precisiated Natural Language (PNL)   Much of human knowledge is expressed in natural language. Traditional theories of natural language are based on bivalent logic. The problem is that natural languages are intrinsically imprecise. Imprecision of natural languages is rooted in imprecision of perceptions. A natural language is basically a system for describing perceptions. Perceptions are intrinsically imprecise, reflecting the bounded ability of human sensory organs, and ultimately the brain, to resolve detail and store information. Imprecision of perceptions is passed on to natural languages.

Bivalent logic is intolerant of imprecision, partiality of truth and partiality of possibility. For this reason, bivalent logic is intrinsically unsuited to serve as a foundation for theories of natural language. As the logic of imprecision and approximate reasoning, fuzzy logic is a much better choice [71,80,84,85,86,88].

Computing with words (CW), NL-computation and precisiated natural language (PNL) are closely related formalisms [93,94,95,97]. In conventional modes of computation, the objects of computation are mathematical constructs. By contrast, in computing with words the objects of computation are propositions and predicates drawn from a natural language. A key idea which underlies computing with words involves representing the meaning of propositions and predicates as generalized constraints. Computing with words opens the door to a wide-ranging enlargement of the role of natural languages in scientific theories [30,31,36,58,59,61].

Computational Theory of Perceptions   Humans have a remarkable capability to perform a wide variety of physical and mental tasks without any measurements and any computations. In performing such tasks humans employ perceptions. To endow machines with this capability what is needed is a formalism in which perceptions can play the role of objects of computation. The fuzzy-logic-based computational theory of perceptions (CTP) serves this purpose [94,95]. A key idea in this theory is that of computing not with perceptions per se, but with their descriptions in a natural language. Representing perceptions as propositions drawn from a natural language opens the door to application of computing with words to computation with perceptions. Computational theory of perceptions is of direct relevance to achievement of human level machine intelligence.

Possibility Theory   Possibility theory is a branch of fuzzy logic [15,79]. Possibility theory and probability theory are distinct theories. Possibility theory may be viewed as a formalization of perception of possibility, whereas probability theory is rooted in perception of likelihood. In large measure, possibility theory and probability theory are complementary rather than competitive. Possibility theory is of direct relevance to, knowledge representation, semantics of natural languages, decision analysis and computation with imprecise probabilities.

Computation with Imprecise Probabilities   Most real-world probabilities are perceptions of likelihood. As such, real-world probabilities are intrinsically imprecise [75]. Until recently, the issue of imprecise probabilities has been accorded little attention in the literature of probability theory. More recently, the problem of computation with imprecise probabilities has become an object of rapidly growing interest [63,99].

Typically, imprecise probabilities occur in an environment of imprecisely defined variables, functions, relations, events, etc. Existing approaches to computation with imprecise probabilities do not address this reality. To address this reality what is needed is fuzzy logic and, more particularly, computing with words and the computational theory of perceptions. A step in this direction was taken in the paper "Toward a perception-based theory of probabilistic reasoning with imprecise probabilities"; [96] followed by the 2005 paper "Toward a generalized theory of uncertainty (GTU) – an outline", [98] and the 2006 paper "Generalized theory of uncertainty (GTU) –principal concepts and ideas" [101].

Fuzzy Logic as a Modeling Language   Science deals not with reality but with models of reality. More often than not, reality is fuzzy. For this reason, construction of realistic models of reality calls for the use of fuzzy logic rather than bivalent logic.

Fuzzy logic is a logic of imprecision, uncertainty and approximate reasoning [82]. It is natural to employ fuzzy logic as a modeling language when the objects of modeling are not well defined [102]. But what is somewhat paradoxical is that in many of its practical applications fuzzy logic is used as a modeling language for systems which are precisely defined. The explanation is that, in general, precision carries a cost. In those cases in which there is a tolerance for imprecision, reduction in cost may be achieved through imprecisiation, e. g., data compression, information compression and summarization. The result of imprecisiation is an object of modeling which is not precisely defined. A fuzzy modeling language comes into play at this

point. This is the key idea which underlies the fuzzy logic gambit. The fuzzy logic gambit is widely used in the design of consumer products – a realm in which cost is an important consideration.

A Glimpse of What Lies Beyond Fuzzy Logic

Fuzzy logic has a much higher level of generality than traditional logical systems – a generality which has the effect of greatly enhancing the problem-solving capability of fuzzy logic compared with that of bivalent logic.

What lies beyond fuzzy logic? What is of relevance to this question is the so-called incompatibility principle [73]. Informally, this principle asserts that as the complexity of a system increases a point is reached beyond which high precision and high relevance become incompatible. The concepts of mm-precisiation and cointension suggest an improved version of the principle: As the complexity of a system increases a point is reached beyond which high cointension and mm-precision become incompatible. What it means in plain words is that in the realm of complex systems – such as economic systems – it may be impossible to construct models which are both realistic and precise.

As an illustration consider the following problem. Assume that *A* asks a cab driver to take him to address *B*. There are two versions of this problem. (a) *A* asks the driver to take him to *B* the shortest way; and (b) *A* asks the driver to take him to *B* the fastest way. Based on his experience, the driver decides on the route to take to *B*. In the case of (a), a GPS system can suggest a route that is provably the shortest way, that is, it can come up with a provably valid (p-valid) solution. In the case of (b) the uncertainties involved preclude the possibility of constructing a model of the system which is cointensive and mm-precise, implying that a p-valid solution does not exist. The driver's solution, based on his experience, has what may be called fuzzy validity (f-validity). Thus, in the case of (b) no p-valid solution exists. What exists is an f-valid solution.

In fuzzy logic, mm-precisiation is a prerequisite to computation. A question which arises is: What can be done when cointensive mm-precisiation is infeasible? To deal with such problems what is needed is referred to as extended fuzzy logic (FL+). In this logic, mm-precisiation is optional rather than mandatory.

Very briefly, what is admissible in FL+ is f-validity. Admissibility of f-validity opens the door to construction of concepts prefixed with f, e. g. f-theorem, f-proof, f-principle, f-triangle, f-continuity, f-stability, etc. An example is f-geometry. In f-geometry, figures are drawn by hand with a spray can. An example of f-theorem in f-geometry is

the f-version of the theorem: The medians of a triangle are concurrent. The f-version of this theorem reads: The f-medians of an f-triangle are f-concurrent. An f-theorem can be proved in two ways. (a) empirically, that is, by drawing triangles with a spray can and verifying that the medians intersect at an f-point. Alternatively, the theorem may be f-proved by constructing an f-analogue of its traditional proof.

At this stage, the extended fuzzy logic is merely an idea, but it is an idea which has the potential for being a point of departure for construction of theories with important applications to the solution of real-world problems.

Bibliography

Primary Literature

1. Aliev RA, Fazlollahi B, Aliev RR, Guirimov BG (2006) Fuzzy time series prediction method based on fuzzy recurrent neural network. In: Neuronal Informatinformation Processing Book. Lecture notes in computer science, vol 4233. Springer, Berlin, pp 860–869
2. Bargiela A, Pedrycz W (2002) Granular computing: An Introduction. Kluwer Academic Publishers, Boston
3. Bardossy A, Duckstein L (1995) Fuzzy rule-based modelling with application to geophysical, biological and engineering systems. CRC Press, New York
4. Bellman RE, Zadeh LA (1970) Decision-making in a fuzzy environment. Manag Sci B 17:141–164
5. Belohlavek R, Vychodil V (2006) Attribute implications in a fuzzy setting. In: Ganter B, Kwuida L (eds) ICFCA (2006) Lecture notes in artificial intelligence, vol 3874. Springer, Heidelberg, pp 45–60
6. Bezdek J, Pal S (eds) (1992) Fuzzy models for pattern recognition – methods that search for structures in data. IEEE Press, New York
7. Bezdek J, Keller JM, Krishnapuram R, Pal NR (1999) Fuzzy models and algorithms for pattern recognition and image processing. In: Zimmermann H (ed) Kluwer, Dordrecht
8. Bouchon-Meunier B, Yager RR, Zadeh LA (eds) (2000) Uncertainty in intelligent and information systems. In: Advances in fuzzy systems – applications and theory, vol 20. World Scientific, Singapore
9. Colubi A, Santos Domínguez-Menchero J, López-Díaz M, Ralescu DA (2001) On the formalization of fuzzy random variables. Inf Sci 133(1–2):3–6
10. Cresswell MJ (1973) Logic and Languages. Methuen, London
11. Dempster AP (1967) Upper and lower probabilities induced by a multivalued mapping. Ann Math Stat 38:325–329
12. Driankov D, Hellendoorn H, Reinfrank M (1993) An Introduction to Fuzzy Control. Springer, Berlin
13. Dubois D, Prade H (1980) Fuzzy Sets and Systems – Theory and Applications. Academic Press, New York
14. Dubois D, Prade H (1982) A class of fuzzy measures based on triangular norms. Int J General Syst 8:43–61
15. Dubois D, Prade H (1988) Possibility Theory. Plenum Press, New York

16. Dubois D, Prade H (1994) Non-standard theories of uncertainty in knowledge representation and reasoning. Knowl Engineer Rev Camb J Online 9(4):pp 399–416

17. Esteva F, Godo L (2007) Towards the generalization of Mundici's gamma functor to IMTL algebras: the linearly ordered case, Algebraic and proof-theoretic aspects of non-classical logics, pp 127–137

18. Filev D, Yager RR (1994) Essentials of Fuzzy Modeling and Control. Wiley-Interscience, New York

19. Gasimov RN, Yenilmez K (2002) Solving fuzzy linear programming problems with linear membership functions. Turk J Math 26:375–396

20. Gerla G (2001) Fuzzy control as a fuzzy deduction system. Fuzzy Sets Syst 121(3):409–425

21. Gerla G (2005) Fuzzy logic programming and fuzzy control. Studia Logica 79(2):231–254

22. Godo LL, Esteva F, García P, Agustí J (1991) A formal semantical approach to fuzzy logic. In: International Symposium on Multiple Valued Logic, ISMVL'91, pp 72–79

23. Goguen JA (1967) L-fuzzy sets. J Math Anal Appl 18:145–157

24. Goodman IR, Nguyen HT (1985) Uncertainty models for knowledge-based systems. North Holland, Amsterdam

25. Hajek P (1998) Metamathematics of fuzzy logic. Kluwer, Dordrecht

26. Hirota K, Sugeno M (eds) (1995) Industrial applications of fuzzy technology in the world. In: Advances in fuzzy systems – applications and theory, vol 2. World Scientific, Singapore

27. Höppner F, Klawonn F, Kruse R, Runkler T (1999) Fuzzy cluster analysis. Wiley, Chichester

28. Jamshidi M, Titli A, Zadeh LA, Boverie S (eds) (1997) Applications of fuzzy logic – towards high machine intelligence quotient systems. In: Environmental and intelligent manufacturing systems series, vol 9. Prentice Hall, Upper Saddle River

29. Jankowski A, Skowron A (2007) Toward rough-granular computing. In: Proceedings of the 11th International Conference on Rough Sets, Fuzzy Sets, Data Mining, and Granular Computing, (RSFDGrC'07), Toronto, Canada, pp 1–12

30. Kacprzyk J, Zadeh LA (eds) (1999) Computing with words in information/intelligent systems part 1. Foundations. Physica, Heidelberg, New York

31. Kacprzyk J, Zadeh LA (eds) (1999) Computing with words in information/intelligent systems part 2. Applications. Physica, Heidelberg, New York

32. Kandel A, Langholz G (eds) (1994) Fuzzy control systems. CRC Press, Boca Raton

33. Klir GJ (2006) Uncertainty and information: Foundations of generalized information theory. Wiley-Interscience, Hoboken

34. Kóczy LT (1992) Fuzzy graphs in the evaluation and optimization of networks. Fuzzy Sets Syst 46(3):307–319

35. Lambert K, Van Fraassen BC (1970) Meaning relations, possible objects and possible worlds. Philosophical problems in logic, pp 1–19

36. Lawry J, Shanahan JG, Ralescu AL (eds) (2003) Modelling with words – learning, fusion, and reasoning within a formal linguistic representation framework. Springer, Heidelberg

37. Lin TY (1997) Granular computing: From rough sets and neighborhood systems to information granulation and computing in words. In: European Congress on Intelligent Techniques and Soft Computing, September 8–12, pp 1602–1606

38. Liu Y, Luo M (1997) Fuzzy topology. In: Advances in fuzzy systems – applications and theory, vol 9. World Scientific, Singapore

39. Mamdani EH, Assilian S (1975) An experiment in linguistic synthesis with a fuzzy logic controller. Int J Man-Machine Stud 7:1–13

40. Mendel J (2001) Uncertain rule-based fuzzy logic systems – Introduction and new directions. Prentice Hall, Upper Saddle River

41. Mordeson JN, Nair PS (2000) Fuzzy graphs and fuzzy hypergraphs. In: Studies in Fuzziness and Soft Computing. Springer, Heidelberg

42. Mukaidono M, Shen Z, Ding L (1989) Fundamentals of fuzzy prolog. Int J Approx Reas 3(2):179–193

43. Nguyen HT (1993) On modeling of linguistic information using random sets. In: Fuzzy sets for intelligent systems. Morgan Kaufmann Publishers, San Mateo, pp 242–246

45. Novak V (2006) Which logic is the real fuzzy logic? Fuzzy Sets Syst 157:635–641

44. Novak V, Perfilieva I, Mockor J (1999) Mathematical principles of fuzzy logic. Kluwer, Boston/Dordrecht

46. Ogura Y, Li S, Kreinovich V (2002) Limit theorems and applications of set-valued and fuzzy set-valued random variables. Springer, Dordrecht

47. Orlov AI (1980) Problems of optimization and fuzzy variables. Znaniye, Moscow

48. Pedrycz W, Gomide F (2007) Fuzzy systems engineering: Toward human-centric computing. Wiley, Hoboken

49. Perfilieva I (2007) Fuzzy transforms: a challenge to conventional transforms. In: Hawkes PW (ed) Advances in images and electron physics, 147. Elsevier Academic Press, San Diego, pp 137–196

50. Puri ML, Ralescu DA (1993) Fuzzy random variables. In: Fuzzy sets for intelligent systems. Morgan Kaufmann Publishers, San Mateo, pp 265–271

51. Ralescu DA (1995) Cardinality, quantifiers and the aggregation of fuzzy criteria. Fuzzy Sets Syst 69:355–365

52. Ross TJ (2004) Fuzzy logic with engineering applications, 2nd edn. Wiley, Chichester

53. Rossi F, Codognet P (2003) Special Issue on Soft Constraints: Constraints 8(1)

54. Rutkowska D (2002) Neuro-fuzzy architectures and hybrid learning. In: Studies in fuzziness and soft computing. Springer

55. Rutkowski L (2008) Computational intelligence. Springer, Polish Scientific Publishers PWN, Warzaw

56. Schum D (1994) Evidential foundations of probabilistic reasoning. Wiley, New York

57. Shafer G (1976) A mathematical theory of evidence. Princeton University Press, Princeton

58. Trillas E (2006) On the use of words and fuzzy sets. Inf Sci 176(11):1463–1487

59. Türksen IB (2007) Meta-linguistic axioms as a foundation for computing with words. Inf Sci 177(2):332–359

60. Wang PZ, Sanchez E (1982) Treating a fuzzy subset as a projectable random set. In: Gupta MM, Sanchez E (eds) Fuzzy information and decision processes. North Holland, Amsterdam, pp 213–220

61. Wang P (2001) Computing with words. Albus J, Meystel A, Zadeh LA (eds) Wiley, New York

62. Wang Z, Klir GJ (1992) Fuzzy measure theory. Springer, New York

63. Walley P (1991) Statistical reasoning with imprecise probabilities. Chapman & Hall, London
64. Wygralak M (2003) Cardinalities of fuzzy sets. In: Studies in fuzziness and soft computing. Springer, Berlin
65. Yager RR, Zadeh LA (eds) (1992) An introduction to fuzzy logic applications in intelligent systems. Kluwer Academic Publishers, Norwell
66. Yen J, Langari R, Zadeh LA (ed) (1995) Industrial applications of fuzzy logic and intelligent systems. IEEE, New York
67. Yen J, Langari R (1998) Fuzzy logic: Intelligence, control and information, 1st edn. Prentice Hall, New York
68. Ying M (1991) A new approach for fuzzy topology (I). Fuzzy Sets Syst 39(3):303–321
69. Ying H (2000) Fuzzy control and modeling – analytical foundations and applications. IEEE Press, New York
70. Zadeh LA (1965) Fuzzy sets. Inf Control 8:338–353
71. Zadeh LA (1972) A fuzzy-set-theoretic interpretation of linguistic hedges. J Cybern 2:4–34
72. Zadeh LA (1972) A rationale for fuzzy control. J Dyn Syst Meas Control G 94:3–4
73. Zadeh LA (1973) Outline of a new approach to the analysis of complex systems and decision processes. IEEE Trans Syst Man Cybern SMC 3:28–44
74. Zadeh LA (1974) On the analysis of large scale systems. In: Gottinger H (ed) Systems approaches and environment problems. Vandenhoeck and Ruprecht, Göttingen, pp 23–37
75. Zadeh LA (1975) The concept of a linguistic variable and its application to approximate reasoning Part I. Inf Sci 8:199–249; Part II. Inf Sci 8:301–357; Part III. Inf Sci 9:43–80
76. Zadeh LA (1975) Calculus of fuzzy restrictions. In: Zadeh LA, Fu KS, Tanaka K, Shimura M (eds) Fuzzy sets and their applications to cognitive and decision processes. Academic Press, New York, pp 1–39
77. Zadeh LA (1975) Fuzzy logic and approximate reasoning. Synthese 30:407–428
78. Zadeh LA (1976) A fuzzy-algorithmic approach to the definition of complex or imprecise concepts. Int J Man-Machine Stud 8:249–291
79. Zadeh LA (1978) Fuzzy sets as a basis for a theory of possibility. Fuzzy Sets Syst 1:3–28
80. Zadeh LA (1978) PRUF – a meaning representation language for natural languages. Int J Man-Machine Stud 10:395–460
81. Zadeh LA (1979) Fuzzy sets and information granularity. In: Gupta M, Ragade R, Yager R (eds) Advances in fuzzy set theory and applications. North-Holland Publishing Co., Amsterdam, pp 3–18
82. Zadeh LA (1979) A theory of approximate reasoning. In: Hayes J, Michie D, Mikulich LI (eds) Machine intelligence 9. Halstead Press, New York, pp 149–194
83. Zadeh LA (1981) Possibility theory and soft data analysis. In: Cobb L, Thrall RM (eds) Mathematical frontiers of the social and policy sciences. Westview Press, Boulder, pp 69–129
84. Zadeh LA (1982) Test-score semantics for natural languages and meaning representation via PRUF. In: Rieger B (ed) Empirical semantics. Brockmeyer, Bochum, pp 281–349
85. Zadeh LA (1983) Test-score semantics as a basis for a computational approach to the representation of meaning. Proceedings of the Tenth Annual Conference of the Association for Literary and Linguistic Computing, Oxford University Press
86. Zadeh LA (1983) A computational approach to fuzzy quantifiers in natural languages. Comput Math 9:149–184
87. Zadeh LA (1984) Precisiation of meaning via translation into PRUF. In: Vaina L, Hintikka J (eds) Cognitive constraints on communication. Reidel, Dordrecht, pp 373–402
88. Zadeh LA (1986) Test-score semantics as a basis for a computational approach to the representation of meaning. Lit Linguist Comput 1:24–35
89. Zadeh LA (1986) Outline of a computational approach to meaning and knowledge representation based on the concept of a generalized assignment statement. In: Thoma M, Wyner A (eds) Proceedings of the International Seminar on Artificial Intelligence and Man-Machine Systems. Springer, Heidelberg, pp 198–211
90. Zadeh LA (1996) Fuzzy logic and the calculi of fuzzy rules and fuzzy graphs. Multiple-Valued Logic 1:1–38
91. Zadeh LA (1997) Toward a theory of fuzzy information granulation and its centrality in human reasoning and fuzzy logic. Fuzzy Sets Syst 90:111–127
92. Zadeh LA (1998) Some reflections on soft computing, granular computing and their roles in the conception, design and utilization of information/intelligent systems. Soft Comput 2:23–25
93. Zadeh LA (1999) From computing with numbers to computing with words – from manipulation of measurements to manipulation of perceptions. IEEE Trans Circuits Syst 45:105–119
94. Zadeh LA (2000) Outline of a computational theory of perceptions based on computing with words. In: Sinha NK, Gupta MM, Zadeh LA (eds) Soft Computing & Intelligent Systems: Theory and Applications. Academic Press, London, pp 3–22
95. Zadeh LA (2001) A new direction in AI – toward a computational theory of perceptions. AI Magazine 22(1):73–84
96. Zadeh LA (2002) Toward a perception-based theory of probabilistic reasoning with imprecise probabilities. J Stat Plan Inference 105:233–264
97. Zadeh LA (2004) Precisiated natural language (PNL). AI Magazine 25(3)74–91
98. Zadeh LA (2005) Toward a generalized theory of uncertainty (GTU) – an outline. Inf Sci 172:1–40
99. Zadeh LA (2005) From imprecise to granular probabilities. Fuzzy Sets Syst 154:370–374
100. Zadeh LA (2006) From search engines to question answering systems – The problems of world knowledge, relevance, deduction and precisiation. In: Sanchez E (ed) Fuzzy logic and the semantic web, Chapt 9. Elsevier, pp 163–210
101. Zadeh LA (2006) Generalized theory of uncertainty (GTU)– principal concepts and ideas. Comput Stat Data Anal 51:15–46
102. Zadeh LA (2008) Is there a need for fuzzy logic? Inf Sci 178:(13)2751–2779
103. Zimmermann HJ (1978) Fuzzy programming and linear programming with several objective functions. Fuzzy Sets Syst 1:45–55

Books and Reviews

Aliev RA, Fazlollahi B, Aliev RR (2004) Soft computing and its applications in business and economics. In: Studies in fuzziness and soft computing. Springer, Berlin
Dubois D, Prade H (eds) (1996) Fuzzy information engineering: A guided tour of applications. Wiley, New York

Gupta MM, Sanchez E (1982) Fuzzy information and decision processes. North-Holland, Amsterdam

Hanss M (2005) Applied fuzzy arithmetic: An introduction with engineering applications. Springer, Berlin

Hirota K, Czogala E (1986) Probabilistic sets: Fuzzy and stochastic approach to decision, control and recognition processes, ISR. Verlag TUV Rheinland, Köln

Jamshidi M, Titli A, Zadeh LA, Boverie S (1997) Applications of fuzzy logic: Towards high machine intelligence quotient systems. In: Environmental and intelligent manufacturing systems series. Prentice Hall, Upper Saddle River

Kacprzyk J, Fedrizzi M (1992) Fuzzy regression analysis. In: Studies in fuzziness. Physica 29

Kosko B (1997) Fuzzy engineering. Prentice Hall, Upper Saddle River

Mastorakis NE (1999) Computational intelligence and applications. World Scientific Engineering Society

Pal SK, Polkowski L, Skowron (2004) A rough-neural computing: Techniques for computing with words. Springer, Berlin

Ralescu AL (1994) Applied research in fuzzy technology, international series in intelligent technologies. Kluwer Academic Publishers, Boston

Reghis M, Roventa E (1998) Classical and fuzzy concepts in mathematical logic and applications. CRC-Press, Boca Raton

Schneider M, Kandel A, Langholz G, Chew G (1996) Fuzzy expert system tools. Wiley, New York

Türksen IB (2005) Ontological and epistemological perspective of fuzzy set theory. Elsevier Science and Technology Books

Zadeh LA, Kacprzyk J (1992) Fuzzy logic for the management of uncertainty. Wiley

Zhong N, Skowron A, Ohsuga S (1999) New directions in rough sets, data mining, and granular-soft computing. In: Lecture Notes in Artificial Intelligence. Springer, New York

Fuzzy Logic, Type-2 and Uncertainty

ROBERT I. JOHN[1], JERRY M. MENDEL[2]
[1] Centre for Computational Intelligence,
  School of Computing, De Montfort University,
  Leicester, United Kingdom
[2] Signal and Image Processing Institute,
  Ming Hsieh Department of Electrical Engineering,
  University of Southern California, Los Angeles, USA

Article Outline

Glossary

Type-1 fuzzy sets Are the underlying component in fuzzy logic where uncertainty is represented by a number between one and zero.

Type-2 fuzzy sets Are where the uncertainty is represented by a type-1 fuzzy set.

Interval type-2 fuzzy sets Are where the uncertainty is represented by a type-1 fuzzy set where the membership grades are unity.

Definition of the Subject

Type-2 fuzzy logic was first defined in 1975 by Zadeh and is an increasingly popular area for research and applications. The reason for this is because it appears to tackle the fundamental problem with type-1 fuzzy logic in that it is unable to handle the many uncertainties in real systems. Type-2 fuzzy systems are conceptually and mathematically more difficult to understand and implement but the proven applications show that the effort is worth it and type-2 fuzzy systems are at the forefront of fuzzy logic research and applications. These systems rely on the notion of a type-2 fuzzy set where the membership grades are type-1 fuzzy sets.

Introduction

Fuzzy sets [1] have, over the past forty years, laid the basis for a successful method of modeling uncertainty, vagueness and imprecision in a way that no other technique has been able. The use of fuzzy sets in real computer systems is extensive, particularly in consumer products and control applications.

Fuzzy logic (a logic based on fuzzy sets) is now a mainstream technique in everyday use across the world. The number of applications is many, and growing, in a variety of areas, for example, heat exchange, warm water pressure, aircraft flight control, robot control, car speed control, power systems, nuclear reactor control, fuzzy memory devices and the fuzzy computer, control of a cement kiln, focusing of a camcorder, climate control for buildings, shower control and mobile robots. The use of fuzzy logic is not limited to control. Successful applications, for example, have been reported in train scheduling, system modeling, computing, stock tracking on the Nikkei stock exchange and information retrieval.

Type-1 fuzzy sets represent uncertainty using a number in $[0, 1]$ whereas type-2 fuzzy sets represent uncertainty by a function. This is discussed in more detail later in the article. Essentially, the more imprecise or vague the data is, then type-2 fuzzy sets offer a significant improve-
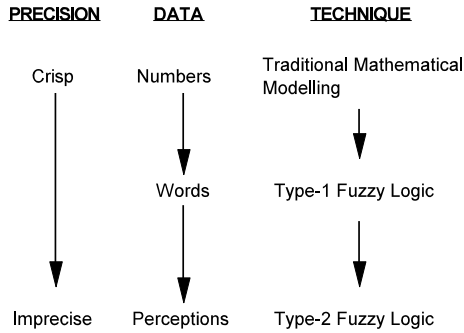
Fuzzy Logic, Type-2 and Uncertainty, Figure 1
Relationships between imprecision, data and fuzzy technique

ment on type-1 fuzzy sets. Figure 1 shows the view taken here of the relationships between levels of imprecision, data and technique.

As the level of imprecision increases then type-2 fuzzy logic provides a powerful paradigm for potentially tackling the problem. Problems that contain crisp, precise data do not, in reality, exist. However some problems can be tackled effectively using mathematical techniques where the assumption is that the data is precise. Other problems (for example, in control) use imprecise terminology that can often be effectively modeled using type-1 fuzzy sets. Perceptions, it is argued here, are at a higher level of imprecision and type-2 fuzzy sets can effectively model this imprecision.

The reason for this lies in some of the problems associated with type-1 fuzzy logic systems. Although successful in the control domain they have not delivered as well in systems that attempt to replicate human decision making. It is our view that this is because a type-1 fuzzy logic system (FLS) has some uncertainties which cannot be modeled properly by type-1 fuzzy logic. The sources of the uncertainties in type-1 FLSs are:

- The meanings of the words that are used in the antecedents and consequents of rules can be uncertain (words mean different things to different people).
- Consequents may have a histogram of values associated with them, especially when knowledge is extracted from a group of experts who do not all agree.
- Measurements that activate a type-1 FLS may be noisy and therefore uncertain.
- The data that are used to tune the parameters of a type-1 FLS may also be noisy.

The uncertainties described all essentially have to do with the uncertainty contained in a type-1 fuzzy set. A type-1 fuzzy set can be defined in the following way:

Let X be a universal set defined in a specific problem, with a generic element denoted by x. A fuzzy set A in X is a set of ordered pairs:

$$A = \{(x, \mu_A(x) \mid x \in X)\},$$

where $\mu_A : X \to [0, 1]$ is called the membership function A of and $\mu_A(x)$ represents the degree of membership of the element x in A.

The key points to draw from this definition of a fuzzy set are:

- The members of a fuzzy set are members to some degree, known as a *membership grade* or degree of membership.
- A fuzzy set is fully determined by the membership function.
- The membership grade is the degree of belonging to the fuzzy set. The larger the number (in $[0, 1]$) the more the degree of belonging.
- The translation from x to $\mu_A(x)$ is known as *fuzzification*.
- A fuzzy set is either continuous or discrete.
- Graphical representation of membership functions is very useful. For example, the fuzzy set 'Tall' might be represented as shown in Fig. 2 where someone who is of height five feet has a membership grade of zero while someone who is of height seven feet is tall to degree one, with heights in between having membership grade between one and zero. The example shown is linear but, of course, it could be any function.

Fuzzy sets offer a practical way of modeling what one might refer to as 'fuzziness'. The real world can be characterized by the fact that much of it is imprecise in one form or other. For a clear exposition (important to the notion of, and argument for, type-2 sets) two ideas of 'fuzziness'
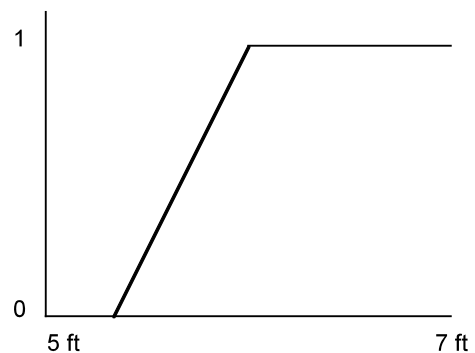
Fuzzy Logic, Type-2 and Uncertainty, Figure 2
The fuzzy set 'Tall'
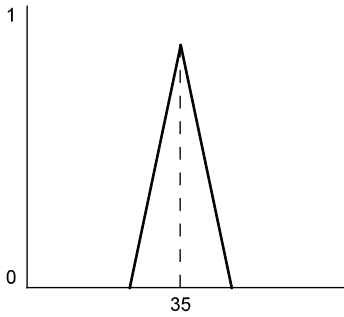
Fuzzy Logic, Type-2 and Uncertainty, Figure 3
The fuzzy number 'About 35'

can be considered important – imprecision and vagueness (linguistic uncertainty).

Imprecision

As has already been discussed, in many physical systems measurements are never precise (a physical property can always be measured more accurately). There is imprecision inherent in measurement. Fuzzy numbers are one way of capturing this imprecision by having a fuzzy set represent a real number where the numbers in an interval near to the number are in the fuzzy set to some degree. So, for example, the fuzzy number 'About 35' might look like the fuzzy set in Fig. 3 where the numbers closer to 35 have membership nearer unity than those that are further away from 35.

Vagueness or Linguistic Uncertainty

Another use of fuzzy sets is where words have been used to capture imprecise notions, loose concepts or perceptions. We use words in our everyday language that we, and the intended audience, know what we want to convey but the words cannot be precisely defined. For example, when a bank is considering a loan application somebody may be assessed as a good risk in terms of being able to repay the loan. Within the particular bank this notion of a good risk is well understood. It is not a black and white decision as to whether someone is a good risk or not – they are a good risk *to some degree*.

Type-2 Fuzzy Systems

A type-1 fuzzy system uses type-1 fuzzy sets in either the antecedent and/or the consequent of type-1 fuzzy if-then rules and a type-2 fuzzy system deploys type-2 fuzzy sets in either the antecedent and/or the consequent of type-2 fuzzy rules.

Fuzzy systems usually have the following features:

- The *fuzzy sets* as defined by their membership functions. These fuzzy sets are the basis of a fuzzy system. They capture the underlying properties or knowledge in the system.
- The *if-then rules* that combine the fuzzy sets – in a rule set or knowledge base.
- The fuzzy *composition* of the rules. Any fuzzy system that has a set of if-then rules has to combine the rules.
- Optionally, the defuzzification of the solution fuzzy set. In many (most) fuzzy systems there is a requirement that the final output be a 'crisp' number. However, for certain fuzzy paradigms the output of the system is a fuzzy set, or its associated word. This solution set is 'defuzzified' to arrive at a number.

Type-1 fuzzy sets are, in fact, crisp and not at all fuzzy, and are two dimensional. A domain value x is simply represented by a number in $[0, 1]$ – the membership grade. The methods for combining type-1 fuzzy sets in rules are also precise in nature.

Type-2 fuzzy sets, in contrast, are three dimensional. The membership grade for a given value in a type-2 fuzzy set is a type-1 fuzzy set. A formal definition of a type-2 fuzzy set is given in the following:

A type-2 fuzzy set, $\tilde{A}$, is characterized by a type-2 membership function $\mu_{\tilde{A}}(x, u)$, where $x \in X$ and $u \in J_x$ subset $JM[0, 1]$, and J_x is called the primary membership, i. e.

$$\tilde{A} = \{((x, u), \mu_{\tilde{A}}(x, u)) \mid \forall x \in X, \forall J_x \text{ subset } JM[0, 1]\}. \tag{1}$$

A useful way of looking at a type-2 fuzzy set is by considering its Footprint of Uncertainty (FOU). This is a two dimensional view of a type-2 fuzzy set. See Fig. 4 for a simple example. The shaded area represents the Union of all the J_x.
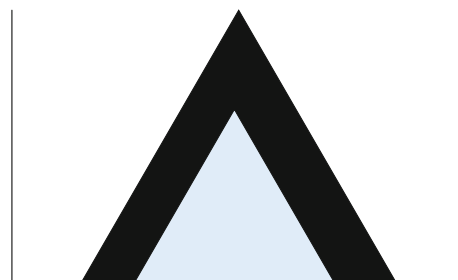
Fuzzy Logic, Type-2 and Uncertainty, Figure 4
A typical FOU of a type-2 set

An effective way to compare type-1 fuzzy sets and type-2 fuzzy sets is by use of a simple example. Suppose, for a particular application, we wish to describe the imprecise concept of 'tallness'. One approach would be to use a type-1 fuzzy set $tall_1$. Now suppose we are only considering three members of this set – Michael Jordan, Danny Devito and Robert John. For the type-1 fuzzy approach one might say that Michael Jordan is $tall_1$ to degree 0.95, Danny Devito to degree 0.4 and Robert John to degree 0.6. This can be written as

$$tall_1 = 0.95/Michael\ Jordan$$
$$+ \ 0.4/Danny\ Devito + 0.6/Robert\ John \ .$$

A type-2 fuzzy set ($tall_2$) that models the concept of 'tallness' could be

$$tall_2 = High_1/Michael\ Jordan$$
$$+ \ Low_1/Danny\ Devito + Medium_1/Robert\ John \ ,$$

where $High_1$, Low_1 and $Medium_1$ are type-1 fuzzy sets.

Figure 5 shows what the sets $High_1$, Low_1 and $Medium_1$ might look like if represented graphically. As can be seen the x axis takes values between 0 and 1, as does the y axis (μ).

Type-1 sets have an x axis representing the *domain* – in this case the height of an individual. Type-2 sets employ type-1 sets as the membership grades. Therefore, these fuzzy sets of type-2 allow for the idea that a fuzzy approach does not necessarily have membership grades $[0, 1]$ in but the degree of membership for the member is itself a type-1 fuzzy set. As can be seen, by the simple example, there is an inherent extra fuzziness offered by type-2 fuzzy sets over and above a type-1 approach. So, a type-2 fuzzy set could be called a fuzzy-fuzzy set.

Real situations do not allow for precise numbers in $[0, 1]$. In a control application, for instance, can we say that a particular temperature, t, belongs to the type-1 fuzzy set
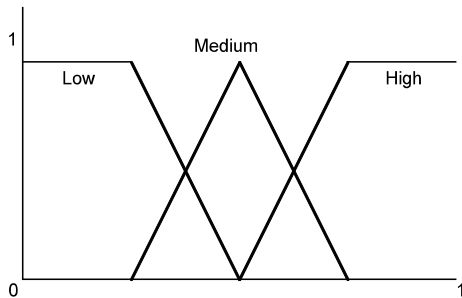
Fuzzy Logic, Type-2 and Uncertainty, Figure 5
The Fuzzy Sets $High_1$, Low_1 and $Medium_1$

hot_1 with a membership grade x precisely? No. Firstly it is highly likely that the membership could just as well be $x - 0.1$ for example. Different experts would attach different membership grades and, indeed, the same expert might well give different values on different days! On top of this uncertainty there is always some uncertainty in the measurement of t. So, we have a situation where an uncertain measurement is matched *precisely* to another uncertain value!! Type-2 fuzzy sets on the other hand, for certain appropriate applications, allow for this uncertainty to be modeled by not using precise membership grades but imprecise type-1 fuzzy sets.

So that type-2 sets can be used in a fuzzy system (in a similar manner to type-1 fuzzy sets) a method is required for computing the intersection (AND) and union (OR) of two type-2 sets. Suppose we have two type-2 fuzzy sets, $\tilde{A}$ and $\tilde{B}$ in X and $\mu_{\tilde{A}}(x)$ and $\mu_{\tilde{B}}(x)$ are two secondary membership functions of $\tilde{A}$ and $\tilde{B}$ respectively, represented as:

$$\mu_{\tilde{A}}(x) = f(u_1)/u_1 + f(u_2)/u_2 + \cdots + f(u_n)/u_n$$
$$= \sum_i f(u_i)/u_i \ ,$$
$$\mu_{\tilde{B}}(x) = g(w)/w_1 + f(w_2)/w_2 + \cdots + f(w_m)/w_m$$
$$= \sum_j g(w_j)/w_j \ ,$$

where the functions f and g are membership functions of fuzzy grades and $\{u_i, i = 1, 2, \ldots, n\}$, $\{w_j, j = 1, 2, \ldots, m\}$, are the members of the fuzzy grades.

Union of Type-2 Fuzzy Sets

The union ($\cup$) of two type-2 fuzzy sets ($\tilde{A}$, $\tilde{B}$) corresponding to $\tilde{A}$ OR $\tilde{B}$ is given by:

$$\tilde{A} \cup \tilde{B} \Leftrightarrow \mu_{\tilde{A} \cup \tilde{B}}(x) = \tilde{A} \text{ join } JM \ \tilde{B}$$
$$= \sum_{ij} \frac{(f(u_i)JMg(w_j))}{(u_iJMw_j)} \ .$$

Intersection of Type-2 Fuzzy Sets

The intersection ($\cap$) of two type-2 fuzzy sets ($\tilde{A}$, $\tilde{B}$) corresponding to $\tilde{A}$ AND $\tilde{B}$ is given by:

$$\tilde{A} \cap \tilde{B} \Leftrightarrow \mu_{\tilde{A} \cap \tilde{B}}(x) = \tilde{A} \text{ meet } JM \ \tilde{B}$$
$$= \sum_{ij} \frac{(f(u_i)JMg(w_j))}{(u_iJMw_j)} \ ,$$

where join JM denotes *join* and meet JM denotes *meet*.

Thus the join and meet allow for us to combine type-2 fuzzy sets for the situation where we wish to 'AND' or 'OR' two type-2 fuzzy sets. Join and meet are the building blocks for type-2 fuzzy relations and type-2 fuzzy inferencing with type-2 if-then rules.

Type-2 fuzzy if-then rules (type-2 rules) are similar to type-1 fuzzy if-then rules. An example type-2 if-then rule is given by

$$\text{IF } x \text{ is } \tilde{A} \text{ and } y \text{ is } \tilde{B} \text{ then } z \text{ is } \tilde{C} . \tag{2}$$

Obviously the rule could have a more complex antecedent connected by AND. Also the consequent of the rule could be type-1 or, indeed, crisp.

Type-2 output processing can be done in a number of ways through type-reduction (e. g. Centroid, Centre of Sums, Height, Modified Height and Center-of-Sets) that produces a type-1 fuzzy set, followed by defuzzification of that set.

Generalized Type-2 Fuzzy Systems

So far we have been discussing type-2 fuzzy systems where the secondary membership function can take any form – these are known as *generalized* type-2 fuzzy sets. Historically these have been difficult to work with because the complexity of the calculations is too high for real applications. Recent developments [2,3,4] mean that it is now possible to develop type-2 fuzzy systems where, for example, the secondary membership functions are triangular in shape. This is relatively new but offers an exciting opportunity to capture the uncertainty in real applications. We expect the interest in this area to grow considerably but for the purposes of this article we will concentrate on the detail of interval type-2 fuzzy systems where the secondary membership function is always unity.

Interval Type-2 Fuzzy Sets and Systems

As of this date, interval type-2 fuzzy sets (IT2 FSs) and interval type-2 fuzzy logic systems (IT2 FLSs) are most widely used because it is easy to compute using them. IT2 FSs are also known as *interval-valued FSs* for which there is a very extensive literature (e. g., [1], see the many references in this article, [5,6,16,28]). This section focuses on IT2 FSs and IT2 FLSs.

Interval Type-2 Fuzzy Sets

An IT2 FS $\tilde{A}$ is characterized as (much of the background material in this sub-section is taken from [23]; see

also [17]):

$$\tilde{A} = \int_{x \in X} \int_{u \in J_x \subseteq [0,1]} 1/(x, u)$$
$$= \int_{x \in X} \left[\int_{u \in J_x \subseteq [0,1]} 1/u \right] / x , \tag{3}$$

where x, the *primary variable*, has domain X; $u \in U$, the *secondary variable*, has domain J_x at each $x \in X$; J_x is called the *primary membership* of x and is defined below in (9); and, the *secondary grades* of $\tilde{A}$ all equal 1. Note that for continuous X and U, (3) means: $\tilde{A}: X \to \{[a, b] : 0 \leqslant a \leqslant b \leqslant 1\}$.

The bracketed term in (3) is called the *secondary MF*, or *vertical slice*, of $\tilde{A}$, and is denoted $\mu_{\tilde{A}}(x)$, i. e.

$$\mu_{\tilde{A}}(x) = \int_{u \in J_x \subseteq [0,1]} 1/u , \tag{4}$$

so that $\tilde{A}$ can be expressed in terms of its vertical slices as:

$$\tilde{A} = \int_{x \in X} \mu_{\tilde{A}}(x)/x . \tag{5}$$

Uncertainty about $\tilde{A}$ is conveyed by the union of all the primary memberships, which is called the *footprint of uncertainty* (FOU) of $\tilde{A}$ (see Fig. 6), i. e.

$$\text{FOU}(\tilde{A}) = \underset{\forall x \in X}{\cup} J_x = \{(x, u): u \in J_x \subseteq [0, 1]\} . \tag{6}$$

The *upper membership function* (UMF) and *lower membership function* (LMF) of $\tilde{A}$ are two type-1 MFs that bound the FOU (Fig. 6). UMF($\tilde{A}$) is associated with the upper bound of FOU($\tilde{A}$) and is denoted $\bar{\mu}_{\tilde{A}}(x)$, $\forall x \in X$, and LMF($\tilde{A}$) is associated with the lower bound of FOU($\tilde{A}$) and is denoted $\underline{\mu}_{\tilde{A}}(x)$, $\forall x \in X$, i. e.

$$\text{UMF}(\tilde{A}) \equiv \bar{\mu}_{\tilde{A}}(x) = \overline{\text{FOU}(\tilde{A})} \quad \forall x \in X , \tag{7}$$

$$\text{LMF}(\tilde{A}) \equiv \underline{\mu}_{\tilde{A}}(x) = \underline{\text{FOU}(\tilde{A})} \quad \forall x \in X . \tag{8}$$
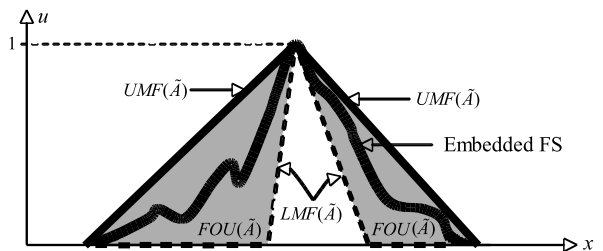
Fuzzy Logic, Type-2 and Uncertainty, Figure 6
FOU (*shaded*), LMF (*dashed*), UMF (*solid*) and an embedded FS (*wavy line*) for IT2 FS $\tilde{A}$

Note that J_x is an *interval set*, i. e.

$$J_x = \left[\underline{\mu}_{\tilde{A}}(x), \bar{\mu}_{\tilde{A}}(x) \right] . \tag{9}$$

This set is discrete when U is discrete and is continuous when U is continuous. Using (9), the $FOU(\tilde{A})$ in (6) can also be expressed as

$$\text{FOU}(\tilde{A}) = \bigcup_{\forall x \in X} \left[\underline{\mu}_{\tilde{A}}(x), \bar{\mu}_{\tilde{A}}(x) \right] . \tag{10}$$

A very compact way to describe an IT2 FS is [22]:

$$\tilde{A} = 1/\text{FOU}(\tilde{A}) , \tag{11}$$

where this notation means that the secondary grade equals 1 for all elements of $\text{FOU}(\tilde{A})$.

For continuous universes of discourse X and U, an *embedded* IT2 FS $\tilde{A}_e$ is

$$\tilde{A}_e = \int_{x \in X} [1/u]/x \quad u \in J_x . \tag{12}$$

Note that (12) means: $\tilde{A}_e \colon X \to \{u \colon 0 \leqslant u \leqslant 1\}$. The set $\tilde{A}_e$ is embedded in $\tilde{A}$ such that at each x it only has one secondary variable (i. e., one primary membership whose secondary grade equals 1). Examples of $\tilde{A}_e$ are $1/\bar{\mu}_{\tilde{A}}(x)$ and $1/\underline{\mu}_{\tilde{A}}(x)$, $\forall x \in X$. In this notation it is understood that the secondary grade equals 1 at all elements in $\underline{\mu}_{\tilde{A}}(x)$ or $\bar{\mu}_{\tilde{A}}(x)$.

For discrete universes of discourse X and U, in which x has been discretized into N values and at each of these values u has been discretized into M_i values, an *embedded* IT2 FS $\tilde{A}_e$ has N elements, where $\tilde{A}_e$ contains exactly one element from $J_{x_1}, J_{x_2}, \ldots, J_{x_N}$, namely $u_1, u_2, \ldots, u_N$, each with a secondary grade equal to 1, i. e.,

$$\tilde{A}_e = \sum_{i=1}^{N} [1/u_i]/x_i , \tag{13}$$

where $u_i \in J_{x_i}$. Set $\tilde{A}_e$ is embedded in $\tilde{A}$, and, there are a total of $n_A = \prod_{i=1}^{N} M_i$ embedded T2 FSs.

Associated with each $\tilde{A}_e$ is an *embedded* T1 FS A_e, where

$$A_e = \int_{x \in X} u/x \quad u \in J_x . \tag{14}$$

The set A_e, which acts as the domain for $\tilde{A}_e$ (i. e., $\tilde{A}_e = 1/A_e$) is the union of all the primary memberships of the set $\tilde{A}_e$ in (12). Examples of A_e are $\bar{\mu}_{\tilde{A}}(x)$ and $\underline{\mu}_{\tilde{A}}(x)$, $\forall x \in X$.

When the universes of discourse X and U are continuous then there is an uncountable number of embedded

IT2 and T1 FSs in $\tilde{A}$. Because such sets are only used for theoretical purposes and are not used for computational purposes, this poses no problem.

For discrete universes of discourse X and U, an *embedded* T1 FS A_e has N elements, one each from $J_{x_1}, J_{x_2}, \ldots, J_{x_N}$, namely $u_1, u_2, \ldots, u_N$, i. e.,

$$A_e = \sum_{i=1}^{N} u_i/x_i . \tag{15}$$

Set A_e is the union of all the primary memberships of set A_e and, there are a total of $\prod_{i=1}^{N} M_i$ embedded T1 FSs.

Theorem 1 (Representation Theorem (RT) [19] Specialized to an IT2 FS [22]) *For an IT2 FS, for which X and U are discrete, $\tilde{A}$ is the union of all of its embedded IT2 FSs. Equivalently, the domain of $\tilde{A}$ is equal to the union of all of its embedded T1 FSs, so that $\tilde{A}$ can be expressed as*

$$\tilde{A} = \sum_{j=1}^{n_A} \tilde{A}_e^j = 1/\text{FOU}(\tilde{A}) = 1/\sum_{j=1}^{n_A} A_e^j = 1/\sum_{i=1}^{N} u_i^j/x_i$$

$$= 1/ \bigcup_{\forall x \in X} \left\{ \underline{\mu}_{\tilde{A}}(x), \ldots, \bar{\mu}_{\tilde{A}}(x) \right\} \tag{16}$$

and if X is a continuous universe, then the infinite set $\{\underline{\mu}_{\tilde{A}}(x), \ldots, \bar{\mu}_{\tilde{A}}(x)\}$ is replaced by the interval set $[\underline{\mu}_{\tilde{A}}(x), \bar{\mu}_{\tilde{A}}(x)]$.

This RT is arguably the most important result in IT2 FS theory because it can be used as the starting point for solving all problems involving IT2 FSs. It expresses an IT2 FS in terms of T1 FSs, so that all results for problems that use IT2 FSs can be solved using T1 FS mathematics [22]. Its use leads to the structure of the solution to a problem, after which efficient computational methods must be found to implement that structural solution. Table 1 summarizes set theoretic operations and uncertainty measures all of which were computed using the RT. Additional results for similarity measures are in [25].

Interval Type-2 Fuzzy Logic Systems

An interval type-2 fuzzy logic system (IT2 FLS), which is a FLS that uses at least one IT2 FS, contains five components – fuzzifier, rules, inference engine, type-reducer and defuzzifier – that are inter-connected, as shown in Fig. 7 (the background material in this sub-section is taken from [23]). The IT2 FLS can be viewed as a mapping from inputs to outputs (the path in Fig. 7, from "Crisp Inputs" to "Crisp Outputs"), and this mapping can be expressed quantitatively as $y = f(\mathbf{x})$, and is also known as *interval*

Fuzzy Logic, Type-2 and Uncertainty, Table 1
Results for IT2 FSs

Set Theoretic Operations [22]	
Union	$\tilde{A} \cup \tilde{B} = 1/\bigcup_{\forall x \in X} \left[\underline{\mu}_{\tilde{A}}(x) \vee \underline{\mu}_{\tilde{B}}(x), \, \bar{\mu}_{\tilde{A}}(x) \vee \bar{\mu}_{\tilde{B}}(x) \right]$
Intersection	$\tilde{A} \cap \tilde{B} = 1/\bigcup_{\forall x \in X} \left[\underline{\mu}_{\tilde{A}}(x) \wedge \underline{\mu}_{\tilde{B}}(x), \, \bar{\mu}_{\tilde{A}}(x) \wedge \bar{\mu}_{\tilde{B}}(x) \right]$
Complement	$\bar{\tilde{A}} = 1/\bigcup_{\forall x \in X} \left[1 - \underline{\mu}_{\tilde{A}}(x), \, 1 - \bar{\mu}_{\tilde{A}}(x) \right]$
Uncertainty Measures [9,21,24]	
Centroid	$C_{\tilde{A}} = [c_l(\tilde{A}), c_r(\tilde{A})] = \left[\dfrac{\sum_{i=1}^{L} x_i \bar{\mu}_{\tilde{A}}(x_i) + \sum_{i=L+1}^{N} x_i \underline{\mu}_{\tilde{A}}(x_i)}{\sum_{i=1}^{L} \bar{\mu}_{\tilde{A}}(x_i) + \sum_{i=L+1}^{N} \underline{\mu}_{\tilde{A}}(x_i)}, \, \dfrac{\sum_{i=1}^{R} x_i \underline{\mu}_{\tilde{A}}(x_i) + \sum_{i=R+1}^{N} x_i \bar{\mu}_{\tilde{A}}(x_i)}{\sum_{i=1}^{R} \underline{\mu}_{\tilde{A}}(x_i) + \sum_{i=R+1}^{N} \bar{\mu}_{\tilde{A}}(x_i)} \right]$ L and R computed using the KM Algorithms in Table 2.
Cardinality	$P_{\tilde{A}} = [p_l(\tilde{A}), p_r(\tilde{A})] = [p(\underline{\mu}_{\tilde{A}}(x)), p(\bar{\mu}_{\tilde{A}}(x))], \quad p(B) = |X| \sum_{i=1}^{N} \mu_B(x_i)/N$
Fuzziness	$F_{\tilde{A}} = [f_1(\tilde{A}), f_2(\tilde{A})] = [f_1(A_{e1}), f_2(A_{e2})], \quad f(A) = h\left(\sum_{i=1}^{N} g(\mu_A(x_i)) \right)$ $A_{e1} : \mu_{A_{e1}}(x) = \begin{cases} \bar{\mu}_{\tilde{A}}(x) & \bar{\mu}_{\tilde{A}}(x) \text{ is further away from 0.5 than } \underline{\mu}_{\tilde{A}}(x) \\ \underline{\mu}_{\tilde{A}}(x) & \text{otherwise} \end{cases}$ $A_{e2} : \mu_{A_{e2}}(x) = \begin{cases} \bar{\mu}_{\tilde{A}}(x) & \text{both } \bar{\mu}_{\tilde{A}}(x) \text{ and } \underline{\mu}_{\tilde{A}}(x) \text{ are below 0.5} \\ \underline{\mu}_{\tilde{A}}(x) & \text{both } \bar{\mu}_{\tilde{A}}(x) \text{ and } \underline{\mu}_{\tilde{A}}(x) \text{ are above 0.5} \\ 0.5 & \text{otherwise} \end{cases}$
Variance	$V_{\tilde{A}} = [v_l(\tilde{A}), v_r(\tilde{A})] = \left[\min_{\forall A_e} v_{\tilde{A}}(A_e), \max_{\forall A_e} v_{\tilde{A}}(A_e) \right]$ $v_{\tilde{A}}(A_e) = \sum_{i=1}^{N} [x_i - c(\tilde{A})]^2 \mu_{A_e}(x_i) / \sum_{i=1}^{N} \mu_{A_e}(x_i), \quad c(\tilde{A}) = [c_l(\tilde{A}) + c_r(\tilde{A})]/2$ KM Algorithms are used to compute $v_l(\tilde{A})$ and $v_r(\tilde{A})$
Skew	$S_{\tilde{A}} = [s_l(\tilde{A}), s_r(\tilde{A})] = \left[\min_{\forall A_e} s_{\tilde{A}}(A_e), \max_{\forall A_e} s_{\tilde{A}}(A_e) \right]$ $s_{\tilde{A}}(A_e) = \sum_{i=1}^{N} [x_i - c(\tilde{A})]^3 \mu_{A_e}(x_i) / \sum_{i=1}^{N} \mu_{A_e}(x_i), \quad c(\tilde{A}) = [c_l(\tilde{A}) + c_r(\tilde{A})]/2$ KM Algorithms are used to compute $s_l(\tilde{A})$ and $s_r(\tilde{A})$
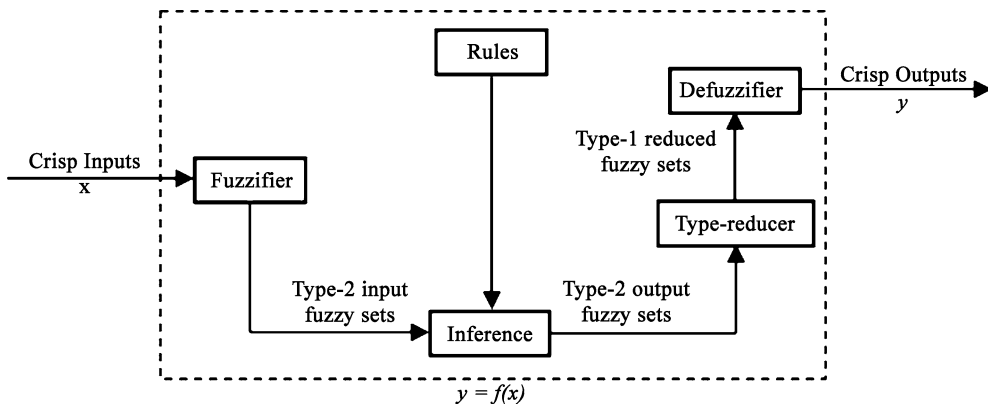
Fuzzy Logic, Type-2 and Uncertainty, Figure 7
Type-2 fuzzy logic system

type-2 fuzzy logic controller (IT2 FLC) [7], *interval type-2 fuzzy expert system, or interval type-2 fuzzy model.*

The inputs to the IT2 FLS prior to fuzzification may be certain (e. g., perfect measurements) or uncertain (e. g., noisy measurements). T1 or IT2 FSs can be used to model the latter measurements.

The IT2 FLS works as follows: the crisp inputs are first fuzzified into either type-0 (known as *singleton fuzzifica-*

tion), type-1 or IT2 FSs, which then activate the inference engine and the rule base to produce output IT2 FSs. These IT2 FSs are then processed by a type-reducer (which combines the output sets and then performs a centroid calculation), leading to an interval T1 FS called the *type-reduced set*. A defuzzifier then defuzzifies the type-reduced set to produce crisp outputs.

Rules are the heart of a FLS, and may be provided by experts or can be extracted from numerical data. In either case, rules can be expressed as a collection of IF-THEN statements. A *multi-input multi-output* (MIMO) rule base can be considered as a group of *multi-input single-output* (MISO) rule bases; hence, it is only necessary to concentrate on a *MISO* rule base. Consider an IT2 FLS having p inputs $x_1 \in X_1, \ldots, x_p \in X_p$ and one output $y \in Y$. We assume there are M rules where the ith rule has the form

$$R^i: \text{ IF } x_1 \text{ is } \tilde{F}_1^i \text{ and } \ldots \text{ and } x_p \text{ is } \tilde{F}_p^i,$$
$$\text{THEN } y \text{ is } \tilde{G}^i \quad i = 1, \ldots, M. \quad (17)$$

This rule represents a T2 relation between the input space, $X_1 \times \cdots \times X_p$, and the output space, Y, of the IT2 FLS. Associated with the p antecedent IT2 FSs, $\tilde{F}_k^i$, are the IT2 MFs $\mu_{\tilde{F}_k^i}(x_k)(k = 1, \ldots, p)$, and associated with the consequent IT2 FS $\tilde{G}^i$ is its IT2 MF $\mu_{\tilde{G}^i}(y)$.

The major result for an interval singleton T2 FLS, i.e. an IT2 FLS in which inputs are modeled as perfect measurements (type-0 FSs, singleton defuzzification) is summarized in the following:

Theorem 2 [11,15] *In an interval singleton T2 FLS using product or minimum t-norm, for input $\mathbf{x} = \mathbf{x}'$:*

(a) The result of the input and antecedent operations, is an IT1 set called the firing set, i.e.,

$$F^i(\mathbf{x}') = [\underline{f}^i(\mathbf{x}'), \overline{f}^i(\mathbf{x}')] \equiv [\underline{f}^i, \overline{f}^i]$$
$$= \left[\underline{\mu}_{\tilde{F}_1^i}(x_1') * \cdots * \underline{\mu}_{\tilde{F}_p^i}(x_p'), \bar{\mu}_{\tilde{F}_1^i}(x_1') * \cdots * \bar{\mu}_{\tilde{F}_p^i}(x_p') \right]. \quad (18)$$

(b) The rule R^i fired output consequent set, $\mu_{\tilde{B}^i}(y)$, is the IT2 FS

$$\mu_{\tilde{B}^i}(y) = \int_{b^i \in \left[\underline{f}^i * \underline{\mu}_{\tilde{G}^i}(y), \overline{f}^i * \bar{\mu}_{\tilde{G}^i}(y) \right]} 1/b^i, \quad y \in Y, \quad (19)$$

where $\underline{\mu}_{\tilde{G}^i}(y)$ and $\bar{\mu}_{\tilde{G}^i}(y)$ are the lower and upper membership grades of $\mu_{\tilde{G}^i}(y)$.

(c) Suppose that N of the M rules in the IT2 FLS fire, where $N \le M$, and the combined output fuzzy set, $\mu_{\tilde{B}}(y)$, is obtained by combining the fired output consequent sets

by taking the union of the rule R^i fired output consequent sets; then,

$$\mu_{\tilde{B}}(y) = \int_{b \in} \left[\begin{array}{c} \left[\underline{f}^1 * \underline{\mu}_{\tilde{G}^1}(y) \right] \vee \cdots \vee \left[\underline{f}^N * \underline{\mu}_{\tilde{G}^N}(y) \right], \\ \left[\overline{f}^1 * \bar{\mu}_{\tilde{G}^1}(y) \right] \vee \cdots \vee \left[\overline{f}^N * \bar{\mu}_{\tilde{G}^N}(y) \right] \end{array} \right]^{1/b},$$
$$y \in Y. \quad (20)$$

We do not necessarily advocate taking the union of these sets. Part (c) of this theorem merely illustrates the calculations if one chooses to do this. Generalizations of this theorem to an input that is a T1 or an IT2 FS are also given in [11,15] and [22].

In Fig. 7, the *type-reduced set* provides an interval of uncertainty for the output of an IT2 FLS, in much the same way that a confidence interval provides an interval of uncertainty for a probabilistic system. The more uncertainties that occur in an IT2 FLS, which translate into more uncertainties about its MFs, the larger will be the type-reduced set, and vice-versa.

Five different type-reduction (TR) methods are described in [10,15]. Each is inspired by what is done in a T1 FLS [when the (combined) output of the inference engine is defuzzified using a variety of defuzzification methods that all do some sort of centroid calculation] and are based on computing the *centroid of an* IT2 FS. Center-of-sets, centroid, center-of-sums, and height type-reduction can all be expressed as

$$Y_{\text{TR}}(\mathbf{x}') = [y_l(\mathbf{x}'), y_r(\mathbf{x}')] \equiv [y_l, y_r]$$
$$= \int_{y^1 \in [y_l^1, y_r^1]} \cdots \int_{y^M \in [y_l^M, y_r^M]} \int_{f^1 \in \left[\underline{f}^1, \overline{f}^1 \right]} \quad (21)$$
$$\cdots \int_{f^M \in \left[\underline{f}^M, \overline{f}^M \right]} 1 / \frac{\sum_{i=1}^M f^i y^i}{\sum_{i=1}^M f^i},$$

where the multiple integral signs denote the union operation. For a detailed explanation of (21) see [15] and [10]. The most widely used TR is center-of-sets (COS) TR, for which: y_l^i and y_r^i are the left and right end points of the centroid of the consequent of the ith rule [the centroids of all consequent IT2 FSs can be pre-computed using the KM Algorithms (Table 2) and stored for COS TR]; $\underline{f}^i$ and $\overline{f}^i$ are the lower and upper firing degrees of the ith rule, computed using (18); and M is the number of fired rules. For other kinds of TR methods, $y_l^i, y_r^i, \underline{f}^i, \overline{f}^i$ and M have different meanings, and are summarized in Table I of [23].

The defuzzified output of the IT2 FLS is simply the average of y_l and y_r, i.e.

$$y(\mathbf{x}') = [y_l(\mathbf{x}') + y_r(\mathbf{x}')]/2. \quad (22)$$

Fuzzy Logic, Type-2 and Uncertainty, Table 2

KM Algorithms for computing the centroid end-points of an IT2 FS, $\tilde{A}$, and their properties [9,15,20]. Note that $x_1 \leqslant x_2 \leqslant \cdots \leqslant x_N$

Step	KM Algorithm for c_l $c_l = \min_{\forall \theta_i \in [\underline{\mu}_{\tilde{A}}(x_i), \bar{\mu}_{\tilde{A}}(x_i)]} \left(\sum_{i=1}^{N} x_i \theta_i / \sum_{i=1}^{N} \theta_i \right)$	KM Algorithm for c_r $c_r = \max_{\forall \theta_i \in [\underline{\mu}_{\tilde{A}}(x_i), \bar{\mu}_{\tilde{A}}(x_i)]} \left(\sum_{i=1}^{N} x_i \theta_i / \sum_{i=1}^{N} \theta_i \right)$
1	Initialize θ_i by setting $\theta_i = \left[\underline{\mu}_{\tilde{A}}(x_i) + \bar{\mu}_{\tilde{A}}(x_i)\right]/2$, $i = 1, \ldots, N$ (or $\theta_i = \underline{\mu}_{\tilde{A}}(x_i), i \leqslant \lfloor (n+1)/2 \rfloor$ and $\theta_i = \bar{\mu}_{\tilde{A}}(x_i), i > \lfloor (n+1)/2 \rfloor$, where $\lfloor \bullet \rfloor$ denotes the first integer equal to or smaller than $\bullet$), and then compute $c' = c(\theta_1, \ldots, \theta_N) = \sum_{i=1}^{N} x_i \theta_i / \sum_{i=1}^{N} \theta_i$	
2	Find $k(1 \leqslant k \leqslant N - 1)$ such that $x_k \leqslant c' \leqslant x_{k+1}$	
3	Set $\theta_i = \bar{\mu}_{\tilde{A}}(x_i)$ when $i \leqslant k$, and $\theta_i = \underline{\mu}_{\tilde{A}}(x_i)$ when $i \geqslant k + 1$, and then compute $c_l(k) \equiv \frac{\sum_{i=1}^{k} x_i \bar{\mu}_{\tilde{A}}(x_i) + \sum_{i=k+1}^{N} x_i \underline{\mu}_{\tilde{A}}(x_i)}{\sum_{i=1}^{k} \bar{\mu}_{\tilde{A}}(x_i) + \sum_{i=k+1}^{N} \underline{\mu}_{\tilde{A}}(x_i)}$	Set $\theta_i = \underline{\mu}_{\tilde{A}}(x_i)$ when $i \leqslant k$, and $\theta_i = \bar{\mu}_{\tilde{A}}(x_i)$ when $i \geqslant k + 1$, and then compute $c_r(k) \equiv \frac{\sum_{i=1}^{k} x_i \underline{\mu}_{\tilde{A}}(x_i) + \sum_{i=k+1}^{N} x_i \bar{\mu}_{\tilde{A}}(x_i)}{\sum_{i=1}^{k} \underline{\mu}_{\tilde{A}}(x_i) + \sum_{i=k+1}^{N} \bar{\mu}_{\tilde{A}}(x_i)}$
4	Check if $c_l(k) = c'$. If yes, stop and set $c_l(k) = c_l$ and call k k_L. If no, go to Step 5	Check if $c_r(k) = c'$. If yes, stop and set $c_r(k) = c_r$ and call k k_R. If no, go to Step 5
5	Set $c' = c_l(k)$ and go to Step 2	Set $c' = c_r(k)$ and go to Step 2
Properties of the KM Algorithms [20]		
Convergence is monotonic and super-exponentially fast.		

Because TR is iterative, it may not be possible to use it in a real-time application. Wu and Mendel [26] introduced a method to approximate the TR set by minimax uncertainty bounds. Doing this avoids the computational overheads associated with TR and, as shown in [26] and [12], provides very similar outputs to the IT2 FLSs using TR. These uncertainty bounds are $\underline{y}_l(\mathbf{x}') \leqslant y_l(\mathbf{x}') \leqslant \bar{y}_l(\mathbf{x}')$ and $\underline{y}_r(\mathbf{x}') \leqslant y_r(\mathbf{x}') \leqslant \bar{y}_r(\mathbf{x}')$, where:

$$\bar{y}_l(\mathbf{x}') = \min \left\{ \frac{\sum_{i=1}^{M} \underline{f}^i y_l^i}{\sum_{i=1}^{M} \underline{f}^i}, \frac{\sum_{i=1}^{M} \bar{f}^i y_l^i}{\sum_{i=1}^{M} \bar{f}^i} \right\}, \quad (23)$$

$$\bar{y}_r(\mathbf{x}') = \max \left\{ \frac{\sum_{i=1}^{M} \underline{f}^i y_r^i}{\sum_{i=1}^{M} \underline{f}^i}, \frac{\sum_{i=1}^{M} \bar{f}^i y_r^i}{\sum_{i=1}^{M} \bar{f}^i} \right\}, \quad (24)$$

$$\underline{y}_l(\mathbf{x}') = \bar{y}_l(\mathbf{x}') - \left[\frac{\sum_{i=1}^{M} (\bar{f}^i - \underline{f}^i)}{\sum_{i=1}^{M} \bar{f}^i \sum_{i=1}^{M} \underline{f}^i} \right.$$

$$\left. \times \frac{\sum_{i=1}^{M} \underline{f}^i (y_l^i - y_l^1) \sum_{i=1}^{M} \bar{f}^i (y_l^M - y_l^i)}{\sum_{i=1}^{M} \underline{f}^i (y_l^i - y_l^1) + \sum_{i=1}^{M} \bar{f}^i (y_l^M - y_l^i)} \right], \quad (25)$$

$$\bar{y}_r(\mathbf{x}') = \underline{y}_r(\mathbf{x}') + \left[\frac{\sum_{i=1}^{M} (\bar{f}^i - \underline{f}^i)}{\sum_{i=1}^{M} \bar{f}^i \sum_{i=1}^{M} \underline{f}^i} \right.$$

$$\left. \times \frac{\sum_{i=1}^{M} \bar{f}^i (y_r^i - y_r^1) \sum_{i=1}^{M} \underline{f}^i (y_r^M - y_r^i)}{\sum_{i=1}^{M} \bar{f}^i (y_r^i - y_r^1) + \sum_{i=1}^{M} \underline{f}^i (y_r^M - y_r^i)} \right]. \quad (26)$$

Observe that the four bounds in (23)–(26) can be computed without having to perform TR. Wu and Mendel [26] then approximate the TR set, as $[y_l(\mathbf{x}'), y_r(\mathbf{x}')] \approx [(\underline{y}_l(\mathbf{x}') + \bar{y}_l(\mathbf{x}'))/2, (\underline{y}_r(\mathbf{x}') + \bar{y}_r(\mathbf{x}'))/2]$ and compute the output of the IT2 FLS as

$$y(\mathbf{x}') = \frac{1}{2} \left[\frac{\underline{y}_l(\mathbf{x}') + \bar{y}_l(\mathbf{x}')}{2} + \frac{\underline{y}_r(\mathbf{x}') + \bar{y}_r(\mathbf{x}')}{2} \right] \quad (27)$$

(instead of as in (22)). So, by using the uncertainty bounds, they obtain both an approximate TR set as well as a defuzzified output.

Wu and Mendel [26] still use TR during the design of an IT2 FLS. They define a new objective function that trades off some RMSE with not having to perform TR during the real-time operation of the IT2 FLS. The drawback to this approach is that TR is still performed during the design step. Lynch et al. [12,13] abandon TR completely where they replace all of the IT2 FLS computations with those in (23)–(26) which gave very similar outputs to the IT2 FLSs using TR. Doing this leads to an IT2 FLS real-time architecture.

Applications

Applications for IT2 FLSs or IT2 FSs are very numerous. Those that have appeared in the literature prior to 2001 can be found in (see pp 13–14 in [15]), and those that have appeared between 2001 and 2006 can be found in [16] and (see Table 24.8 in [18]). It is worth mentioning that applications have now appeared for the following general classes of problems: approximation, clustering, communications, control, databases, decision making embedded agents, health care, hidden Markov models, knowledge mining, neural networks, pattern classification, quality control, scheduling, signal and image processing, and

spatial query. Control applications, which were the original bread-and-butter ones for T1 FLSs, are now a major focus of attention for IT2 FLSs, and even general T2 FLSs. [8] provides three important applications that demonstrate that an IT2 FLS can significantly outperform a T1 FLS. Recently, IT2 FLSs have also been implemented in hardware [14]; this should make them more attractive and accessible for industrial applications.

Future Directions

There is much work still to be done in type-2 fuzzy research. But, we see that particular areas for fruitful and interesting work will include:

1. Applications. The two broad areas that fuzzy logic is used we can categorize as control and non-control. As discussed there is already some work in using type-2 for control and we expect this to grow. However, the weakness in type-1 fuzzy logic applications is in non-control where we are trying to emulate human expertise. Type-2 fuzzy sets are well placed to help here.
2. Generalized type-2 fuzzy systems. The relatively new approaches for allowing generalized type-2 fuzzy systems is exciting and we expect many researchers to exploit these approaches in interesting applications and provide new theoretical results.
3. Computing with Words. Because words mean different things to different people, we strongly believe that type-2 fuzzy sets must be used to model words when implementing Zadeh's Computing with Words paradigm.

Bibliography

1. Bustince H (2000) Indicator of inclusion grade for interval-valued fuzzy sets: application to approximate reasoning based on interval-valued fuzzy sets. Int J Approx Reason 23:137–209
2. Coupland S (2007) Type-2 Fuzzy Sets: Geometric Defuzzification and Type-Reduction. In: Proc FOCI, pp 622–629
3. Coupland S, John RI (2005) Towards More Efficient Type-2 Fuzzy Logic Systems. In: Proc Fuzz-IEEE, pp 236–241
4. Coupland S, John RI (2007) Geometric Type-1 and Type-2 Fuzzy Logic Systems. IEEE Trans Fuzzy Syst 15(1):3–15
5. Gorzalczany MB (1987) Decision making in signal transmission problems with interval-valued fuzzy sets. Fuzzy Sets Syst 23:191–203
6. Gorzalczany MB (1988) Interval-valued fuzzy controller based on verbal model of object. Fuzzy Sets Syst 28:45–53
7. Hagras H (2004) A hierarchical type-2 fuzzy logic control architecture for autonomous mobile robots. IEEE Trans Fuzzy Syst 12:524–539
8. Hagras H (2007) Type-2 FLCs: a new generation of fuzzy controllers. IEEE Comput Intel Mag 2:30–43
9. Karnik NN, Mendel JM (2001) Centroid of a type-2 fuzzy set. Inf Sci 132:195–220
10. Karnik NN, Mendel JM, Liang Q (1999) Type-2 fuzzy logic systems. IEEE Trans Fuzzy Syst 7:643–658
11. Liang Q, Mendel JM (2000) Interval type-2 fuzzy logic systems: theory and design. IEEE Trans Fuzzy Syst 8:535–550
12. Lynch C, Hagras H, Callaghan V (2005) Embedded type-2 FLC for real-time speed control of marine and traction diesel engines. In: Proc FUZZ-IEEE, pp 347–353
13. Lynch C, Hagras H, Callaghan V (2006) Using uncertainty bounds in the design of embedded real-time type-2 neuro-fuzzy speed controller for marine diesel engines. In: Proc FUZZ-IEEE, pp 7217–7224
14. Melgarejo M, Pena-Reyes CA (2007) Implementing interval type-2 fuzzy processors. IEEE Comput Intel Mag 2:63–71
15. Mendel JM (2001) Uncertain rule-based fuzzy logic systems: introduction and new directions. Prentice-Hall, Upper Saddle River
16. Mendel JM (2007) Advances in type-2 fuzzy sets and systems. Inf Sci 177:84–110
17. Mendel JM (2007) Type-2 fuzzy sets and systems: an overview. IEEE Comput Intel Mag 2:20–29
18. Mendel JM (2008) On type-2 fuzzy sets as granular models of words. In: Pedrycz W, Skowron A, Kreinovich V (eds) Granular Computing Handbook. Wiley, London
19. Mendel JM, John RI (2002) Type-2 fuzzy sets made simple. IEEE Trans Fuzzy Syst 10:117–127
20. Mendel JM, Liu F (2007) Super-exponential convergence of the Karnik-Mendel algorithms for computing the centroid of an interval type-2 fuzzy set. IEEE Trans on Fuzzy Syst 15:309–320
21. Mendel JM, Wu H (2007) New results about the centroid of an interval type-2 fuzzy set, including the centroid of a fuzzy granule. Inf Sci 177:360–377
22. Mendel JM, John RI, Liu F (2006) Interval type-2 fuzzy logic systems made simple. IEEE Trans Fuzzy Syst 14:808–821
23. Mendel JM, Hagras H, John RI (2006) Standard background material about interval type-2 fuzzy logic systems that can be used by all authors. IEEE Computational Intelligence Society standard. http://ieee.cis.org/standards
24. Wu D, Mendel JM (2007) Uncertainty measures for interval type-2 fuzzy sets. Inf Sci 177:5378–5393
25. Wu D, Mendel JM (2008) A vector similarity measure for linguistic approximation: interval type-2 and type-1 fuzzy sets, vol 177. Inf Sci 178:381–402
26. Wu H, Mendel JM (2002) Uncertainty bounds and their use in the design of interval type-2 fuzzy logic systems. IEEE Trans Fuzzy Syst 10:622–639
27. Zadeh LA (1965) Fuzzy Sets. Inf Control 8:338–353
28. Zadeh LA (1975) The concept of a linguistic variable and its application to approximate reasoning–1. Inf Sci 8:199–249

Fuzzy Optimization

WELDON A. LODWICK, ELIZABETH A. UNTIEDT
Department of Mathematical Sciences, University of Colorado Denver, Denver, USA

Article Outline

Glossary

Glossary

Fuzzy set A set whose membership is characterized by gradualness and uniquely described by a membership function which measures the degree of membership that domain values possess with respect to belonging to the set in question.

Fuzzy set theory The theory of uncertainty associated with sets characterized by gradual membership.

Possibility theory The theory of uncertainty associated with deficiency of information.

Fuzzy number A fuzzy set whose membership function is upper semi-continuous with a unique modal value and whose domain is the set of real numbers.

Possibilistic number A variable described by a possibilistic distribution whose domain is the set of real numbers.

Optimization The mathematical field that studies normative processes.

Definition of the Subject

Fuzzy optimization is normative and as a mathematical model deals with transitional uncertainty and information deficiency uncertainty. Some literature calls these uncertainties *vagueness* and *ambiguity* respectively. Transitional uncertainty is the domain of *fuzzy set theory* while uncertainty resulting from information deficiency is the domain of *possibility theory*. Suppose one is told by one's employer to go to the airport and meet a tall female visitor at the baggage claim of the airport. The notion of "tall" is transitional uncertainty. As the passengers arrive at the baggage claim, one compares each female passenger to the ascribed characteristic (tall woman) as determined by one's evaluation of what the boss' function is for "tall woman" and all the information one might have regarding "tall women". The resulting function is used to obtain a possibility value for each female that appears at the baggage claim. This uncertainty arises from information deficiency.

An example of fuzzy and possibilistic uncertainty in optimization is in the radiation treatment plan of tumors which is a problem that seeks to deposit a tumorcidal dose to cells that are cancerous while sparing all other cells and at the same time minimizing the total amount of radiation used for the treatment. In this context fuzzy uncertainty arises in cell classification because a cell may be precancerous in which case it is both healthy and cancerous at the same time to some degree. Typically, radiation treatment modeling discretizes a CT-scan of a patient into pixels (voxels in three dimensions) where a pixel is a position in space relative to a fixed coordinate system containing the patient and the radiation machine. The nature of a particular pixel in space may be two (or more) things at once. That is, a particular pixel may be precancerous and thus cancerous and healthy at the same time. Classification into discrete states leaves open transitional states. Continuous states are often intractable. That is, given a discrete classification of cells, what type of cell a pixel models is often transitional and thus fuzzy. Optimization models that incorporate uncertainty of these types are confronted with a wider spectrum of uncertainty than merely uncertainty due to probability, that is, frequency.

Possibilistic uncertainty arises when there is the specification, "deposit 60 units of radiation at every tumor cell". The minimal radiation that kills a tumor cell is considered 60 units of radiation by the community of radiation oncologists. The number 60 is derived from mathematical models, research, experience, and expert knowledge. It (60 units) represents the best available information as to a minimum radiation dose that will kill a tumor cell. Of course, if a radiation oncologist were able to attain 59.999 or 60.001 but not 60 units at a tumor cell, this would undoubtedly be satisfactory. The 60 units is "informational" since it is derived from research results and experience whose value as a single mathematical entity, the real number 60, is not precise in fact. It is not just a matter of measurement or probabilistic/statistical uncertainty though both statistics and probability theory may play a part in the determination of the number 60. The number 60 is possibilistic rather than probabilistic (or fuzzy) since the uncertainty associated with the death of a cancer cell given a precise amount of radiation, 60 in this case, as cause/effect (certainly killing a cancerous cell as a result of delivering precisely 60 units of radiation to the cell) is derived from sources beyond frequency analysis. The death of a cancer cell, as a result of radiation, certainly must depend on the type of cancer, its stage of growth, personal genetic characteristics, amount of food in the blood supply at the time of radiation, and so on.

Mathematical models may be considered as being of two types – descriptive (such as simulation) and normative. Optimization models and problems, including fuzzy optimization, are normative in that they impose upon the mathematical system criteria that seek to extract a "best".

This exposition will clearly identify fuzzy optimization as a distinct optimization approach. While it is not exhaustive, this presentation will focus on the salient features of fuzzy optimization most related to optimization under uncertainty. In particular, fuzzy multiobjective programming, fuzzy stochastic programming, and fuzzy dynamic programming are not discussed (see, for example [62,65,86,87,112]). Since intervals (and real numbers) may be considered as fuzzy sets, interval optimization is not covered separately but considered within the family of fuzzy optimization.

Mathematical analyzes that include the normative must embody the idea of order, measure, and distance. The notion of "best" requires an order and measuring with respect to that order. Professor Lothar Collatz in a lecture titled, "Monotonicity and Optimization", given to the Mathematics Department at Oregon State University, August 5, 1974, stated, "The idea of *ordering* is more fundamental than *distance* because from ordering we can get distance but not the other way around". (my notes) The real number system contains within itself the most fundamental mathematical order. Optimal control, stochastic dominance, [97,138], and mean-variance, [89,97,117], are approaches to order functions. Since fuzzy intervals from their definition (see below) relate themselves to *sets* of real numbers, that is, graphs in $\mathbb{R}^2$ (membership functions), the order of fuzzy intervals will need to be derived from that associated with real-valued functions. Moreover, since fuzzy sets model uncertainty on one hand and amplification and flexibility on the other, it is clear that the idea of order and its derived distance (measure) generated will be flexible, that is, require choices, as we shall see. The choice that needs to be made is dependent on the semantics of the problem to a greater extent than in the deterministic setting.

Mathematical modeling involves simplification, the transformation from reality to symbols. Even after a symbolic representation of a system, the process being modeled may, for example, be nonlinear, but is only tractable if linearized. Moreover, a nonlinear process may be the correct model but its linear counterpart may yield acceptable solutions. In the case of radiation therapy models (see [82,83]), scatter, which causes nonlinearities, is typically ignored in developing radiation therapy models that determine the angles and intensities for a given radiation machine and a given patient being treated. Without scatter, the resulting model is a linear model whose results usually produce acceptable treatment plans as long as the breathing of the patient is ignored. On the other hand, obtaining the mathematical model for the location of a cell (in a fixed coordinate system that includes both the patient and the radiation machine) of a patient undergoing radiation therapy, is important to include (while scatter may not be as important in the determination of acceptable angles and intensities). Location of points in the body that are or have been affected by breathing is not a deterministic model. Nor is it a probabilistic model.

The mathematical model development for radiation therapy planning of a particular patient tumor for a particular machine not only is designed to "do the job" (kill all tumor cells while sparing healthy cells) but often adds various "normative" criteria. For example, a radiation oncologist might seek a mathematical model which will, when used, kill the tumor cells, spare the healthy cells, *and* minimize total radiation used to do this. The "minimize total radiation used" is a normative criterion. Of course, there could be many other normative criteria imposed, such as "minimize the probability that healthy cells will become cancerous by the radiation deposited from the radiation treatment".

The thesis of this presentation is two-fold. First, not all uncertainty that occurs and is incorporated in optimization models can be described by the frequency with which its parameters take on various values (probability). Fuzzy and possibility theory are necessary components in some cases (such as in radiation of tumor models). Secondly, optimization under uncertainty models at their most general symbolic level try to capture the true complexity of the systems being modeled so that it is crucial to model transitional processes as fuzzy sets, frequencies as probabilistic distributions, and information deficiencies as possibility distributions. Once this is done, a further simplification may be (usually is) involved, but at the highest symbolic level, it is crucial to be faithful to the nature of the uncertainty, since this will ultimately determine the correct semantics and correct simplifying assumptions. For example, a normative probability model may be translated into a stochastic recourse model which is transformed into a real-valued nonlinear programming problem. Knowing that the underlying process is probabilistic means that the input data must be faithful to the laws of probability even though in the end, the model is a real-valued nonlinear programming model. The solution semantics in this case are those arising from probability and thus frequency based. A fuzzy linear optimization model is often translated into a real-valued linear programming model. Knowing that the underlying processes being modeled are fuzzy means that the nature of both the input data and the semantic interpretation of the output solution are based on the laws of fuzzy set theory and not of probability. Fuzzy optimization models are frequently solved as a real-valued linear or nonlinear programming problems just as in the

case of stochastic optimization. This process, insuring that at the highest symbolic level of mathematical modeling the correct uncertainty is used and only then transforming it into a real-value model, is analogous to first modeling a nonlinear system as a nonlinear system of relations and then linearizing it. It is crucial to create the nonlinear model first (at least in principle) and then to linearize it so that the approximation that is being used is explicit. When one is clear about these steps, one is able to take into account the correct associated approximation errors and underlying assumptions which enable the appropriate interpretation of the solution output.

Likewise, normative models of fuzzy processes require, at the level of mathematical abstraction, fuzzy modeling *before* approximation and transformation to an equivalent real-valued model because the fuzzy model is most faithful to the underlying uncertainty and thus is able to adhere to the basic assumptions associated with its uncertainty type. Moreover, if the starting point is a fuzzy model, when approximations and translations are made, the sources and magnitudes of associated errors are explicit. Lastly, the solution semantics are determined by the context of the input uncertainty. The modeling of normative processes that contain transitional and/or information deficiency uncertainty should utilize fuzzy and/or possibilistic optimization. The symbolic representation faithfully executed according to the associated axioms governing the uncertainty type and its semantics is a necessary first step. That is, mathematics (or any science) tries to bare all of its underlying assumptions. Since mathematical models objectify relationships occurring in reality via symbols, the nature of uncertainty must first be made explicit and then approximated, not the other way around.

Fuzzy optimization, which for this exposition encompasses models affected by both fuzzy and possibilistic uncertainty, is one of the newest optimization fields. Its place is along side stochastic optimization within the field of optimization under uncertainty. Fuzzy optimization began in 1970 with the publication of the seminal Bellman and Zadeh paper [5]. It took three years before the next fuzzy optimization article was published in 1973 by H. Tanaka, T. Okuda, and K. Asai, [1,119], (with the full version [120]). These researchers seem to have been the first to realize the importance of alpha-levels in the mathematical analysis of fuzzy sets in general and fuzzy optimization in particular. The Tanaka, Okuda, Asai article operationalized the theoretical approach developed by Bellman and Zadeh. Independently, in 1974, H.-J. Zimmermann presented a paper at the *ORSA/TIMS* conference in Puerto Rico [147] (with the full version [148])

that not only operationalized the Bellman and Zadeh approach, but greatly simplified and clarified fuzzy optimization, so much so that Zimmermann's approach is a standard to this day. In this same period, the book by C.V. Negoita and D.A. Ralescu [92] contained a description of fuzzy optimization. C.V. Negoita and M. Sularia published in 1976 a set containment approach to fuzzy optimization [93]. From this beginning, fuzzy optimization has become a field of study in its own right with a journal devoted to the subject, *Fuzzy Optimization and Decision Making*, whose first issue came out in February of 2002. Moreover, there have been special sessions devoted solely to fuzzy/possibilistic optimization at the international fuzzy society meetings (*IFSA05*, July 2005, Beijing, China and *IFSA07* June 2007, Cancun, Mexico). There have been two special editions of the journal *Fuzzy Sets and Systems* dealing with fuzzy/possibilistic optimization, the latest being [73]. Two books with edited articles have appeared – [12,54] and there are at least six authored books devoted to fuzzy optimization – [4,60,62,65,103,116].

Thirty-five years of fuzzy optimization research has yielded a wide-ranging set of applications. It is beyond the scope of this exposition to cover applications, but, the interested reader may wish to consult the following set of references: Chap. 6 and 8 in [12], Chap. III in [54], and [4,28,33,36,41,47,53,70,72,73,75,82,83,104,105,107, 112,123,125,129,132,133].

The general fuzzy optimization model considered in this presentation is to find a fuzzy optimum of a fuzzy objective function subject to a fuzzy constraint,

$$\widetilde{\mathrm{opt}}_x \, \tilde{z} = f(\tilde{c}, x) \tag{1}$$

$$x \, \tilde{\in} \, \tilde{X} \tag{2}$$

where the tilde, ~, represents fuzzy and/or possibilistic entities or relationships which is made clear by the context and assumptions of the problem. When fuzzy uncertainty needs to be distinguished from possibilistic uncertainty, the tilde, ~, will denote fuzzy uncertainty and the circumflex, ⌢, will denote possibilistic uncertainty. For (1), $\tilde{c}$ is considered to be a known vector of uncertainty parameters characterized by a fuzzy membership function or a possibilistic distribution. The variables (unknown quantities whose values are to be determined by the model) which are denoted here by x, are often called the "decision variables" because in optimization models, it is the value of x that is being computed. For example, x may denote a quantity to be produced by a manufacturing process, or the amount to be transported from a production point to a consuming point, or the intensity of radiation for the angle represented by the variable, and so on.

Introduction

It is assumed that the reader has a basic knowledge of fuzzy set theory at the level of [59] though the basics that are needed for this exposition are set forth. Much of the introductory exposition can also be found in [71].

Basics of Fuzzy Set Theory

Fuzzy set and possibility theory were defined and developed by L. Zadeh beginning with [142] and subsequently [143], and [144]. As is now well-known, the idea was to mathematize and develop analytical tools to solve problems whose uncertainty went beyond probability theory. Classical mathematical sets, for example a set A, have the property that either an element $x \in A$ or $x \notin A$ but not both. There are no other possibilities for classical sets which are also called *crisp* sets. An interval is a classical set. L. Zadeh's idea was to relax this "all or nothing" membership in a set to allow for grades of belonging to a set. When grades of belonging are used, a fuzzy set ensues. To each fuzzy set, $\tilde{A}$, L. Zadeh associated a real-valued *membership function* $\mu_{\tilde{A}}(x)$, which takes x in the domain of interest, the universe Ω, to a value in the interval $[0, 1]$. The membership function $\mu_{\tilde{A}}(x)$ quantifies the degree to which x belongs to $\tilde{A}$ where a value of zero means that x certainly does not belong to $\tilde{A}$ and a value of one means that x certainly belongs to $\tilde{A}$.

$$\mu_{\tilde{A}}(x) \colon \mathbb{R} \to [0, 1] \,.$$

Another way of looking at a fuzzy set is as a set in $\mathbb{R}^2$ as follows.

Definition 1 A **fuzzy set** $\tilde{A}$, as a crisp set in $\mathbb{R}^2$, is the set of ordered pairs

$$\tilde{A} = \{(x, \mu_{\tilde{A}}(x))\} \subseteq \{(-\infty, \infty) \times [0, 1]\} \,. \tag{3}$$

The $\alpha - cut$ of a fuzzy set is the set

$$\tilde{A}_\alpha = \{x \mid \mu_{\tilde{A}}(x) \ge \alpha\} \,.$$

Definition 2 A **modal value** of a membership function is a domain value at which the membership function is one. A fuzzy set with at least one modal value is called **normal**. *The **support** of a membership function is the closure of $\{x \mid \mu_{\tilde{A}}(x)) > 0\}$.*

Definition 3 (see [29]) A **fuzzy interval**, $\tilde{M}$, defined by its membership function $\mu_{\tilde{M}}(\cdot)$, is a fuzzy continuous subset of the real line such that, if $x, y, z \in \mathbb{R}$, $z \in [x, y]$, then

$$\mu_{\tilde{M}}(z) \ge \min\{\mu_{\tilde{M}}(x), \mu_{\tilde{M}}(y)\} \,.$$

Like a fuzzy set, a fuzzy interval M is said to be **normal** if $\exists x \in \mathbb{R}$ such that $\mu_{\tilde{M}}(x) = 1$. The set $\{x \mid \mu_{\tilde{M}}(x) = 1\}$ is called the **core** (of the fuzzy interval).

Definition 4 A **fuzzy number** is a fuzzy interval with a unique modal value, that is, the core is a singleton.

For all that follows, fuzzy intervals will be assumed to be normal fuzzy intervals with upper semi-continuous membership functions. This means that the $\alpha - cut$ of a fuzzy interval, M_α, is a closed interval. Let $M_1 = \{x \mid \mu_{\tilde{M}}(x) = 1\} = [m_1^-, m_1^+]$, be the core of a fuzzy interval $\tilde{M}$ and the open support $M_0 = \{x \mid \mu_M(x) > 0\} = (m_0^-, m_0^+)$. For a fuzzy interval M, $\mu_M(x)$, is non-decreasing for $x \in (-\infty, m_1^-]$ and $\mu_M(x)$ is non-increasing for $x \in [m_1^+, \infty)$.

The fact that we have closed intervals at each $\alpha - cut$ means that fuzzy arithmetic can be defined by interval arithmetic (see [90]) on each $\alpha - cut$. Unbounded intervals can be handled by extended interval arithmetic. In fact, when dealing with fuzzy intervals, the operations and analysis can be considered as interval operations and analysis on $\alpha - cuts$ [71]. We define the most common types of fuzzy intervals next.

Definition 5 A **triangular fuzzy number**, $\tilde{A} = (\alpha, \beta, \gamma)$, has a membership function μ_A centered at a value α, with a support (β, γ) such that

$$\mu_A(x) = \begin{cases} 1, & x = \alpha, \\ 0, & x \le \alpha - \beta, \\ 0, & x \ge \alpha + \gamma, \\ 1 - \frac{x-\alpha}{\beta}, & x \in (\alpha - \beta, \alpha) \\ 1 - \frac{\gamma-x}{\gamma}, & x \in (\alpha, \alpha + \gamma) \,. \end{cases}$$

Definition 6 A **symmetric triangular fuzzy number**, $\tilde{A} = (\alpha, \beta)$, has a membership function μ_A centered at a value α, with a spread β such that

$$\mu_A(x) = \begin{cases} 1, & x = \alpha, \\ 0, & x \le \alpha - \beta, \\ 0, & x \ge \alpha + \beta, \\ 1 - \frac{|x-\alpha|}{\beta}, & x \in (\alpha - \beta, \alpha + \beta) \,. \end{cases}$$

Definition 7 A **trapezoidal fuzzy interval**, $\tilde{A} = (\alpha, \beta, \gamma, \delta)$, has a membership function μ_A with a core (α, β) and a support (γ, δ) such that

$$\mu_A(x) = \begin{cases} 1, & x \in [\alpha, \beta], \\ 0, & x \notin (\gamma, \delta), \\ 1 - \frac{\alpha-x}{\alpha-\gamma}, & x \in (\gamma, \alpha), \\ 1 - \frac{x-\beta}{\delta-\beta}, & x \in (\beta, \delta) \,. \end{cases}$$

Definition 8   A **L-R fuzzy interval**, $\tilde{A} = (\alpha^R, \alpha^L, \beta^R, \beta^L)_{LR}$ has a membership function μ_A and reference functions L and R. $L: [0, \inf) \to [0, 1]$ and $R: [0, \inf) \to [0, 1]$ are upper semi-continuous and strictly decreasing in the range $(0,1]$, and μ_A is defined as follows:

$$\mu_A(x) = \begin{cases} 1, & x \in [\alpha^L, \alpha^R], \\ 0, & x \notin (\alpha^L - \beta^L, \alpha^R + \beta^R), \\ L\left(\frac{\alpha^L - x}{\beta^L}\right), & x \in [\alpha^L - \beta^L, \alpha^L], \\ R\left(\frac{x - \alpha^R}{\beta^R}\right), & x \in [\alpha^R, \beta + \alpha^R]. \end{cases}$$

The following relations are applicable to fuzzy intervals:

Definition 9 (Equality)   Two fuzzy sets, $\tilde{A}$ and $\tilde{B}$ are said to be equal if and only if $\mu_A(x) = \mu_B(x)$ for all $x \in X$.

This definition, given by Bellman and Zadeh [5], is generally accepted. We explore broader interpretations of $\tilde{A} = \tilde{B}$ in Subsect. "Fuzzy Relations".

Definition 10 (Containment)   A fuzzy set $\tilde{A}$ is said to be a subset of fuzzy set $\tilde{B}$ if and only if $\mu_A(x) \le \mu_B(x)$ for all $x \in X$.

Definition 11 (Intersection)   The *intersection* of $\tilde{A}$ and $\tilde{B}$ is defined as the largest fuzzy set contained in both $\tilde{A}$ and $\tilde{B}$. The membership function of $\tilde{A} \wedge \tilde{B}$ is given by

$$\mu_{A \wedge B}(x) = \min(\mu_A(x), \mu_B(x)), \quad x \in X.$$

Minimum is only one of a continuum of operators that define intersection, but this discussion is beyond the scope of this paper. The $t - norms$ (see [59]) define a family of intersection operators.

Definition 12 (Relation)   A *fuzzy relation R* in the product space $X \times Y$ is a fuzzy set characterized by a membership function μ_R which associates with each ordered pair a grade of membership $\mu_R(x, y)$ in $\mathbb{R}$.

Definition 13 (Non-Interactivity)   Consider a fuzzy number $\tilde{C}$ where $C \subseteq \mathbb{R}^2$ which is a direct product of two fuzzy numbers $\tilde{A} \in \mathbb{R}$ and $\tilde{B} \in \mathbb{R}$ such that for all $(x, y) \in \mathbb{R}^2$,

$$\mu_C(x, y) = \mu_A(x) \wedge \mu_B(y). \tag{4}$$

If condition (4) holds for $\tilde{A}$ and $\tilde{B}$, then $\tilde{A}$ and $\tilde{B}$, are said to be non-interactive. Semantically, two numbers are non-interactive if they can be assigned values independently of each other [18]. Every fuzzy and possibilistic model examined in this paper operates on the assumption of non-interactivity of all uncertain entities.

A different and more recent approach to fuzzy entities is possible. Instead of considering a fuzzy interval as a specialized fuzzy set over the set of real numbers, $\mathbb{R}$, Dubois and Prade (see [26]) and Fortin, Dubois, and Fargier (see [29]) introduce the concept of *gradual numbers* to revise the theory of fuzzy intervals so that a (real-valued) fuzzy interval is to a (real-valued) interval what a fuzzy set is to a (classical) set. This restores the algebraic structure of the real numbers to fuzzy arithmetic. (In particular, the usual fuzzy arithmetic which uses interval arithmetic on $\alpha - cuts$ [57] lacks an additive identity, a multiplicative identity, and the distributive law, all properties of algebra of real numbers.) An earlier approach to fuzzy arithmetic which also restores the algebraic structure of real numbers to fuzzy arithmetic is constraint interval arithmetic on $\alpha - cuts$ [69,71].

Gradual numbers were developed by [26,29] not only to restore a richer algebraic structure to fuzzy intervals, but to put fuzzy intervals on a firmer foundation. The theory of gradual numbers is one in which the relationship between real-valued intervals and fuzzy intervals is made quite apparent, clear, and compelling. Special (extreme) gradual numbers serve as endpoints of fuzzy intervals in the same way that real numbers serve as the endpoints of real intervals. The theory of gradual numbers also allows one to deal with the concept of a fuzzy element (of a fuzzy set) in a theoretically sound and meaningful way. The special gradual numbers associated with the endpoints of a fuzzy interval may be thought of as being composed of two parts. The left gradual number endpoint of a fuzzy interval $\tilde{A}$ is the inverse of the membership function $\mu_{\tilde{A}}$ restricted to $(-\infty, m_1^-]$. That is, it is the inverse of

$$\mu_{\tilde{A}}^-(x) = \mu_{\tilde{A}}(x), \quad x \in (-\infty, m_1^-], \tag{5}$$

where $[m_1^-, m_1^+]$ is the core as before, which is non-empty by definition. The second part, the right gradual number endpoint of $\tilde{A}$, is the inverse of the membership function restricted to $[m_1^+, \infty)$. That is, it is the inverse of

$$\mu_{\tilde{A}}^+(x) = \mu_{\tilde{A}}(x), \quad x \in [m_1^+, \infty). \tag{6}$$

Definition 14 (see [29])   A **gradual number** $\tilde{r}$ is defined by an assignment $A_{\tilde{r}}$ from $(0, 1]$ to $\mathbb{R}$.

Note that for a fuzzy interval, $\tilde{A}$, the functions that define the endpoints, $(\mu_{\tilde{A}}^-)^{-1}(\alpha)$ (5), and $(\mu_{\tilde{A}}^+)^{-1}(\alpha)$ (6) are special cases of gradual numbers. Briefly, a gradual number in fuzzy interval $\tilde{A}$ is simply the inverse relation of any unique assignment $r(x): x \in \tilde{A}$ such that $\mu_{\tilde{A}}^-(x) \le r(x) \le \mu_{\tilde{A}}^+(x)$. For the purpose of optimization, continuous strictly-monotonic assignment functions are of interest. Fuzzy intervals may be defined from the point

of view of gradual numbers and and in this context, we have the following.

Definition 15 (see [29]) Using the notion of gradual number, a **fuzzy interval** M is an ordered pair of gradual numbers $(\tilde{m}^-, \tilde{m}^+)$ where $\tilde{m}^-$ is called the fuzzy lower bound or **left profile** and $\tilde{m}^+$ is called the fuzzy upper bound or **right profile**.

To ensure that the left and right profiles adhere to what has been previously defined as a *fuzzy interval*, several properties of $\tilde{m}^-$ and $\tilde{m}^+$ must hold. In particular (see [29]):

1. The domains of the assignment functions, $\tilde{A}_{\tilde{m}^-}$ and $\tilde{A}_{\tilde{m}^+}$, must be in $(0, 1]$.
2. $\tilde{A}_{\tilde{m}^-}$ must be increasing and $\tilde{A}_{\tilde{m}^+}$ must be decreasing.
3. $\tilde{m}^-$ and $\tilde{m}^+$ must be such that $\tilde{A}_{\tilde{m}^-} \leq \tilde{A}_{\tilde{m}^+}$.

Remark 16 Fuzzy intervals with properties 1–3 above possess well-defined inverses which are functions. Note that the endpoints of a crisp interval $[a, b]$ have constant assignments, that is, $\tilde{A}_{\tilde{m}^-}(\alpha) = a$ and $\tilde{A}_{\tilde{m}^+}(\alpha) = b, 0 < \alpha \leq 1$. These are the left and right profiles of the fuzzy interval membership function of a real-valued interval,

$$\mu_{[a,b]}(x) = \begin{cases} 1 & \text{for } x \in [a, b] \\ 0 & \text{otherwise}. \end{cases}$$

Since it is constant, it has no fuzziness in it and is simply an interval, not a fuzzy interval. An interval that possesses fuzziness has a non-decreasing, non-constant left profile and a non-increasing, non-constant right profile, but a crisp interval has no fuzziness in the left/right profiles (they are horizontal line segments as gradual numbers). That is, the left and right membership function segments are vertical line segments (no fuzziness), whose inverses are horizontal line segments. Each gradual number that is a horizontal line segment, $y = f(x) = a.0 \leq x \leq 1$, represents the real number (written as a fuzzy set),

$$\mu_a(x) = \begin{cases} 1 & \text{for } x = a \\ 0 & \text{otherwise}. \end{cases}$$

The application of gradual numbers to optimization is just beginning. Two of these are [56] and [127].

Extension Principles

Fuzzy extension principles show how to transform real-valued functions into functions of fuzzy sets on one hand and how to compute fuzzy algebraic or fuzzy transcendental expressions on the other. The meaning of fuzzy arithmetic depends directly on the extension principle since arithmetic operations are (continuous) func-

tions over the reals, assuming division by zero is not allowed, and over the extended real numbers, when division by zero is allowed. The fuzzy arithmetic that results from Zadeh's extension principle [142] and its relationship to interval analysis has an extensive recent development (see [70,71,75]). Moreover, there is an intimate interrelationship between the extension principle being used and the analysis that ensues. Since *t-norms* and *t-conorms* capture the way trade-offs among decisions are made in constrained optimization (see [58]), the way one extends union and intersection via *t-norms* and *t-conorms* will determine the constraint set.

The extension principle within the context of fuzzy set theory was first proposed, developed, and defined in [142] and [144].

Definition 17 (Extension Principle of L. Zadeh [142, 144]) Given a real-valued function $f: X \to Y$, the function over fuzzy sets $f: S(X) \to S(Y)$, where $S(X)$ (respectively $S(Y)$) is the set of all fuzzy sets of X (respectively Y) is given by

$$\mu_{f(\tilde{A})}(y) = \sup\{\mu_{\tilde{A}}(x) \mid y = f(x)\} \tag{7}$$

for all fuzzy subsets A of $S(X)$. In particular, if $(X_1, \ldots, X_n)$ is a vector of fuzzy intervals, and $f(x_1, \ldots, x_n)$ a real-valued function, then

$$\mu_{f(X_1,\ldots,X_n)} = \sup_{(x_1,\ldots,x_n)} \min_{1 \leq i \leq n} \{\mu_{X_i}(x_i)\} \mid z$$
$$= f(x_1, \ldots, x_n). \tag{8}$$

The extension principle is used to define functions of fuzzy sets by taking a function over the real numbers and extending it to a fuzzy-set-valued function with fuzzy set arguments in place of the real numbers. If $X \subset \mathbb{R}$, the set-valued function $F: A \subseteq P(X) \to P(Y)$, where $P(X)$ is the power-set of X and $P(Y)$ is the power-set of Y, is an *extension function* of real-valued function $f: X \to Y \subset \mathbb{R}$ if:

$$F(A) = \{f(x) \mid x \in A\}, \tag{9}$$

and

$$\left[\inf_{x \in A} f(x), \sup_{x \in A} f(x)\right] \subseteq F(A). \tag{10}$$

This latter condition (10) needs to be imposed when these set-valued extension functions are approximated on a computer so that even computationally, it is always true that when a set is "retracted" to a point,

$$F(\{x\}) = f(x) \quad \forall x \in X$$

where F is the (set-valued) extension of (the real-valued) f. Now, if we have a set of fuzzy subsets $\tilde{A}$ of X, to ob-

tain a resulting fuzzy set within the set of all fuzzy subsets of Y under the mapping f, the extension function of f over fuzzy sets is denoted $\tilde{F}$ and the membership function (recall that a fuzzy set is uniquely defined by its membership function) $\mu_{\tilde{F}(\tilde{A})}$ is defined by:

$$\mu_{\tilde{F}(\tilde{A})}(y) = \sup\{\mu_{\tilde{A}}(x) \mid y = f(x), x \in X, y \in Y\}$$

The definition (8) of the extension principle has led to fuzzy arithmetic. Moreover, it is one of the main mechanisms used for fuzzy (interval) analysis. Various researchers have dealt with the issue of the extension principle and amplified its applicability. H. Nguyen [98] pointed out, in his 1978 paper, that a fuzzy set needs to be defined to be what Dubois and Prade later called a fuzzy interval (see [26,29]) in order that

$$[f(\tilde{A}, \tilde{B})]_\alpha = f(A_\alpha, B_\alpha)$$

where the function f is assumed to be continuous. In particular A_α and B_α need to be compact (that is closed/bounded intervals) for each α-cut. Thus, H. Nguyen defined a fuzzy number as one whose membership function is upper semi-continuous and for which the closure of the support is compact. In this case, the α-cuts generated are closed and bounded (compact) sets, that is, real-valued intervals. This is a well-known result in real analysis. That is, when f is continuous, the decomposition by $\alpha - cuts$ can be used to compute $f(\tilde{X}_1, \ldots, \tilde{X}_n)$ via interval analysis by [98] as

$$[f(\tilde{X}_1, \ldots, \tilde{X}_n)]_\alpha = f([X_1]_\alpha, \ldots, [X_n]_\alpha).$$

It should be noted that the gradual number representation of fuzzy intervals allows use of the extension principle without resorting to $\alpha - cuts$.

R. Yager [141] pointed out that by looking at functions as graphs (in the Euclidean plane), the extension principle could be extended to include all graphs thus allowing for analysis of what he calls "non-deterministic" mappings, that is, graphs which are not functions. Now, "non-determinism" as is used by Yager can be considered as point-to-set mappings. Thus, Yager implicitly restores the extension principle to a more general setting of point-to-set mappings.

J. Ramik [100] points out that we can restore L. Zadeh's extension principle to its most general setting of set-to-set mappings explicitly. In fact, a fuzzy mapping is indeed a set-to-set mapping. He defines the image of a fuzzy set-to-set mapping as being the set of α's generated by the function on the $\alpha - cuts$ of the domain.

Lastly, T.Y. Lin's paper [61] is concerned with determining the function space in which the fuzzy set generated by the extension principle "lives". That is, to what

space does the range of the fuzzy function belong? The extension principle generates a resultant membership function in the range space. Suppose one is interested in stable controls, then one way to extend is to generate resultant (range-space) membership functions that are continuous. The definition of continuous function states that small perturbations in the input, that is, domain, cause small perturbations in the output, that is, range, which is one way to view the definition of stability. T.Y. Lin develops conditions that are necessary in order that the range membership function has some desired characteristics (such as continuity or smoothness).

The essential purpose of fuzzy extension principles is to define functions over fuzzy sets so that the resulting range maintains various properties of interest specific to both the function and its fuzzy set input. In optimization, this is crucial in computing the output of objective and constraint functions (1) and (2). The theory and computational methods associated with distribution arithmetic including fuzzy and possibilistic arithmetic is discussed in detail in [71].

Basics of Possibility Theory

Possibilistic distributions (of fuzzy intervals or sets) encapsulate the most knowledgeable estimate of the possible values of an entity given the available information. This theory was articulated in [145]. Fuzzy membership function values (of fuzzy intervals or sets) describe the degree to which an entity is that value. Note that if the possibility distribution at x is one, this signifies that the best evidence available indicates x is indeed the entity that the distribution describes. Possibility distributions constructed from first principles require nested sets (see, for example [51]) and normalization. Possibility distributions are normalized since their semantics are tied to existent entities. Since the entity exists, it is always possible for at least one x. For example, if one is hiking from an elevation of 2,000 m to an elevation of 3,000 m, one must traverse the 2,500 m isoline (at least once). That there is a spot at which this occurs is not in question. The location of the spot is and will be dependent on the information at hand (accuracy of the maps, possibly of a global positioning system, compass, altimeter, expert knowledge). Nevertheless, one knows with certainty in this case *that* the 2,500 m isoline is traversed but not *where* it is traversed.

On the other hand, normalization is not required of fuzzy membership functions. Thus, *not all fuzzy sets can give rise to possibility distributions*.

Possibility theory may be derived in at least one of the following ways:

1. Via normalized fuzzy sets (see [145]),
2. Axiomatically from fuzzy measures g that satisfy $g(A \cup B) = \max\{g(A), g(B)\}$ (see [25,59] for example),
3. Via belief functions of Dempster–Shafer theory whose focal elements are normalized and nested (see [59]),
4. By construction (via nested sets with normalization, for example nested $\alpha - level$ sets, see [25,51] and [59]).

The most general derivation of possibility theory (method 2 above) sets up an order among variables with respect to their being an entity. The *magnitudes* associated with this ordering have no significance other than an indication of order. Thus, if $possibility_A(x) = 0.75$ and $possibility_A(y) = 0.25$ all that can be said is that x more likely to be the entity A than y. One *cannot* conclude that x is three times more likely to be A than y is. This means that for optimization problems, if the possibility distributions were constructed using the most general assumptions, then comparisons among several distributions is problematic. In particular, setting the possibility level to be greater than or equal to a certain fixed value, say $0 \leq \alpha \leq 1$, does not have the same meaning as setting a probability level or a membership function to be at least α. In the former case the α has no inherently meaning (other than if one has a $\beta > \alpha$ one prefers the decision that generated β to that which generated α) whereas in the former, the value of α is meaningful. Because of this, optimization methods must assume that the possibility distributions are constructed according to probability based possibility (see [51]) since the possibilities that are so constructed do have meaningful distribution value levels. That is, if the possibility level is α, the the value of α has a quantitative (not just order) meaning in relation to all values in $[0, 1]$.

There is a companion set-valued function to possibility called *necessity*, when the measure of the underlying space is finite. The necessity set-valued function is defined by

$$\text{necessity}(\tilde{A}) = 1 - \text{possibility}(\tilde{A}^C),$$

where $\tilde{A}^C$ denotes the complement of the fuzzy set $\tilde{A}$. Semantically, the necessity of an event measures the impossibility of the opposite event.

Semantics of Fuzzy Sets and Possibility Distributions in Fuzzy Optimization

This study restricts uncertainty to parameters (input data) whose fuzzy membership functions or possibilistic distributions are over intervals of real numbers, that is, fuzzy or possibilistic intervals. That semantics is crucial in the context of fuzzy sets was known early in the development of fuzzy/possibility theory [19] and subsequently elaborated [23,27].

Fuzzy optimization is distinguished from possibilistic optimization by both semantics and optimization procedures. As will be seen below, fuzzy optimization optimizes over sets of real numbers while possibilistic optimization optimizes over sets of (possibility) distributions. In addition, optimization procedures are influenced by the fact that fuzzy and possibilistic distributions have different development when they are derived from first principles.

The semantic distinctions between fuzzy and possibilistic optimization can be found in [37,45] and [47] where it is noted that, for optimization models, *ambiguity* in the coefficients of the model leads to possibility optimization, while *vagueness* in the decision maker's preference is modeled by fuzzy optimization. When this vagueness represents a willingness on the part of the decision maker to relax his or her requirements in order to attain better results, this type of fuzzy optimization is sometimes called *flexible programming*. It has also been said that, in the context of optimization, possibilistic uncertainty is information-based uncertainty and fuzzy uncertainty is preference-based uncertainty [60]. Another point of view on the semantic distinction between fuzzy and possibilistic uncertainty in optimization is evinced in [45].

> The membership grade of a fuzzy goal (fuzzy constraint) represents the *degree of satisfaction*, whereas that of a possibility distribution represents the *degree of occurrence*.

The semantics of an optimization problem are also influenced by where in the optimization problem the uncertainty occurs. Suppose an optimization problem is known to have fuzzy uncertainty in the inequality constraints. In the case of soft constraints, it is the inequality (or equality) itself which is viewed to be fuzzy (i. e. $Ax \tilde{\leq} b$ for a linear programming problem). This is distinct from the case in which the right hand side has vague value, in which case the right hand side is viewed to be fuzzy (i. e. $Ax \leq \tilde{b}$). To illuminate the difference between fuzzy inequalities and fuzzy right hand sides, consider the example of a computer dating service. Suppose woman A specifies that she would like to date a "medium-height" man. Woman A defines "medium" as a fuzzy set characterized by a triangular membership function, centered at $69''$, with a spread of $3''$. This is a hard constraint because the man is *required* to be medium-height, but medium-height is a fuzzy set. This is, therefore, an example of a fuzzy right-hand side (i. e. $Ax = \tilde{b}$). Now consider woman B, who desires to date a man of about $69''$ in height. Unlike woman A, woman B is willing to compromise a little on the height requirement in order to be matched with a date who meets some of her other requirements. Her satisfaction level with a $69''$ man

is 1, with a 68″ or 70″ man is $\frac{2}{3}$, and with a 67″ or 71″ man is $\frac{1}{3}$. This is an example of a soft constraint, represented by a fuzzy equality (i. e. $Ax \doteq b$). Notice that the membership functions in the two fuzzy cases are the same (symmetric triangular centered at 69″ with a spread of 3″), but the semantics are different.

One result of the distinction between fuzziness and possibilistic uncertainty is that they manifest themselves in different regions of the optimization problem. Given the basic linear program,

$$\min z = c^{\mathrm{T}} x$$
$$\text{subject to:} \quad Ax \leq b , \quad\quad (11)$$
$$x \geq 0 ,$$

fuzziness can occur in the right-hand-side($\tilde{b}$), and/or in the inequality ($\tilde{\leq}$). To date, no model has been proposed which deals with fuzzy objective function coefficients of fuzzy constraint matrix coefficients. However, suppose a constraint, $a_{ij} x \leq b_i$, $i \in [1, n]$, is meant to apply only to members of a particular fuzzy set, $\tilde{Y}$. Now suppose that the element represented by row i of the constraint matrix is a member of set Y to a degree defined by $\mu_Y(y)$. Then the constraint i should be multiplied by the membership value of y_i, resulting in fuzzy coefficients. A similar argument can be applied for objective function coefficients. Possibilistic values can occur in the objective function coefficients $\hat{c}$, in the constraint matrix coefficients $\hat{A}$, and/or in the right-hand-side $\hat{b}$. To date, there is neither semantic nor model for a possibilistic inequality.

Fuzzy Relations

Optimization models usually involve constraints consisting of equalities, inequalities, or both. For deterministic optimization, the meaning and computation of the constraint set is clear. In optimization under uncertainty, however, the meaning and computation of "equality" and "inequality" must be determined. To this end, Dubois and Prade, [22,24], give a comprehensive analysis of fuzzy relations with four possible interpretations of fuzzy equalities, called modalities, and four possible interpretations of fuzzy inequalities. Inuiguchi [45] adds two more modalities. However, only the original four modalities are outlined below, using fuzzy intervals $\tilde{M} = [m^-(\alpha), m^+(\alpha)]$, $0 \leq \alpha \leq 1$, and $\tilde{N} = [n^-(\alpha), n^+(\alpha)]$, $0 \leq \alpha \leq 1$.

The statement $\tilde{M} \geq \tilde{N}$ can be interpreted in any of the four following ways:

i. $\forall x \in \tilde{M}, \ \forall y \in \tilde{N}, \ x > y$.
This is equivalent to $m^-(\alpha) > n^+(\alpha)$.

ii. $\forall x \in \tilde{M}, \ \exists y \in \tilde{N}, \ x \geq y$.
This is equivalent to $m^-(\alpha) \geq n^-(\alpha)$.

iii. $\exists x \in \tilde{M}, \ \forall y \in \tilde{N}, \ x > y$.
This is equivalent to $m^+(\alpha) > n^+(\alpha)$.

iv. $\exists(x, y) \in \tilde{M} \times \tilde{N}, \ x \geq y$.
This is equivalent to $m^+(\alpha) \geq n^-(\alpha)$.

Inequality relation (i) indicates that x is necessarily greater than y, This is the pessimistic view. The decision maker who requires that $m^- > n+$ in order to satisfy $\tilde{M} > \tilde{N}$ is taking no chances [67]. (iv) indicates that x is possibly greater than y. This is the optimistic view. The decision maker who merely requires that $m^+ > n^-$ in order to satisfy $\tilde{M} > \tilde{N}$ has a hopeful outlook. Inequality relations (ii) and (iii) fall somewhere between the optimistic and pessimistic views.

The statement $\tilde{M} = \tilde{N}$ can be interpreted in any of the following four ways:

i.   Zadeh's fuzzy set equality: $\mu_M = \mu_N$

ii. $\forall x \in \tilde{M}, \ \exists y \in \tilde{N}, \ x = y$
(which is equivalent to $\tilde{M} \subseteq \tilde{N}$).

iii. $\forall y \in \tilde{N}, \ \exists x \in \tilde{M}, \ x = y$
(which is equivalent to $\tilde{N} \subseteq \tilde{M}$).

iv. $\exists(x, y) \in \tilde{M} \times \tilde{N}, \ x = y$
(which is equivalent to $\tilde{N} \cap \tilde{M} \neq \emptyset$).

Equality relation (i) indicates that x is necessarily equal to y (the pessimistic view), (iv) indicates that x is possibly equal to y (the optimistic view), and (ii) and (iii) fall somewhere in between.

Basics of Fuzzy Sets and Possibility Theory in Optimization

The general fuzzy optimization model (1), (2), has two parts – the *objective* (normative criteria), $\widetilde{\mathrm{opt}}\,\tilde{z} = f(\tilde{c}, x)$ (1), and the *constraints*, $x \tilde{\in} \tilde{X}$ (2). This corresponds to what Kacprzyk and Orlovski state (p. 50 in [55]):

> The analysis of real decision making situation is virtually based on two types of information:

- information on feasible alternative decisions (options, choices, alternatives, variants, …),
- information making possible the comparison of alternative decisions with each other in terms of "better", "worse", "indifferent", etc.

The set of "feasible alternative decisions" is defined by the constraints while "the comparison of alternative decisions with each other" is accomplished by the objec-

tive. When the objective is a *utility function* that takes a given fuzzy set (or possibility distribution) and maps it to a subset of the real numbers and the constraints are transformed into equations and/or inequalities, the optimization model is a mathematical programming model. "Mathematical programming problems can be considered as decision making problems in which preferences between alternatives are described by means of objective function(s) on a set of alternatives given by constraints in such a way that more preferable alternatives have higher values" [102].

A key component of any mathematical programming problem is the input data. In our radiation therapy planning example, the input data includes the amount of radiation required to kill a cancerous cell, the maximum radiation a healthy cell can tolerate, and how much radiation a beam at some angle will deposit at a particular location in the body. When the input parameters of a mathematical programming model are described by uncertainty distributions (membership functions or possibility distributions), the mathematical programming problem becomes a fuzzy or possibilistic optimization problem. Recall that the decision-maker's flexibility is modeled by fuzzy relations, which results in fuzzy optimization. In addition, we sometimes see fuzzy and possibilistic parameters combined in the same problem, or possibilistic parameters occurring in the statement of fuzzy goals. These situations result in a mixed fuzzy and possibilistic optimization problem. We note that for the models considered here, while parameters, relations, and even the value of the objective function might be fuzzy or possibilistic, all decisions are "crisp". In practice, the solution to a mathematical programming problem is useless if it is not implementable.

There have been many superb surveys of the area of fuzzy optimization. Among all of these, the following are noted: [7,12,38,39,40,41,44,46,55,60,65,85,102,108,109, 112,125,130,137], and [150].

It is clear that the fuzzy optimization model (1), (2) is ill-defined. The resolution of this ill-definition depends upon the function space in which the fuzzy optimization problem is solved. To date, two types of function spaces in which fuzzy optimization is "housed" have been used.

1. Fuzzy Banach Spaces (see [16,17,48,113])
2. Real Euclidean Space $\mathbb{R}^n$ (all other approaches)

Each of these two approaches has its own way of mapping the associated transitional and/or information deficiency uncertainty onto an ordered field. This order is necessary for the determination of optimality given the inherent normativeness of optimization.

Key Issues for Fuzzy Optimization Mapped to Fuzzy Banach Spaces   Methods for solving optimization problems often involve iteration and/or approximation. For fuzzy optimization problems, a Banach space facilitates the convergence analysis of iterative or approximation algorithms. When each successive iterate or approximation remains a fuzzy entity in a fuzzy Banach space, the convergent entity or approximation is also in the fuzzy Banach space – it is again a fuzzy entity. Once convergence is achieved, a decision can be based on this fuzzy entity. When the problem is solved in a fuzzy Banach space, the fuzzy sets in the optimization problem remain fuzzy sets throughout the solution process and no translation is necessary except that of inclusion into an appropriate Banach space. Diamond and Kloeden [16,17] use a space in which their set of fuzzy sets, which they denoted $\boldsymbol{E}^n$, over $\mathbb{R}^n$ is endowed with a neighborhood system and metric that renders it a Banach space. They then find the Karush–Kuhn–Tucker (KKT) conditions for optimality (see [17]). Saito and Ishii [113] also develop the KKT conditions for optimization over fuzzy numbers. Diamond [15] and Jamison [48,49] consider equivalence classes in developing a Banach space of fuzzy sets. Jamison [48] goes on to use the Banach space developed from the equivalence classes to solve optimization problems. After the fuzzy optimization problem embedded in the fuzzy Banach space is solved, crisp decisions must be made based on fuzzy solution. This mapping from a fuzzy solution to the decision space is often called *defuzzification*. But up to the point of defuzzification all fuzzy entities from the model (1), (2) retain their fuzziness. Further discussion of fuzzy Banach space methods in fuzzy optimization is beyond the scope of this presentation.

Key Issues for Fuzzy Optimization Mapped to Real Euclidean Spaces   The ill-definition of (1), (2) may be resolved by mapping the ambiguity and/or vagueness into a complete ordered lattice, namely, the real numbers. This is accomplished by translating the fuzzy objective and fuzzy constraint into real numbers or vectors, and real-valued relations. The various possible mappings distinguish among the types of fuzzy optimization. For a particular optimization problem in the form (1), (2), the answers to the following questions will determine which mapping is used.

1. Is the optimization fuzzy, possibilistic or a mixture?
2. What is meant by fuzzy/possibilistic function $f(\tilde{c}, x)$ and how does one compute $\tilde{z} = f(\tilde{c}, x)$?
3. What is meant by a fuzzy/possibilistic relation $x \tilde{\in} \tilde{X}$? How does one compute the resulting constraint set from fuzzy relations?

4. What is meant by fuzzy/possibilistic optimization of a fuzzy-valued function $\widetilde{\text{opt}}\ \tilde{z}$?

Is the Optimization Fuzzy, Possibilistic, or a Mixture? The first issue is the nature of the uncertainty itself. This depends upon the semantics of the problem, as discussed in Subsect. "Semantics of Fuzzy Sets and Possibility Distributions in Fuzzy Optimization". Recall that fuzzy entities are sets with non-sharp boundaries in which there is a transition between elements that belong and elements that don't belong to the set. Possibilistic entities are known to exist but the evidence associated with whether a particular element belongs to the set or not is incomplete or hard to obtain.

With these definitions in mind, we briefly consider fuzzy, possibilistic, and mixed decision making. Much of what is presented next can be found in [77] and [81].

Fuzzy Decision Making: Given the set of (crisp) decisions, Ω, and fuzzy sets, $\{\tilde{F}_i \mid i = 1 \text{ to } n\}$, find the optimal decision in the set Ω. That is,

$$\sup_{x \in \Omega} h\left(\tilde{F}_1(x), \ldots, \tilde{F}_n(x)\right), \tag{12}$$

where $h: [0, 1]^n \to [0, 1]$ is an aggregation operator, often taken to be the min function, and $\tilde{F}_i(x) \in [0, 1]$ is the fuzzy membership of x in fuzzy set $\tilde{F}_i$. Note that the decision space Ω is a **crisp set** (a set of real numbers) and the optimal decision satisfies a mutual membership condition defined by the aggregation operator h. The methods of Bellman and Zadeh [5], Tanaka, Okuda and Asai [120], and Zimmermann [147,148], who were the first to develop fuzzy mathematical programming fall into this category. While the aggregation operator h historically has been the min operator, it can be, for example, any $t - norm$ that is consistent with the context of the problem and/or decision methods, for example, risk aversion (see [58,106], or [136]). For a discussion of aggregation operators see [59].

Possibilistic Decision Making: Given the set of (crisp) decisions, Ω, and the set of possibility distributions representing the uncertain outcomes from selecting decision $\vec{x} = (x_1, \ldots, x_n)^T$ denoted $\Psi_x = \{\hat{F}_x^i, i = 1, \ldots, n\}$, find the optimal decision that produces the best set of possible outcomes with respect to an ordering U of the outcomes. That is,

$$\sup_{\hat{F}_x^i \in \Psi} U(\hat{F}_x^1, \ldots, \hat{F}_x^n), \tag{13}$$

where $U(\hat{F}_x^1, \ldots, \hat{F}_x^n)$ represents a "utility" of the set of distributions of possible outcomes $\Psi = \{\Psi_x \mid x \in \Omega\}$. Note that the decision space Ψ is a **set of (possibility) dis-**

tributions $\hat{F}_x^i \colon \Omega \to [0, 1]$ resulting from taking decision $x \in \Omega$. This semantic is represented by the possibilistic optimization of [37,45,47,50].

Very simply, fuzzy decision making selects from a set of crisp elements ordered by an aggregation operator on corresponding membership functions while possibilistic decision making selects from a set of distributions measured by a utility operator that orders the corresponding distributions. These approaches have two different ordering operators (an aggregation operation like *min* for fuzzy sets and a utility function for possibility) which lead to different optimization methods (see [72]). The underlying sets associated with fuzzy decision making are fuzzy and the decision space consists of crisp elements from operations on these fuzzy sets. The underlying sets associated with possibilistic decision making are crisp and the decision space consists of distributions from operations on crisp sets.

Mixed Fuzzy/Possibilistic Decision Making: The issue of mixed fuzzy and possibility optimization problems has been studied as early as 1989 (see [11,44]). In both these early models the same α-cut defines the level of ambiguity in the possibilistic coefficients and the level at which the decision-maker's requirements are satisfied. We interpret the solution to these models to mean that we have a possibility α of obtaining a solution that satisfies the decision maker to degree α. A recently proposed model [126] allows a trade-off between the fuzzy α-level and the possibilistic α-level. This allows the decision-maker to balance the likelihood of a solution with how satisfactory the solution is.

The use of distinct approaches for mathematical programming problems in which each constraint contains only one single kind of uncertainty is found in [77] and [81]. They aggregate all fuzzy constraint rows according to (12) and find a utility for the possibilistic rows according to (13). If a mixture of uncertainty occurs within one constraint, they apply interval-valued probability measure (IVPM) [78] and [79] to optimization. IVPM applied to optimization is discussed as a separate topic in Sect. "Future Directions".

What is Meant by a Fuzzy-Valued Function and How Does one Compute $\tilde{z} = f(\tilde{c}, x)$? Any function of fuzzy sets (fuzzy-valued function) or possibility distributions must rely on an extension principle, as described in Subsect. "Extension Principles" [71,76], and independently [3]. These authors present practical methods for computing the output of a fuzzy – or possibilistic-valued functions. For this presentation, it is assumed that (8) is used to evaluate fuzzy functions.

What is Meant by a Fuzzy/Possibilistic relation? There are several possible interpretations of fuzzy and possibilistic equalities and inequalities as detailed in Subsect. "Fuzzy Relations". Other comparisons of fuzzy sets may be found in [22,140]. The choice of optimization model for a particular problem depends heavily on which of these relations is used.

What is Meant by Fuzzy Optimization of a Fuzzy-Valued Function? There are three distinct kinds of optimality a fuzzy mathematical program might pursue. *First*, a program might seek to find, among a collection of fuzzy sets, the optimal fuzzy set. This optimality depends on the order dictated by the fuzzy relations described in the previous section.

Secondly, a program may seek to maximize the membership function (or possibility) of the solution chosen for the problem. This typically happens when no fuzzy set satisfies the constraint at a full membership level. Note that $x \tilde{\in} \tilde{X}$ occurs in the transformation of the objective as well as the constraints. This is because some models allow constraint violations at a cost (to the objective function). There are several possible interpretations of fuzzy and possibilistic equalities and inequalities, as detailed in Subsect. "Fuzzy Relations". The choice of optimization model for a particular problem depends heavily on which of these relations is used.

Thirdly, fuzzy optimization problems that use $\mathbb{R}^n$ as the basic space require a mapping of (1) and (2) from fuzzy sets and/or possibility distributions to real numbers and vectors. When the sense of optimization is fuzzy, $\widetilde{\text{opt}}$, optimization is interpreted as a goal, that is, the objective function is considered as a goal. Typically, a target is set and the objective is to come as close as possible to the target. For this presentation, $\widetilde{\text{opt}}$ will be considered as optimization over real numbers. Thus, the general fuzzy optimization problem is:

$$\underset{x}{\text{opt}}\, \tilde{z} = f(\tilde{c}, x) \tag{14}$$

$$x \tilde{\in} \tilde{X}. \tag{15}$$

To transform (14) and (15) to $\mathbb{R}^n$, a mapping $T: E^n \to \mathbb{R}^{n+1}$ is defined by:

$$\underset{x}{\text{opt}}\, T\begin{pmatrix} f(\tilde{c}, x) \\ x \tilde{\in} \tilde{X} \end{pmatrix} = \underset{x}{\text{opt}} \begin{bmatrix} T_1(f(\tilde{c}, x), x \tilde{\in} \tilde{X}) \\ T_2(x \tilde{\in} \tilde{X}) \end{bmatrix}$$
$$= \begin{bmatrix} \underset{x}{\text{opt}}\, F(c, x) \in \mathbb{R} \\ x \in \Omega \subseteq \mathbb{R}^n \end{bmatrix}. \tag{16}$$

Note that $x \tilde{\in} \tilde{X}$ occurs in the transformation of the objective as well as the constraints. This is because some

fuzzy/possibilistic models consider violations of constraints as possible but at a penalty or cost (to the objective function). To make the discussion clearer, the general fuzzy optimization problem (1), (2) is restricted to the fuzzy linear programming model

$$\min \tilde{z} = \tilde{c}^{\mathrm{T}} x$$
$$\text{subject to: } \tilde{A}x \tilde{\leq} \tilde{b} \tag{17}$$
$$x \geq 0.$$

In this context, the *objective* $\widetilde{\text{opt}}\, \tilde{z} = f(\tilde{c}, x)$ becomes

$$\min \tilde{z} = f(\tilde{c}, x) = \tilde{c}^{\mathrm{T}} x. \tag{18}$$

The *constraint* $x \tilde{\in} \tilde{X}$ is

$$\tilde{X} = \{\tilde{A}x \tilde{\leq} \tilde{b}\} \cup \{x \geq 0\}. \tag{19}$$

Interactivity To date, most fuzzy optimization models assume that the fuzzy and possibilistic parameters are non-interactive. Inuiguchi has studied fuzzy optimization with interactivity (dependencies) [39,43]. The issue of interactivity is especially important for mathematical programming models in finance, such as portfolio models where groups of stocks are in fact known to be dependent. However, for this presentation, it is assumed that all uncertainties are non-interactive.

Classical Approaches to Fuzzy Optimization

There have been numerous formulations set forth for modeling imprecise objectives and constraints, each with associated input semantics and solution interpretations. This chapter organizes and reviews a representative selection of fuzzy and possibilistic formulations.

Fuzzy Programming

Vague parameter values leads to fuzzy programming. When the vagueness represents a willingness on the part of the decision-maker to bend the constraints, fuzzy programming can also be called flexible programming. In these cases, the decision-maker is willing to lower the feasibility requirements in order to obtain a more satisfactory objective function value, or, in some cases, simply in order to reach a feasible solution. Such flexible constraints are commonly referred to as *soft constraints*.

Bellman and Zadeh The landmark paper [5] by Bellman and Zadeh is the first to propose an approach to mathematical programming with fuzzy components. Conventional mathematical programming has three principal
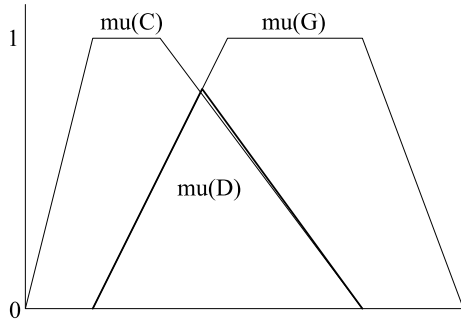
Fuzzy Optimization, Figure 1
Fuzzy Decision Membership Function μ_D

components: A set of alternatives, X, a set of constraints, C, which limit the choice of alternatives, and an objective function, z, which measures the desirability of each alternative. Bellman and Zadeh propose that, in a fuzzy environment, a more natural framework is one in which goal(s) replace the objective function(s), and a symmetry between goals and constraints erases the differences between them. (For this reason, flexible programming is sometimes called symmetric programming.)

Each fuzzy goal, $\tilde{G}_i$, or constraint, $\tilde{C}_j$, is defined as a fuzzy subset of the solution alternatives, X, via a membership function ($\mu_{G_i}(x)$ or $\mu_{C_j}(x)$). Then a fuzzy decision, $\tilde{D}$, is defined as a fuzzy set resulting from the intersection of $\tilde{C}$ and $\tilde{G}$, and is characterized by membership function μ_D as follows:

$$\mu_D(x) = \mu_C(x) \wedge \mu_G(x)$$
$$= \min[\mu_C(x), \mu_G(x)]. \tag{20}$$

An optimal decision, in turn, is one which maximizes μ_D. Figure 1 illustrates and provides a visual interpretation of μ_G, μ_C, and μ_D.

Tanaka, Okuda, and Asai [120] Tanaka, Okuda, and Asai suggested an implementation of Bellman and Zadeh's fuzzy decision using α-cuts [120]. In the literature, α-cut, α-level, and α-set are used synonymously. If $\tilde{C}$ is a fuzzy constraint in X, then an α-cut of $\tilde{C}$, denoted by C_α, is the following crisp set in X:

$$C_\alpha = \{x \mid \mu_C(x) \geq \alpha\} \quad \text{for } \alpha \in (0, 1]$$
$$C_\alpha = cls\{x \mid \mu(x) > 0\} \quad \text{for } \alpha = 0 \tag{21}$$

where $cls\{A\}$ denotes the closure of set $\{A\}$. Tanaka, Okuda and Asai [120] make the following remarkable observation:

$$\sup_x \mu_D(x) = \sup_\alpha [\alpha \wedge \max_{C_\alpha} \mu_G(x)].$$

For illustration, consider two fuzzy sets, $\tilde{C}$ and $\tilde{G}$, depicted in Fig. 2.

In case (i), $\alpha = \alpha_1$, and C_α is the interval between the end-points of the α-cut, $[C^-(\alpha_1), C^+(\alpha_1)]$. The maximum μ_G in this interval is shown in the example. In this case, $\alpha_1 < \max_{C_{\alpha_1}} \mu_G(x)]$, so $[\alpha_1 \wedge \max_{C_{\alpha_1}} \mu_G(x)] = \alpha_1$. In case (ii), $\alpha_2 > \max_{C_{\alpha_2}} \mu_G(x)$, so $[\alpha_2 \wedge \max_{C_{\alpha_2}} \mu_G(x)] = \max_{C_{\alpha_2}} \mu_G(x)$. In case (iii), $\alpha^* = \max_{C_{\alpha^*}} \mu_G(x)$. It should be apparent from Fig. 2 that $\alpha^* = \sup_x \mu_D$. In case (iii), $\alpha = \alpha^*$ is also $\sup_\alpha [\alpha \wedge \max_{C_\alpha} \mu_G(x)]$. For any $\alpha < \alpha_*$, we have case (i), where $[\alpha_1 \wedge \max_{C_{\alpha_1}} \mu_G(x)] = \alpha_1 < \alpha^*$; and for any $\alpha > \alpha_*$, we have case (ii), where $[\alpha_2 \wedge \max_{C_{\alpha_2}} \mu_G(x)] = \max_{C_{\alpha_2}} \mu_G(x) < \alpha^*$. The formal proof, which follows the reasoning illustrated here pictorially, is omitted for brevity's sake. However, the interested reader is referred to Tanaka, Okuda, and Asai [120].

This result allows the following reformulation of the fuzzy mathematical problem:

Determine (α^*, x^*)
$$\alpha^* \wedge f(x^*) = \sup_\alpha [\alpha \wedge \max_{x \in C_\alpha} f(x)]. \tag{22}$$
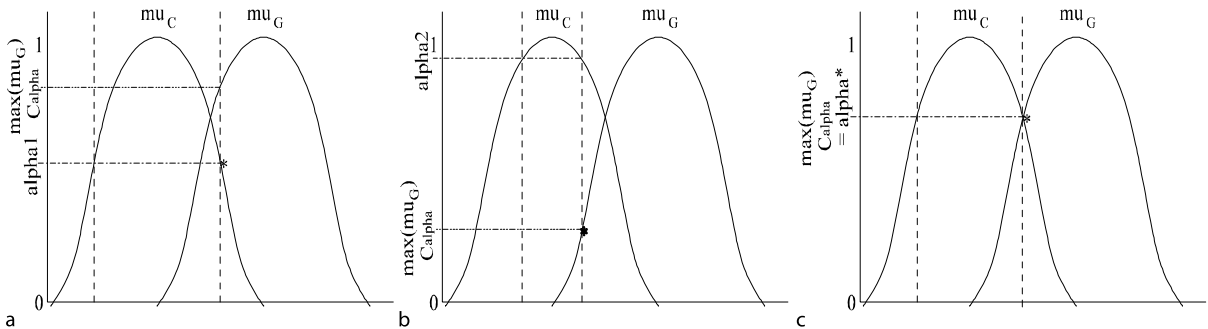
Fuzzy Optimization, Figure 2
Illustration of a Bellman/Zadeh Fuzzy Decision Set

The researchers suggest an iterative algorithm which solves the problem. The algorithm cycles through a series of steps, each of which brings it closer to a solution to the relation $\alpha^* = \max_{C_\alpha^*} f(X)$. When α^* is determined to within a tolerable degree of uncertainty, it is used to find x^* such that $f(x*) = \max_{C_\alpha^*} f(X)$.

This cumbersome solution method is impractical for large-scale problems. It should be noted that Negoita [93] suggested an alternative, but similarly complex, algorithm for the flexible programming problem based on Tanaka's findings.

Zimmermann   Just two years after Tanaka, Okuda, and Asai [120] suggested the use of α-cuts to solve the fuzzy mathematical problem, Zimmermann published a linear programming equivalent to the α-cut formulation.

Beginning with the crisp programming problem,

$$\min Z = cx$$
$$\text{subject to: } Ax \leq b \tag{23}$$
$$x \geq 0\,,$$

the decision maker introduces flexibility in the constraints, and sets a target value for the objective function, Z,

$$cx \lesssim Z$$
$$Ax \lesssim b \tag{24}$$
$$x \geq 0\,.$$

A linear membership function μ_i is defined for each flexible right hand side, including the goal Z as

$$\mu_i \left(\sum_j a_{ij} x_j \right)$$
$$= \begin{cases} 1 & \sum_j a_{ij} x_j \leq b_i, \\ \frac{1 - \left(\sum_j a_{ij} x_j - b_i \right)}{d_i} & \sum_j a_{ij} x_j \in (b_i, b_i + d_i), \\ 0 & \sum_j a_{ij} x_j \geq b_i + d_i. \end{cases}$$

where d_i is the decision maker's maximum allowable violation of constraint (or goal) i.

According to the convention set forth by Bellman and Zadeh, the fuzzy decision D is characterized by

$$\mu_D = \min_i \mu_i \left(\sum_j a_{ij} x_j \right), \tag{25}$$

and

$$\max_{x \geq 0} \min_i \mu_i \left(\sum_j a_{ij} x_j \right) \tag{26}$$

is the decision with the highest degree of membership.

The problem of finding the solution is therefore

$$\max \mu_D(x)$$
$$\text{subject to: } x \geq 0. \tag{27}$$

In the membership functions μ_i from (25), Zimmermann substitutes b_i' for b_i/d_i and B_i' for $\sum_j a_{ij}/d_i$. He also drops the 1 (which does not change the solution to the problem) to obtain the simplification $\mu_i = b_i' - B_i'x$. Equation (27) then becomes

$$\min_i (b_i' - B_i'x)$$
$$x \geq 0. \tag{28}$$

This is equivalent to the following linear program:

$$\max \alpha$$
$$\alpha \leq b_i' - B_i'x \quad \forall\, i \tag{29}$$
$$x \geq 0, \quad 0 \leq \alpha \leq 1\,.$$

Equation (29) can easily be solved via the simplex or interior point methods.

Verdegay   The solutions examined so far for the fuzzy programming problem have been crisp solutions. Ralescu [99] first suggested that a fuzzy problem should have a fuzzy solution and Verdegay [130] proposes a method for obtaining a fuzzy solution. Verdegay considers a problem with fuzzy constraints,

$$\max z = f(x)$$
$$\text{subject to: } x \subseteq \tilde{C}\,, \tag{30}$$

where the set of constraints have a membership function $\mu_{\tilde{C}}$, with alpha-cuts $\tilde{C}_\alpha$.

Verdegay defines x_α as the set of solutions that satisfy constraints $\tilde{C}_\alpha$. Then a fuzzy solution to the fuzzy linear programming problem is

$$\max_{x \in \tilde{C}_\alpha} z = f(x)$$
$$\forall\, \alpha \in [0, 1]\,. \tag{31}$$

Verdegay proposes solving (31) parametrically for $\alpha \in [0, 1]$ to obtain a fuzzy solution $\tilde{X}$, with α-cut x_α, which yields fuzzy objective value $\tilde{z}$, with α-cut z_α.

It should be noted that the (crisp) solution obtained via Zimmermann's method corresponds to Verdegay's solution in the following way: If the objective function is transformed into a goal, with membership function μ_G, then Zimmermann's optimal solution, x^*, is equal to Verdegay's optimal value $x(\alpha)$ for the value of α which satisfies

$$\mu_G(z_\alpha^* = cx_\alpha^*) = \alpha\,.$$

In other words, when a particular α-cut of the fuzzy solution, (x_α) yields an objective value $(z_\alpha = c^T x_\alpha)$ whose membership level for goal G, $(\mu_G(z_\alpha))$ is equal to *the same* α, then that solution x_α corresponds to Zimmermann' optimal solution, x^*.

Surprise   Recently, Neumaier suggested a way to model fuzzy right-hand side values with crisp inequality constraints ($Ax \le \tilde{b}$) based on the concept of surprise [83,94]. Neumaier defines a surprise function, $s(x \mid E)$, which corresponds to the amount of surprise a variable x produces, given a statement E. The range of s is $[0, \infty)$, with $s = 0$ corresponding to an entirely true statement E, and $s = \infty$ corresponding to an entirely false statement E.

The most plausible values are those values of x for which $s(x \mid E)$ is minimal, so Neumaier proposes that the best compromise solution for an optimization problem with fuzzy goals and constraints can be found by minimizing the sum of the surprise functions of the goals and constraints. Each fuzzy constraint,

$$(Ax)_i \le \tilde{b}_i \,,$$

is translated into a fuzzy equality constraint,

$$(Ax)_i = \tilde{\xi}_i \,,$$

where the membership function $\mu_i(\xi)$ of $\tilde{\xi}_i$ is the possibility that $\tilde{b}_i \ge \xi$. These membership functions are subsequently translated into surprise functions by

$$s_i(\xi) = (\mu_i(\xi)^{-1} - 1)^2;$$

and the contribution of all constraints are added to give the total surprise

$$\sum_i s_i(\xi) = \sum_i s_i((Ax)_i) \,.$$

Thus, the fuzzy optimization problem is

$$\min z = \sum_i s_i((Ax)_i) \tag{32}$$

subject to: $x \ge 0$.

Figure 3 illustrates a surprise function associated with a corresponding fuzzy goal. For triangular and trapezoidal numbers, the surprise function is simple, smooth, and convex, leading to a tractable non-linear programming problem.

The surprise approach is a fuzzy optimization method since the optimization is over sets of crisp values derived from fuzzy sets. In contrast to flexible programming, the constraints are not restricted such that **all** satisfy a minimal level. The salient feature is that surprise uses a dynamic penalty for falling outside distribution/membership values of one. Because the individual penalties are convex functions which become infinite as the values approach the endpoints of the support, this method lends itself to convex programming solution techniques.

It is noted that the surprise approach may be used to handle soft constraints of flexible programming since these soft constraints can be viewed as fuzzy numbers (trapezoidal fuzzy numbers when linear interpolation is used). However, if soft constraints are handled using surprise functions, the **sum** of the failure to meet the constraints is minimized rather than forcing each constraint to meet a minimal feasibility level. One could add a hard constraint to the surprise approach to attain this minimal level of feasibility. The modeler might choose to translate soft constraints to surprise functions (with perhaps a fixed minimal feasibility level) because surprise is usually more computationally efficient than Zimmermann's method of handling soft constraints (see [72]). Therefore, the surprise approach is quite flexible both in terms of semantics as well as in computational robustness.
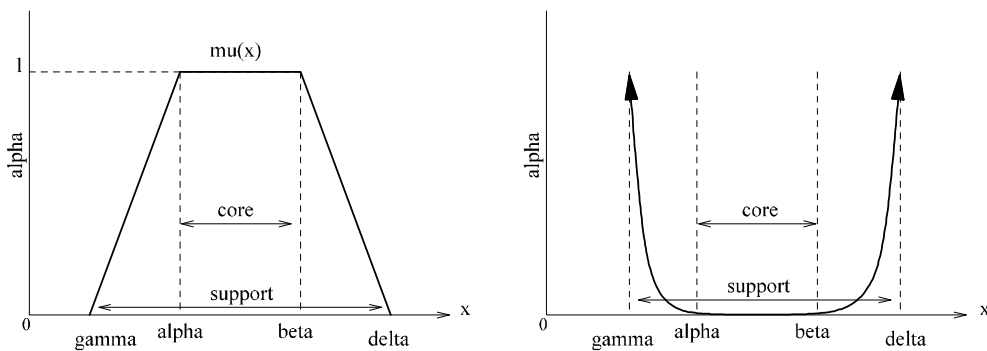
Fuzzy Optimization, Figure 3
Surprise Function

Possibilistic Programming

Recall that fuzzy imprecision arises when elements of a set (for instance, a feasible set) are members of the set to varying degrees, which are defined by the membership function of the set. Possibilistic imprecision arises when elements of a set (say, again, a feasible set) are known to exist as either full members or non-members, but whether they are members or non-members is known with a varying degree of certainty, which is defined by the possibility distribution of the set. This possibilistic uncertainty arises from a lack of information. In this section, we examine possibilistic programming formulations.

Buckley J.J. Buckley has suggested an algorithm for dealing with possibilistic cost coefficients, constraint coefficients, *and* right hand sides. Consider the possibilistic linear program

$$\min Z = \hat{c}x$$
$$\text{subject to: } \hat{A}x \geq \hat{b}, x \geq 0. \tag{33}$$

where $\hat{A} = [\tilde{a}_{ij}]$ is an $m \times n$ matrix of trapezoidal possibilistic intervals $\hat{a}_{ij} = (a_{ij\alpha}, a_{ij\beta}, a_{ij\gamma}, a_{ij\delta})$, $\hat{b} = (\tilde{b}_1, \ldots, \tilde{b}_m)^{\mathrm{T}}$ is an $m \times 1$ vector of trapezoidal fuzzy numbers $\hat{b}_i = (b_{i\alpha}, b_{i\beta}, b_{i\gamma}, b_{i\delta})$, and $\hat{c} = (\tilde{c}_1, \ldots, \tilde{c}_n)$ is a $1 \times n$ vector of trapezoidal possibilistic intervals $\tilde{c}_i = (c_{i\alpha}, c_{i\beta}, c_{i\gamma}, c_{i\delta})$. The possibilistic intervals are the possibility distributions associated with the variables, and place a restriction on the possible values the variables may assume [145]. For example, $\text{Poss}[\tilde{a}_{ij} = a] = \mu_a(\tilde{a}_{ij})$ is the possibility that $\tilde{a}_{ij}$ is equal to a. Stated another way, $\text{Poss}[\tilde{a}_{ij} = a] = x$ means that, given the current state of knowledge about the value of a_{ij}, we believe that x is the possibility that variable a_{ij} could take on value a.

Because this is a possibilistic linear programming problem, the objective function will be governed by a possibilistic distribution, $\text{Poss}[Z = z]$. Let us consider this simple example as we follow Buckley's derivation of $\text{Poss}[Z = z]$:

$$\min z = \tilde{d}x_1 + \tilde{e}x_2$$
$$\text{subject to } \tilde{f}x_1 + \tilde{g}x_2 \leq \tilde{h}$$
$$\tilde{i}x_1 + \tilde{j}x_2 \leq 0 \tag{34}$$
$$x_1, x_2 \geq 0$$

To derive the possibility function, $\text{Poss}[Z = z]$, Buckley first specifies the possibility that x satisfies the ith constraint. Let

$$\Pi(\hat{a}_i, \hat{b}_i) = \min(\mu_{a_{i1}}(\tilde{a}_{i1}), \ldots, \mu_{a_{in}}(\tilde{a}_{in}), \mu_{b_i}(\tilde{b}_i)), \tag{35}$$

which is the simultaneous distribution of $\tilde{a}_{ij}, 1 \leq j \leq n$, and $\tilde{b}_i$. In our example (34) this corresponds to

$$\Pi(\hat{f}, \hat{g}, \hat{h}) = \min(\mu_F(f), \mu_G(g), \mu_H(h)).$$

Then the possibility that $x \geq 0$ is feasible with respect to the ith constraint is:

$$\text{Poss}[x \in \mathcal{F}_i] = \sup_{a_i, b_i}(\Pi(a_i, b_i) \mid a_i x \geq b_i).$$

In our example (34) this corresponds to

$$\text{Poss}[x \in \mathcal{F}_1] = \sup_{f, g, h}(\Pi(f, g, h) \mid fx_1 + gx_2 \leq h)$$
$$\text{Poss}[x \in \mathcal{F}_2] = \sup_{i, j, k}(\Pi(i, j, k) \mid ix_1 + jx_2 \leq k). \tag{36}$$

Now the possibility that $x \geq 0$ is feasible with respect to *all* constraints is

$$[x \in \mathcal{F}] = \min_{1 \leq i \leq m}([x \in \mathcal{F}_i])$$

In our example (34),

$$[x \in \mathcal{F}] = \min([x \in \mathcal{F}_1], [x \in \mathcal{F}_2])$$

Buckley next constructs $\text{Poss}[Z = z \mid x]$, which is the conditional possibility that the objective function Z obtains a particular value z, given values for x. The joint distribution of the possibilistic cost coefficients $\tilde{c}_j$ is

$$\Pi(c) = \min(\mu_{c_1}(\tilde{c}_1), \ldots, \mu_{c_n}(\tilde{c}_n)). \tag{37}$$

In our example (34),

$$\Pi(c) = \min(\mu_D(\tilde{d}), \mu_E(\tilde{e})).$$

Therefore

$$\text{Poss}[Z = z \mid x] = \sup_c(\Pi(c) \mid cx = z). \tag{38}$$

In our example (34), to obtain the possibility that the objective function will take on a particular value z, given the solution x, (i.e. $\text{Poss}[Z = z \mid x]$), we consider all the values of d and e for which $dx_1 + ex_2 = z$. Of these, the pair with the highest simultaneous possibility values will yield an objective function value of z with the same possibility level. For example, consider the case illustrated in Fig. 4.

When $\tilde{d} = (2, 2)$ and $\tilde{e} = (3, 2)$ are symmetric triangular fuzzy numbers, $(x_1, x_2) = (1, 1)$ and $z = 7$. We consider values of d and e such that $d \times 1 + e \times 1 = 7$. We could select $d = 2$ and $e = 5$ with $\mu_D(2) = 1$ and
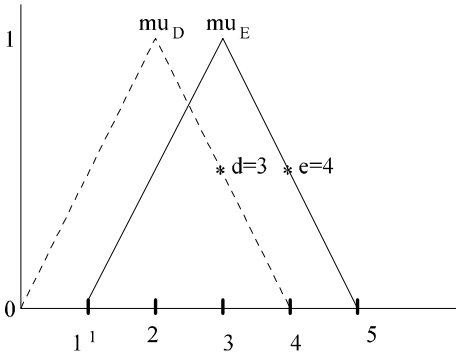
Fuzzy Optimization, Figure 4
Buckley Example

$\mu_E(5) = 0$. The joint possibility of $d = 2$ and $e = 5$ is $\min(\mu_D(2), \mu_E(5)) = 0$. The maximum joint possibility of d and e such that $d \times 1 + e \times 1 = 7$ occurs when $d = 3$ and $e = 4$. The joint possibility that $d = 3$ and $e = 4$, which is $\min(\mu_D(3), \mu_E(4)) = .5$, we set equal to the possibility that $z = 7$ given that $(x_1, x_2) = (1, 1)$. So $\text{Poss}[z = 7 \,|\, (1, 1)] = .5]$. A combination of (37) and (38) yields the possibility distribution of the objective function:

$$\text{Poss}[Z = z] = \sup_{x \geq 0}[\min(\text{Poss}[Z = z \,|\, x], \text{Poss}[x \in \mathcal{F}])]. \tag{39}$$

In our example (34), to obtain the possibility that the objective function will take on a particular value z (that is $\text{Poss}[Z = z]$, we consider $\text{Poss}[Z = z \,|\, x]$, as described above, for all positive x's, and select the x which maximizes $\text{Poss}[Z = z \,|\, x]$.

This definition of the possibility distribution motivates Buckley's solution method. Recall that because we are dealing with a possibilistic problem, the solution will be governed by a possibilistic distribution. Buckley's method depends upon a static α, chosen a priori. The decision maker defines an acceptable level of uncertainty in the objective outcome, $0 < \alpha \leq 1$. For a given α, we define the left and right end-points of the α-cut of a fuzzy number $\tilde{x}$ as $x^-(\alpha)$ and $x^+(\alpha)$, respectively. Using these, Buckley defines a new objective function:

$$\min Z(\alpha) = c^-(\alpha)x$$
$$\text{subject to: } A^+(\alpha)x \geq b^-(\alpha). \tag{40}$$

Since both x and α are variables, this is a non-linear programming problem. When α is fixed in advance, however, it becomes linear. We can use either the simplex method or an interior point method to solve for a given

a priori chosen value of α. If a maximal value for α is desired, the linear programming method must be applied iteratively.

It should be noted that this linear program is constrained upon the best-case scenario. That is, for a given α-level, each variable is multiplied by the largest possible coefficient $a_{ij}^+(\alpha)$, and is required to be greater than the the smallest possible right hand side $b_i^-(\alpha)$. We should interpret $z(\alpha)$ accordingly. If the solution to the linear program is implemented, the possibility that the objective function will attain the level $z(\alpha)$ is given by α. Stated another way, the best-case scenario is that the objective function attains a value of $z(\alpha)$, and the possibility of the best case scenario occurring is α.

Tanaka, Asai, Ichihashi In the mid 1980s, Tanaka and Asai [118] and Tanaka, Ichahashi, and Asai [121] proposed a technique for dealing with ambiguous coefficients and right hand sides based upon a possibilistic definition of "greater than zero". The reader will note that this approach bears many similarities to the flexible programming proposed by Tanaka, Okuda, and Asai a decade earlier, which was discussed in Subsect. "Tanaka, Okuda, and Asai". Indeed, the 1984 publications refer to *fuzzy* variables. This approach has subsequently been classified [44,60] as possibilistic programming because the imprecision it represents stems from a lack of information about the values of the coefficients.

Consider a programming problem with non-interactive possibilistic $\hat{A}$, $\hat{b}$, and $\hat{c}$, whose possible values are defined by fuzzy matrix $\tilde{A}$, fuzzy vector $\tilde{b}$, and fuzzy vector $\tilde{c}$, respectively:

$$\min \tilde{z} = \tilde{c}x$$
$$\text{subject to: } \tilde{A}x \leq \tilde{b} \tag{41}$$
$$x \geq 0.$$

Tanaka, Asai, and Ichihashi transform the problem in several steps. First, the objective function is viewed as a goal. As in flexible programming, the goal becomes a constraint, with the aspiration level for the objective function on the right-hand-side of the inequality. Next, a new decision variable x_0 is added. Finally, the b's are moved to the left hand side, so that the possibilistic linear programming problem is

$$\tilde{a}_i' x_i' \geq 0, \quad i = 1, 2, \ldots, m$$
$$x' \geq 0 \tag{42}$$

where $x' = (1, x^T)^T = (1, x_1, x_2, \ldots, x_n)^T$, and $\tilde{a}_i = (\tilde{b}_i, \tilde{a}_{i1}, \ldots \tilde{a}_{in})$.

Note that all the parameters, A, b, and c are now grouped together in the new constraint matrix $\tilde{A}$. Because the objective function(s) have become goals, the cost coefficients, c, are part of the constraint coefficient matrix A. Furthermore, because the right-hand side values have been moved to the left-hand side, the b's are also part of the constraint coefficient matrix. The right-hand-sides corresponding the former objective functions are the aspiration levels of the goals.

Each constraint becomes

$$\tilde{Y}_i = \tilde{a}_i x \; \geq \; 0 \,,$$

where

$$\tilde{a}_i = (\tilde{b}_i, \tilde{a}_{i1}, \dots, \tilde{a}_{in}) \,.$$

"$\tilde{Y}_i$ is almost positive", denoted by $\tilde{Y}_i \gtrsim 0$, is defined by

$$
\begin{aligned}
\tilde{Y}_i &\gtrsim 0 \\
&\Leftrightarrow \\
\mu_{Y_i}(0) &\leq 1 - h, \\
x^T \alpha_i &\geq 0.
\end{aligned}
\tag{43}
$$

The measure of the non-negativity of $\tilde{Y}_i$ is h: The greater the value of h, the stronger the meaning of "almost positive" (see Fig. 5). Actually, h is $1 - \alpha$, where α is the level of membership used by Bellman and Zadeh.

Tanaka and Asai [118] developed this theory for triangular fuzzy numbers, and Tanaka and Asai extended it to trapezoidal fuzzy numbers in [122]. Inuiguchi, Ichihashi, and Tanaka, [44] generalized the approach for *L-R* fuzzy numbers. For the sake of simplicity, our discussion here will deal with trapezoidal fuzzy numbers, denoted
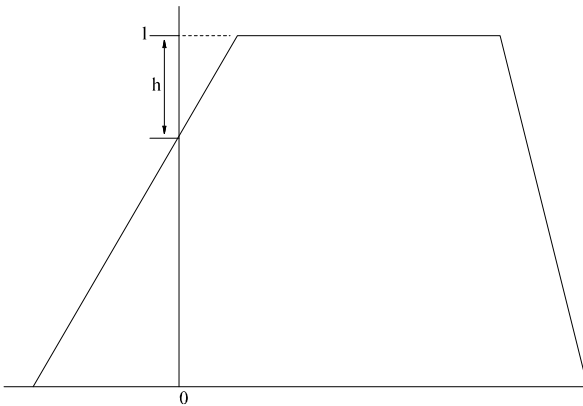
Fuzzy Optimization, Figure 5
h-Level

$\tilde{x} = (\gamma, \alpha, \beta, \delta)$, where (γ, δ) is the support of $\tilde{x}$, and (α, β) is the core of $\tilde{x}$.

Using (43), we can rewrite each constraint from (42) as

$$
\begin{aligned}
\mu_{Y_i}(0) &= 1 - \frac{\alpha_i^T x}{(\alpha_i - \gamma_i)^T x} \\
&\leq 1 - h \\
\alpha_i^T x &\geq 0 \,,
\end{aligned}
\tag{44}
$$

where $x > 0$. Then (44) reduces to

$$(\alpha_i - h(\alpha_i - \gamma_i))^T x \geq 0 \,.$$

Since we wish to find the largest h that satisfies these conditions, the linear program becomes

$$
\begin{aligned}
\max z &= h \\
\text{subject to: } (\alpha_i - h(\alpha_i - \gamma_i))^T x &\geq 0 \,, \quad \forall i \\
h &\in [0, 1] \,.
\end{aligned}
\tag{45}
$$

Since both x and h are variables, this is a non-linear programming problem. When h is fixed, it becomes linear. We can use the simplex method or an interior point algorithm to solve for a given value of h. If the decision maker wishes to maximize h, the linear programming method must be applied iteratively.

Fuzzy Max   Dubois and Prade [19] suggested that the concept of "fuzzy max" could be applied to constraints with fuzzy parameters. The "fuzzy max" concept was used to solve possibilistic linear programs with triangular possibilistic coefficients by Tanaka, Ichihashi, and Asai [121]. Ramik and Rimanek [101] applied the same technique to *L-R* fuzzy numbers. For consistency, we will discuss the fuzzy max technique with respect to trapezoidal numbers.

The fuzzy max, illustrated in Fig. 6 where $C = \max[A, B]$, is the extended maximum operator between real numbers, and defined by the extension principle (8) as

$$\mu_C(c) = \max_{\{a,b \,: \, c = max(a,b)\}} \min[\mu_{\tilde{A}}(a), \mu_{\tilde{B}}(b)] \,.$$

Using fuzzy max, we can define an inequality relation as

$$
\begin{aligned}
\tilde{A} &\geq \tilde{B} \\
&\Leftrightarrow \\
\max(\tilde{A}, \tilde{B}) &= \tilde{A} \,.
\end{aligned}
\tag{46}
$$

Applying (46) to a fuzzy inequality constraint:

$$
\begin{aligned}
f(x, \tilde{a}) &\geq g(x, \tilde{b}) \\
&\Leftrightarrow \\
\max(f(x, \tilde{a}), g(x, \tilde{b})) &= f(x, \tilde{a}) \,.
\end{aligned}
\tag{47}
$$
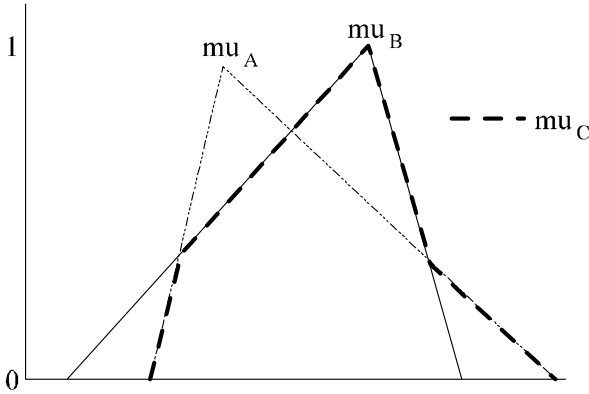
Fuzzy Optimization, Figure 6
Illustration of Fuzzy Max
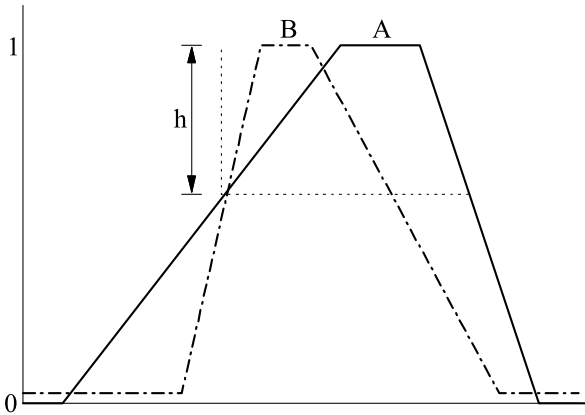
Fuzzy Optimization, Figure 7
Illustration of $\tilde{A} \geq \tilde{B}$

Observe that the inequality relation defined by (46) yields only a partial ordering. That is, it is sometimes the case that neither $\tilde{A} \geq \tilde{B}$ nor $\tilde{B} \geq \tilde{A}$ holds. To improve this, Tanaka, Ichihashi, and Asai, introduce a level h, corresponding to the decision maker's degree of optimism. They define an h-level set of $\tilde{A} \geq \tilde{B}$ as

$$
\begin{aligned}
&\tilde{A} \geq_h \tilde{B} \\
&\quad \Leftrightarrow \\
&\max[\tilde{A}]_\alpha \geq \max[\tilde{B}]_\alpha \\
&\min[\tilde{A}]_\alpha \geq \min[\tilde{B}]_\alpha \\
&\quad \alpha \in [1-h, 1] \, .
\end{aligned}
\tag{48}
$$

This definition of $\tilde{A} \geq \tilde{B}$ requires that two of Dubois' inequalities from Subsect. "Fuzzy Relations", (ii) and (iii), hold at the same time, and is illustrated in Fig. 7.

Tanaka, Ichihashi, and Asai, [118] suggest a similar treatment for a fuzzy objective function. A problem with the single objective function, Maximize $z(x, \tilde{c})$, becomes a multi-objective problem with objective functions:

$$
\text{maximize} \begin{cases} \inf(z(x, \tilde{c})_\alpha) & \alpha \in [0, 1] \\ \sup(z(x, \tilde{c})_\alpha) & \alpha \in [0, 1]. \end{cases}
\tag{49}
$$

Clearly, since α can assume an infinite number of values, (49) has an infinite number of parameters. Since (49) is not tractable, Inuiguchi, Ichihashi, and Tanaka, [44] suggest using the following approximation using a finite set of $\alpha_i \in [0, 1]$:

$$
\text{maximize} \begin{cases} \min(z(x, \tilde{c}))_{\alpha_i} & i = 1, 2, \ldots, p \\ \max(z(x, \tilde{c}))_{\alpha_i} & i = 1, 2, \ldots, p \end{cases}.
\tag{50}
$$

Jamison and Lodwick Jamison and Lodwick [50,74] develop a method for dealing with possibilistic right hand sides that is a possibilistic generalization of the recourse models in stochastic optimization. Violations of constraints are allowable, at a cost determined a priori by the decision maker.

Jamison and Lodwick choose the utility (that is, valuation) of a given interval of possible values to be its expected average (a concept defined by Yager [139].) The expected average (or EA) of a possibilistic distribution of $\tilde{a}$ is defined to be

$$
EA(\tilde{a}) = \frac{1}{2} \int_0^1 (\tilde{a}^-(\alpha) + \tilde{a}^+(\alpha)) \, d\alpha.
\tag{51}
$$

It should be noted that the expected average of a crisp value is the value itself, since $\tilde{a}^-(\alpha) = \tilde{a}^+(\alpha) = a$, $EA(a) = \frac{1}{2} \int_0^1 (a + a) dx = a \int_0^1 dx = a$.

Jamison and Lodwick start from the following possibilistic linear program:

$$
\begin{aligned}
\max z &= c^{\mathrm{T}} x \\
Ax &\leq \hat{b} \, , \\
x &\geq 0 \, .
\end{aligned}
\tag{52}
$$

By subtracting a penalty term from the objective function, they transform (52) into the following possibilistic nonlinear program:

$$
\begin{aligned}
\max z &= c^{\mathrm{T}} x + p^{\mathrm{T}} \max(0, Ax - \hat{b}) \\
x &\geq 0 \, .
\end{aligned}
\tag{53}
$$

The "max" in the penalty is taken component-wise and each $p_i < 0$ is the cost per unit violation of the right hand side of constraint i. The utility, which is the expected average, of the objective function is chosen to be minimized.

The possibilistic programming problem becomes

$$\max z = c^{\mathrm{T}} x + p E A(\max(0, Ax - \hat{b}))$$
$$x \in [0, U].$$
(54)

A closed-form objective function (for the purpose of differentiating when solving) is achieved in [72] by replacing

$$\max(0, Ax - \hat{b})$$

with

$$\frac{\sqrt{(Ax - \hat{b}) + \epsilon^2} + Ax - \hat{b}}{2}.$$

Jamison and Lodwick's method can be extended, [50], to account for possibilistic values for A, b, c, and even the penalty coefficients p with the following formulation:

$$EA\tilde{f}(x) = \frac{1}{2} \int_0^1 \left\{ \hat{c}^-(\alpha)^{\mathrm{T}} x + \hat{c}^+(\alpha)^{\mathrm{T}} x \right.$$
$$- \left[\hat{p}^+(\alpha) \max(0, \hat{A}^-(\alpha) x - \hat{b}^+(\alpha) \right]$$
$$\left. - \left[\hat{p}^-(\alpha) \max(0, \hat{A}^+(\alpha) x - \hat{b}^-(\alpha) \right] \right\} d\alpha. \quad (55)$$

This approach differs significantly from the others we've examined in several regards. First, many of the approaches we've seen have incorporated the objective function(s) as goals into the constraints in the Bellman and Zadeh tradition. Jamison and Lodwick, on the other hand, incorporate the constraints into the objective function. Bellman and Zadeh create a symmetry between constraints and objective, while Jamison and Lodwick temper the objective with the constraints. A second distinction of the expected average approach is the nature of the solution. The other formulations we have examined to this point have produced either (1) a crisp solution for a particular value of α, (namely, the maximal value of α), or, (2) a fuzzy/possibilistic solution which encompasses all possible α values. The Jamison and Lodwick approach provides a crisp solution via the expected average utility which encompasses all alpha values. This may be a desirable quality to the decision maker who wants to account for all possibility levels and still reach a crisp solution.

Luhandjula Luhanjula's [84] formulation of the possibilistic mathematical program depends upon his concept of "more possible" values. He first defines a possibility distribution Π_X with respect to constraint F as

$$\Pi_X = \mu_F(u),$$

where $\mu_F(u)$ is the degree to which the constraint F is satisfied when u is the value assigned to the solution X.

Then the set of more possible values for X, denoted by $V_p(X)$, is given by

$$V_p(X) = \Pi_X^{-1}(\max_u \Pi_X(u)).$$

In other words, $V_p(X)$ contains elements of U which are most compatible with the restrictions defined by Π_X. It follows from intuition, and from Luhanjula's formal proof [84], that when Π_X is convex, $V_p(X)$ is a real-valued interval, and when Π_X is strongly convex, $V_p(X)$ is a single real number.

Luhandjula considers the mathematical program

$$\max \tilde{z} = \hat{c} x$$
$$\text{subject to: } \hat{A}_i \le \hat{b}_i, \quad (56)$$
$$x \ge 0.$$

By replacing the possibilistic numbers $\hat{c}$, $\hat{A}_i$, and $\hat{b}_i$ with their more possible values, $V_p(\hat{c})$, $V_p(\hat{A}_i)$, and $V_p(\hat{b}_i)$, respectively, Luhandjula arrives at a deterministic equivalent to Eq. (56):

$$\max z = kx$$
$$\text{subject to: } k_i \in V_p(\hat{c}_i)$$
$$\sum_i t_i x_i \le s_i$$
$$t_i \in V_p(\hat{a}_{ij}) \quad (57)$$
$$s_i \in V_p(\hat{b}_i)$$
$$x \ge 0.$$

This formulation varies significantly from the other approaches considered thus far. The possibility of each possibilistic component is maximized individually. Other formulations have required that each possibilistic component $\tilde{c}_j$, $\tilde{A}_{ij}$, and $\tilde{b}_i$ achieve the same possibility level defined by α. This formulation also has a distinct disadvantage over the others we've considered, since to date there is no proposed computational method for determining the "more possible" values, V_p, so there is no way to solve the deterministic MP.

Programming with Fuzzy and Possibilistic Components

Sometimes the values of an optimization problem's components are ambiguous *and* the decision-makers are vague (or flexible) regarding feasibility requirements. This sections explores a couple of approaches for dealing with such fuzzy/possibilistic problems.

One type of mixed programming problem that arises (see [20,91]) is a mathematical program with possibilistic constraint coefficients $\hat{a}_{ij}$ whose possible values are defined by fuzzy numbers of the form $\tilde{a}_{ij}$:

$$\max cx$$
$$\text{subject to: } \hat{a}'_i x' \subseteq \tilde{b}_i \tag{58}$$
$$x' = (1, x^t)t \geq 0.$$

Zadeh [142] defines the set-inclusion relation $\tilde{M} \subseteq \tilde{N}$ as $\mu_{\tilde{M}}(r) \leq \mu_{\tilde{N}}(r) \forall r \in R$. Recall that Dubois [18] interprets the set-inclusive constraint $\tilde{a}'_i x' \subseteq \tilde{b}_i$ as a fuzzy extension of the crisp equality constraint. Mixed programming, however, interprets the set-inclusive constraint to mean that the region in which $\tilde{a}'_i x'$ can possibly occur is restricted to $\tilde{b}_i$, a region which is tolerable to the decision maker. Therefore, the left side of (58) is possibilistic, and the right side is fuzzy.

Negoita [91] defines the fuzzy right hand side as follows:

$$\tilde{b}_i = \{r \mid r \geq b_i\}. \tag{59}$$

As a result, we can interpret $\tilde{a}'_i x' \subseteq \tilde{b}_i$ as an extension of an inequality constraint. The set-inclusive constraint (58) is reduced to

$$a_i^+(\alpha)x \leq b_i^+(\alpha)$$
$$a_i^-(\alpha)x \geq b_i^-(\alpha) \tag{60}$$
$$\text{for all } \alpha \in (0, 1].$$

If we abide by Negoita's definition of $\tilde{b}$ (59), $b_i^+ = \infty$ for all values of α, so we can drop the first constraint in (60). Nonetheless, we still have an infinitely (continuum) constrained program, with two constraints for each value of $\alpha \in (0, 1]$. Inuiguchi observes [44] that if the left-hand side of the membership functions for $a_{i0}, a_{i1}, \ldots a_{in}, b_i$ are identical for all i, and the right-hand side of the membership functions for $a_{i0}, a_{i1}, \ldots a_{in}, b_i$ are identical for all i, the constraints are reduced to the finite set,

$$a_{i,\alpha} \geq b_{i,\alpha}$$
$$a_{i,\gamma} \geq b_{i,\gamma}$$
$$a_{i,\beta} \leq b_{i,\beta} \tag{61}$$
$$a_{i,\delta} \leq b_{i,\delta}.$$

As per our usual notation, (γ, δ) is the support of the fuzzy number, and (α, β) is its core. In application, constraint formulation (61) has limited utility because of the narrowly defined sets of memberships functions it admits. For example, if $a_{i0}, a_{i1}, \ldots a_{in}, b_i$ are defined by trapezoidal fuzzy numbers, they must all have the same spread,

and therefore the same slope, on the right-hand side; and they must all have the same spread, and therefore the same slope, on the left-hand side if (61) is to be implemented. Recall that in this kind of mixed programming, the a_{ij}'s are possibilistic, reflecting a lack of information about their values, and the b_i are fuzzy, reflecting the decision maker's degree of satisfaction with their possible values. It is possible that n possibilistic components and 1 fuzzy component will share identically-shaped distribution, but it is not something likely to happen with great frequency.

Delgado, Verdegay and Vila Delgado, Verdegay, and Vila [11] propose the following formulation for dealing with ambiguity in the constraint coefficients and right-hand sides, as well as vagueness in the inequality relationship:

$$\max cx$$
$$\text{subject to: } \hat{A}x \stackrel{\sim}{\leq} \hat{b} \tag{62}$$
$$x \geq 0.$$

In addition to (62), membership functions $\mu_{a_{ij}}$ are defined for the possible values of each possibilistic element of $\hat{A}$, membership functions μ_{b_i} are defined for the possible values of each possibilistic element of $\hat{b}$, and membership function μ_i gives the degree to which the fuzzy constraint i is satisfied. Stated another way, μ_i is the membership function of the fuzzy inequality. The uncertainty in the $\tilde{a}_{ij}$ and the $\tilde{b}_i$ is due to ambiguity concerning the actual value of the parameter, while the uncertainty in the $\stackrel{\sim}{\leq}$ is due to the decision maker's flexibility regarding the necessity of satisfying the constraints in full.

Delgado, Verdegay, and Vila do not define the fuzzy inequality, but leave that specification to decision maker. Any ranking index that preserves ranking of fuzzy numbers when multiplied by a positive scalar is allowed. For instance, one could select any of Dubois' four inequalities from Subsect. "Fuzzy Relations". Once the ranking index is selected, the problem is solved parametrically, as in Verdegay's earlier work [130] (see Subsect. "Verdegay").

To illustrate this approach, let us choose Dubois' pessimistic inequality (i), which interprets $\tilde{A} \geq \tilde{b}$ to mean $\forall x \in A, \forall y \in B, x \geq y$. This is equivalent to $a^+ \geq b^-$. Then (62) becomes

$$\max cx$$
$$\text{subject to: } a_{ij}^+(\alpha)x \leq b_i^-(\alpha) \tag{63}$$
$$x \geq 0.$$

Fuzzy/Robust Programming The approach covered so far in this section, has the same α cut define the level of am-

biguity in the coefficients *and* the level at which the decision-maker's requirements are satisfied. These α, however, mean very different things. The fuzzy α represents the level at which the decision-maker's requirements are satisfied. The possibilistic α, on the other hand, represents the likelihood that the parameters will take on values which will attain that level. The solution is interpreted to mean that for any value $\alpha \in (0, 1]$ there is a possibility α of obtaining a solution that satisfies the decision maker to degree α. Using the same α value for both the possibilistic and fuzzy components of the problem is convenient, but does not necessarily provide a meaningful model of reality.

A recent model [126] based on Markowitz's mean-variance approach to portfolio optimization (see [89,117]) differentiates between the fuzzy α and the possibilistic α. Markowitz introduced an efficient combination, which has the minimum risk for a return greater than or equal to a given level; or one which has the maximum return for a risk less than or equal to a given level. The decision maker can move among these efficient combinations, or along the efficient frontier, according to her/his degree of risk aversion.

Similarly, in mixed possibilistic and fuzzy programming, one might wish to allow a trade-off between the potential reward of the outcome and the reliability of the outcome, with the weights of the two competing objectives determined by the decision maker's risk aversion. The desire is to obtain an objective function like the following:

$$\max[\text{reward} + (\text{risk aversion}) \times (\text{reliability})] . \tag{64}$$

The reward variable is the α-level associated with the fuzzy constraints and goal(s). It tells the decision maker how satisfactory the solution is. The reliability variable is the α-level associated with the possibilistic parameters. It tells the decision maker how likely it is that the solution will actually be satisfactory. To avoid confusion, let us refer to the fuzzy constraint membership parameter as α and the possibilistic parameter membership level as β.

In addition, let $\mu \in [0, 1]$ be an indicator of the decision maker's valuation of reward and risk-avoidance, with 0 indicating that the decision maker cares exclusively about the reward, and 1 indicating the only risk avoidance is important. Using this notation, the desired objective is

$$\max(1 - \mu)\alpha + \mu\beta . \tag{65}$$

Suppose we begin with the mixed problem:

$$\max \hat{c}^{\mathrm{T}} x$$
$$\text{subject to } \hat{A}x \lesssim b \tag{66}$$
$$x \geq 0 .$$

Incorporating fuzziness from soft constraints in the tradition of Zimmermann and incorporating a pessimistic view of possibility results in the following formulation:

$$\max \quad (1 - \mu)\alpha + \mu\beta \tag{67}$$

$$\text{subject to} \quad \alpha \leq -\frac{g}{d_0} + \sum_j \frac{u_j}{d_0} x_j + \sum_j \frac{u_j - w_j}{d_0} x_j \beta$$

$$\alpha \leq \frac{b_i}{d_i} - \sum_j \frac{v_{ij}}{d_i} x_j - \sum_j \frac{z_{ij} - v_{ij}}{d_i} x_j \beta$$

$$x \geq 0$$

$$\alpha, \beta \in [0, 1] .$$
$$\tag{68}$$

The last terms in each of the constraints contain βx, so the system is non-linear. It can be fairly easily reformulated as an optimization program with linear objective function and quadratic constraints, but the feasible set is non-convex, so finding to solution to this mathematical programming problem is very difficult.

Possibilistic, Interval, Cloud, and Probabilistic Optimization Utilizing IVPM

This section is taken from [79,80] and begins by defining what is meant by an IVPM. This generalization of a probability measure includes probability measures, possibility/necessity measures, intervals, and clouds (see [95]) which will allow a mixture of uncertainty within one constraint (in)equality. The previous mixed methods was restricted to a single type of uncertainty for any particular (in)equality and are unable to handle cases in which a mixture of fuzzy and possibilistic parameter occurs in the same constraint (in)equality.

The IVPM set function may be thought of as a method for giving a partial representation for an unknown probability measure. Throughout, arithmetic operations involving set functions are in terms of interval arithmetic [90] and the set of all intervals contained in $[0, 1]$ is denoted, $\text{Int}_{[0,1]} \equiv \{[a, b] \mid 0 \leq a \leq b \leq 1\}$. Moreover, S is used to denote the universal set and a set of subsets of the universal set is denoted as $\mathcal{A} \subseteq S$. In particular $\mathcal{A}$ is a set of subset on which a structure has been imposed on it as will be seen and a generic set of the structure $\mathcal{A}$ is denoted by A.

Definition 18 (Weichselberger [135]) Given measurable space $(S, \mathcal{A})$, an interval valued function $i_m \colon A \subseteq \mathcal{A} \to \text{Int}_{[0,1]}$ is called an **R-probability** if:

(a) $i_m(A) = [i_m^-(A), i_m^+(A)] \subseteq [0, 1]$ with $i_m^-(A) \leq i_m^+(A)$,

(b) $\exists$ a probability measure Pr on $\mathcal{A}$ such that $\forall A \in \mathcal{A}$, $\mathrm{Pr}(A) \in i_m(A)$. By an **R-probability field** we mean the triple $(S, \mathcal{A}, i_m)$.

Definition 19 (Weichselberger [135]) Given an R-probability field $\mathcal{R} = (S, \mathcal{A}, i_m)$ the set

$$\mathcal{M}(\mathcal{R}) = \{\mathrm{Pr} \mid \mathrm{Pr} \text{ is a probability measure on } \mathcal{A}$$
$$\text{such that } \forall A \in \mathcal{A}, \mathrm{Pr}(A) \in i_m(A)\}$$

is called the **structure** of $\mathcal{R}$.

Definition 20 (Weichselberger [135]) An R-probability field $\mathcal{R} = (S, \mathcal{A}, i_m)$ is called an **F-probability field** if $\forall A \in \mathcal{A}$:

(a) $i_m^+(A) = \sup\{\mathrm{Pr}(A) \mid \mathrm{Pr} \in \mathcal{M}(\mathcal{R})\}$,
(b) $i_m^-(A) = \inf\{\mathrm{Pr}(A) \mid \mathrm{Pr} \in \mathcal{M}(\mathcal{R})\}$.

It is interesting to note that given a measurable space $(S, \mathcal{A})$ and a set of probability measures P, then defining $i_m^+(A) = \sup\{\mathrm{Pr}(A) \mid \mathrm{Pr} \in P\}$ and $i_m^-(A) = \inf\{\mathrm{Pr}(A) \mid \mathrm{Pr} \in P\}$ gives an F-probability and that P is a subset of the structure.

The following examples show how intervals, possibility distributions, clouds and (of course) probability measures can define R-probability fields on $\mathcal{B}$, the Borel sets on the real line.

Example 21 (An interval defines an F-probability field) Let $I = [a, b]$ be a non-empty interval on the real line. On the Borel sets define

$$i_m^+(A) = \begin{cases} 1 & \text{if } I \cap A \neq \emptyset \\ 0 & \text{otherwise} \end{cases}$$

and

$$i_m^-(A) = \begin{cases} 1 & \text{if } I \subseteq A \\ 0 & \text{otherwise} \end{cases}$$

then

$$i_m(A) = [i_m^-(A), i_m^+(A)]$$

defines an F-probability field $\mathcal{R} = (R, \mathcal{B}, i_m)$. To see this, simply let P be the set of all probability measures on $\mathcal{B}$ such that $\mathrm{Pr}(I) = 1$.

This example also illustrates that any set A, not just an interval I, can be used to define an F-probability field.

Example 22 (A probability measure is an F-probability field) Let Pr be a probability measure over $(S, \mathcal{A})$. Define

$$i_m(A) = [\mathrm{Pr}(A), \mathrm{Pr}(A)].$$

This definition of a probability as an IVPM is equivalent to having total knowledge about a probability distribution over S. The concept of a cloud was introduced by Neumaier in [95] as follows:

Definition 23 A **cloud** over set S is a mapping c such that:

1) $\forall s \in S$, $c(s) = [\underline{n}(s), \bar{p}(s)]$ with $0 \leq \underline{n}(s) \leq \bar{p}(s) \leq 1$
2) $(0, 1) \subseteq \cup_{s \in S} c(s) \subseteq [0, 1]$ In addition, random variable X taking values in S is said to belong to cloud c (written $X \in c$) iff
3) $\forall \alpha \in [0, 1]$, $\mathrm{Pr}(\underline{n}(X) \geq \alpha) \leq 1 - \alpha \leq \mathrm{Pr}(\bar{p}(X) > \alpha)$

Property 3) above defines when a random variable belongs to a cloud. That any cloud contains a random variable X is proved in section 5 of [96]. This is a significant result as will be seen, since among other things, it means that clouds can be used to define IVPMs. Clouds are closely related to possibility theory.

It is shown in [51] that possibility distributions can be constructed which satisfy the following consistency definition.

Definition 24 ([51]) Let $p: S \to [0, 1]$ be a possibility distribution function with associated possibility measure Pos and necessity measure Nec. Then p is said to be **consistent** with random variable X if $\forall$ measurable sets A, $\mathrm{Nec}(A) \leq \mathrm{Pr}(X \in A) \leq \mathrm{Pos}(A)$.

Possibility distributions constructed in a consistent manner are able to bound (unknown) probabilities of interest. The reason this is significant is twofold. Firstly, possibility and necessity distributions are easier to construct since the axioms they satisfy are more general. Secondly, the algebra on possibility and necessity pairs are much simpler since they are min/max algebras akin to the min/max algebra of interval arithmetic (see, for example, [76]). In particular, they avoid convolutions which are requisite for probabilistic arithmetic. The concept of a cloud can be stated in terms of certain pairs of consistent possibility distributions as shown by the following proposition (which means that clouds may be considered as pairs of consistent possibilities – possibility and necessity pairs, see [51]).

Proposition 25 *Let $\bar{p}, \underline{p}$ be a pair of regular possibility distribution functions over set S such that $\forall s \in S \bar{p}(s) + \underline{p}(s) \geq 1$. Then the mapping $c(s) = [\underline{n}(s), \bar{p}(s)]$ where $\underline{n}(s) = 1 - \underline{p}(s)$ (i.e. the dual necessity distribution function) is a cloud. In addition, if X is a random variable taking values in S and the possibility measures associated with $\bar{p}, \underline{p}$ are consistent with X then X belongs to cloud c. Conversely, every cloud defines such a pair of possibility distribution functions and their associated possibility measures are consistent with every random variable belonging to c.*

Proof (see [52,78,79])   □

Example 26 (A cloud defines an R-probability field) Let c be a cloud over the real line. Let $\text{Pos}^1, \text{Nec}^1, \text{Pos}^2, \text{Nec}^2$ be the possibility measures and their dual necessity measures relating to $\bar{p}(s)$ and $\underline{p}(s)$ (where $\bar{p}$ and $\underline{p}$ are as in Proposition 18). Define

$$i_m(A) = \big[\max\{\text{Nec}^1(A), \text{Nec}^2(A)\},$$
$$\min\{\text{Pos}^1(A), \text{Pos}^2(A)\}\big].$$

Neumaier [96] proved that every cloud contains a random variable X. Since consistency requires that $\Pr(X \in A) \in i_m(A)$, the result that every cloud contains a random variable X shows consistency. Thus every cloud defines an R-probability field because the inf and sup of the probabilities are bounded by the lower and upper bounds of $i_m(A)$.

Example 27 (A possibility distribution defines an R-probability field) Let $p: S \to [0,1]$ be a possibility distribution function and let Pos be the associated possibility measure and Nec the dual necessity measure. Define $i_m(A) = [\text{Nec}(A), \text{Pos}(A)]$. Defining a second possibility distribution, $\underline{p}(x) = 1 \forall x$ means that the pair $p, \underline{p}$ define a cloud for which $i_m(A)$ defines the R-probability. Since a cloud defines an R-probability field, this means that this possibility in turn generates a R-probability.

Note that the above example means that every pair of possibility $p(x)$ and necessity $n(x)$ such that $n(x) \leq p(x)$ has an associated F-probability field. This is because such a pair defines a cloud and in every cloud there exists a probability distribution. The F-probability field can then be constructed from a *inf/sup* over all such enclosed probabilities that are less than or equal to the bounding necessity/possibility distributions.

The application of these concepts to mixed fuzzy and possibilistic optimization is as follows. Suppose the optimization problem is to maximize $f(\vec{x}, \vec{a})$ subject to $g(\vec{x}, \vec{b}) = 0$ (where $\vec{a}$ and $\vec{b}$ are parameters). Assume $\vec{a}$ and $\vec{b}$ are vectors of independent uncertain parameters, each with an associated IVPM. Assume the constraint may be violated at a cost $\vec{p} > 0$ so that the problem becomes one to maximize

$$h(\vec{x}, \vec{a}, \vec{b}) = f(\vec{x}, \vec{a}) - \vec{p}|g(\vec{x}, \vec{b})|.$$

Given the independence assumption, form an IVPM for the product space $i_{\vec{a} \times \vec{b}}$ for the joint distribution (see the example below and [52,78,79]) and calculate the interval-valued expected value (see [48,140]) with respect to this IVPM. The interval-valued expected value is denoted

(there is a lower expected value and an upper expected value)

$$\int_R h(\vec{x}, \vec{a}, \vec{b}) \mathrm{d}i_{\vec{a} \times \vec{b}}. \tag{69}$$

To optimize (69) requires an ordering of intervals, a *valuation function* denoted by $v: \text{Int}_R \to \mathbb{R}^n$. One such ordering is the midpoint of the interval on the principle that in the absence of additional data, the midpoint is the best estimate for the true value so that for this function (midpoint and width), $v: \text{Int}_R \to \mathbb{R}^2$. Thus, for $I = [a, b]$, this particular valuation function is $v(I) = ((a+b)/2, b-a)$. Next a *utility function* of a vector in $\mathbb{R}^n$ is required which is denote by $u: \mathbb{R}^n \to \mathbb{R}$. A utility function operating on the midpoint and width valuation function is $u: \mathbb{R}^2 \to \mathbb{R}$ and particular utility is a weighted sum of the midpoint and width $u(c, d) = \alpha c + \beta d$. Using the valuation and utility functions, the optimization problem is:

$$\max_x u\left(v\left(\int_R h(x, a, b)\mathrm{d}i_{a \times b}\right)\right). \tag{70}$$

Thus there are four steps to obtaining a real-valued optimization problem from an IVPM problem. The first step is to obtain interval probability $h(x, a, b)$. The second step is to obtain the IVPM expected value $\int_R h(\vec{x}, \vec{a}, \vec{b})\mathrm{d}i_{\vec{a} \times \vec{b}}$. The third step is to obtain the vector value of $v: \text{Int}_R \to \mathbb{R}^n$. The fourth step is to obtain the utility function value of a vector $u: \mathbb{R}^n \to \mathbb{R}$.

Example 28 Consider the problem

$$\max f(x, a) = 8x_1 + 7x_2$$
$$\text{subject to:}$$
$$g_1(x, b) = 3x_1 + [1, 3]x_2 - 4 = 0$$
$$g_2(x, b) = \tilde{2}x_1 + 5x_2 - 1 = 0$$
$$\vec{x} \in [0, 2]$$

where $\tilde{2} = 1/2/3$, that is, $\tilde{2}$ is a triangular possibilistic number with support [1,3] and modal value at 2. For $\vec{p} = (1, 1)^{\mathrm{T}}$,

$$h(x, a, b) = 5x_1 - \tilde{2}x_1 + [3, 5]x_2 - 6$$

so that

$$\int_R h(x, a, b)\mathrm{d}i_{a \times b}$$
$$= 5x_1 + \left[\int_0^1 (\alpha - 3)\,\mathrm{d}\alpha, \int_0^1 (-1 - \alpha)\,\mathrm{d}\alpha\right]x_1$$
$$+ [3, 5]x_2 - 6$$
$$= 5x_1 + \left[-\frac{5}{2}, -\frac{3}{2}\right]x_1 + [3, 5]x_2 - 5.$$

Since the constant -5 will not affect the optimization, it will be removed (then added at the end), so that

$$v\left(\int_R h\,(x,a,b)\,\mathrm{d}i_{a\times b}\right) = v\left(\left[\frac{5}{2},\frac{7}{2}\right]x_1 + [3,5]x_2\right)$$
$$= (3,1)x_1 + (4,2)x_2\,.$$

Let $u(\vec{y}) = \sum_{i=1}^{n} y_i$ which for the context of this problem yields

$$\max_x z = u\left(v\left(\int_R h\,(x,a,b)\,\mathrm{d}i_{a\times b}\right)\right)$$
$$= \max_{\vec{x}\in[0,2]} (4x_1 + 6x_2) - 5$$
$$= 20 - 5 = 15$$
$$x_1^* = 2,\ x_2^* = 2\,.$$

Example 29 (see [124]) Consider

$$\max z = \hat{2}x_1 - \bar{0}x_2 + [3,5]x_3$$
$$\hat{4}x_1 + [1,5]x_2 - 2x_3 - [0,2] = 0$$
$$6x_1 - \bar{2}x_2 + 9x_3 - 9 = 0$$
$$-2x_1 - [1,4]x_2 - \hat{8}x_3 + \bar{5} = 0$$
$$0 \le x_1 \le 3, 1 \le x_2, x_3 \le 2\,.$$

Here the $\bar{0}$, $\bar{2}$, and $\bar{5}$ are probability distributions. Note the mixture of three uncertainty types in the third constraint equation. Using the same approach as in the previous example, the optimal values are:

$$z^* = 3.9179$$
$$x_1^* = 0,\quad x_2^* = 0.4355,\quad x_3^* = 1.0121\,.$$

In [124] it is shown that these mixed problems arising from linear programs remain linear programs. Thus, the complexity of mixed problems is equivalent to that of linear programming.

Future Directions

The applications of fuzzy optimization seems to be headed toward "industrial strength" problems. Increasingly, each year there are a greater number of applications that appear. Given that greater attention is being given to the semantics of fuzzy optimization and as fuzzy optimization becomes increasingly used in applications, associated algorithms that are more sophisticated, robust, and efficient will need to be developed to handle these more complex problems. It would be interesting to develop modeling languages like GAMS [6], MODLER [34], or AMPL [30], that support fuzzy data structures. From the theoretical side, the flexibility that fuzzy optimization has with working with uncertainty data that is fuzzy, flexible, and/or possibilistic (or a mixture of these via IVPM), means that fuzzy optimization is able to provide an ample approach to optimization under uncertainty. Further research into the development of more robust methods that use fuzzy Banach spaces would certainty provide a deeper theoretical foundation to fuzzy optimization. Fuzzy optimization that utilize fuzzy Banach spaces have the advantage that the problem remains fuzzy throughout and only when one needs to make a decision or implement the solution does one map the solution to a real number (defuzzify). The methods that map fuzzy optimization problems to their real number equivalent defuzzify first and then optimize. Fuzzy optimization problems that optimize in fuzzy Banach spaces keep the solution fuzzy and defuzzify as a last step.

Continued development of clear input and output semantics of fuzzy optimization will greatly aid fuzzy optimization's applicability and relevance. When fuzzy optimization is used in, for example, an assembly-line scheduling problem and one's solution is a fuzzy three, how does one convey this solution to the assembly-line manager? Lastly, continued research into handling dependencies in an efficient way would amplify the usefulness and applicability of fuzzy optimization.

Bibliography

1. Asai K, Tanaka H (1973) On the fuzzy – mathematical programming, Identification and System Estimation Proceedings of the 3rd IFAC Symposium, The Hague, 12–15 June 1973, pp 1050–1051
2. Audin J-P, Frankkowska H (1990) Set-Valued Analysis. Birkhäuser, Boston
3. Baudrit C, Dubois D, Fargier H (2005) Propagation of uncertainty involving imprecision and randomness. ISIPTA 2005, Pittsburgh, pp 31–40
4. Bector CR, Chandra S (2005) Fuzzy Mathematical Programming and Fuzzy Matrix Games. Springer, Berlin
5. Bellman RE, Zadeh LA (1970) Decision-Making in a Fuzzy Environment. Manag Sci B 17:141–164
6. Brooke A, Kendrick D, Meeraus A (2002) GAMS: A User's Guide. Scientific Press, San Francisco (the latest updates can be obtained from http://www.gams.com/)
7. Buckley JJ (1988) Possibility and necessity in optimization. Fuzzy Sets Syst 25(1):1–13
8. Buckley JJ (1988) Possibilistic linear programming with triangular fuzzy numbers. Fuzzy Sets Syst 26(1):135–138
9. Buckley JJ (1989) Solving possibilistic linear programming problems. Fuzzy Sets Syst 31(3):329–341
10. Buckley JJ (1989) A generalized extension principle. Fuzzy Sets Syst 33:241–242
11. Delgado M, Verdegay JL, Vila MA (1989) A general model for fuzzy linear programming. Fuzzy Sets Syst 29:21–29
12. Delgado M, Kacprzyk J, Verdegay J-L, Vila MA (eds) (1994) Fuzzy Optimization: Recent Advances. Physica, Heidelberg

13. Demster AP (1967) Upper and lower probabilities induced by multivalued mapping. Ann Math Stat 38:325–339

14. Dempster MAH (1969) Distributions in interval and linear programming. In: Hansen ER (ed) Topics in Interval Analysis. Oxford Press, Oxford, pp 107–127

15. Diamond P (1991) Congruence classes of fuzzy sets form a Banach space. J Math Anal Appl 162:144–151

16. Diamond P, Kloeden P (1994) Metric Spaces of Fuzzy Sets. World Scientific, Singapore

17. Diamond P, Kloeden P (1994) Robust Kuhn–Tucker conditions and optimization under imprecision. In: Delgado M, Kacprzyk J, Verdegay J-L, Vila MA (eds) Fuzzy Optimization: Recent Advances. Physica, Heidelberg, pp 61–66

18. Dubois D (1987) Linear programming with fuzzy data. In: Bezdek JC (ed) Analysis of Fuzzy Information, vol III: Applications in Engineering and Science. CRC Press, Boca Raton, pp 241–263

19. Dubois D, Prade H (1980) Fuzzy Sets and Systems: Theory and Applications. Academic Press, New York

20. Dubois D, Prade H (1980) Systems of linear fuzzy constraints. Fuzzy Sets Syst 3:37–48

21. Dubois D, Prade H (1981) Additions of interactive fuzzy numbers. IEEE Trans Autom Control 26(4):926–936

22. Dubois D, Prade H (1983) Ranking fuzzy numbers in the setting of possibility theory. Inf Sci 30:183–224

23. Dubois D, Prade H (1986) New results about properties and semantics of fuzzy set-theoretical operators. In: Wang PP, Chang SK (eds) Fuzzy Sets. Plenum Press, New York, pp 59–75

24. Dubois D, Prade H (1987) Fuzzy numbers: An overview. Tech. Rep. no 219, (LSI, Univ. Paul Sabatier, Toulouse, France). Mathematics and Logic. In: James Bezdek C (ed) Analysis of Fuzzy Information, vol 1, chap 1. CRC Press, Boca Raton, pp 3–39

25. Dubois D, Prade H (1988) Possibility Theory an Approach to Computerized Processing of Uncertainty. Plenum Press, New York

26. Dubois D, Prade H (2005) Fuzzy elements in a fuzzy set. Proceedings of the 11th International Fuzzy System Association World Congress, IFSA 2005, Beijing, July 2005, pp 55–60

27. Dubois D, Moral S, Prade H (1997) Semantics for possibility theory based on likelihoods. J Math Anal Appl 205:359–380

28. Ertuğrul I, Tuş A (2007) Interactive fuzzy linear programming and an application sample at a textile firm. Fuzzy Optim Decis Making 6:29–49

29. Fortin J, Dubois D, Fargier H (2008) Gradual numbers and their application to fuzzy interval analysis. IEEE Trans Fuzzy Syst 16:2, pp 388–402

30. Fourer R, Gay DM, Kerninghan BW (1993) AMPL: A Modeling Language for Mathematical Programming. Scientific Press, San Francisco, CA (the latest updates can be obtained from http://www.ampl.com/)

31. Fullér R, Keresztfalvi T (1990) On generalization of Nguyen's theorem. Fuzzy Sets Syst 41:371–374

32. Ghassan K (1982) New utilization of fuzzy optimization method. In: Gupta MM, Sanchez E (eds) Fuzzy Information and Decision Processes. North-Holland, Netherlands, pp 239–246

33. Gladish B, Parra M, Terol A, Uría M (2005) Management of surgical waiting lists through a possibilistic linear multiobjective programming problem. Appl Math Comput 167:477–495

34. Greenberg H (1992) MODLER: Modeling by Object-Driven Linear Elemental Relations. Ann Oper Res 38:239–280

35. Hanss M (2005) Applied Fuzzy Arithmetic. Springer, Berlin

36. Hsu H, Wang W (2001) Possibilistic programming in production planning of assemble-to-order environments. Fuzzy Sets Syst 119:59–70

37. Inuiguchi M (1992) Stochastic Programming Problems Versus Fuzzy Mathematical Programming Problems. Jpn J Fuzzy Theory Syst 4(1):97–109

38. Inuiguchi M (1997) Fuzzy linear programming: what, why and how? Tatra Mt Math Publ 13:123–167

39. Inuiguchi M (2007) Necessity measure optimization in linear programming problems with fuzzy polytopes. Fuzzy Sets Syst 158:1882–1891

40. Inuiguchi M (2007) On possibility/fuzzy optimization. In: Melin P, Castillo O, Aguilar LT, Kacprzyk J, Pedrycz W (eds) Foundations of Fuzzy Logic and Soft Computing: 12th International Fuzzy System Association World Congress, IFSA 2007, Cancun, June 2007, Proceedings. Springer, Berlin, pp 351–360

41. Inuiguchi M, Ramik J (2000) Possibilistic linear programming: A brief review of fuzzy mathematical programming and a comparison with stochastic programming in portfolio selection problem. Fuzzy Sets Syst 111:97–110

42. Inuiguchi M, Sakawa M (1997) An achievement rate approach to linear programming problems with an interval objective function. J Oper Res Soc 48:25–33

43. Inuiguchi M, Tanino T (2004) Fuzzy linear programming with interactive uncertain parameters. Reliab Comput 10(5):512–527

44. Inuiguchi M, Ichihashi H, Tanaka H (1990) Fuzzy programming: A survey of recent developments. In: Slowinski R, Teghem J (eds) Stochastic versus Fuzzy Approaches to Multiobjective Mathematical Programming Under Uncertainty. Kluwer, Netherlands, pp 45–68

45. Inuiguchi M, Ichihashi H, Kume Y (1992) Relationships Between Modality Constrained Programming Problems and Various Fuzzy Mathematical Programming Problems. Fuzzy Sets Syst 49:243–259

46. Inuiguchi M, Ichihashi H, Tanaka H (1992) Fuzzy Programming: A Survey of Recent Developments. In: Slowinski R, Teghem J (eds) Stochastic versus Fuzzy Approaches to Multiobjective Mathematical Programming under Uncertainty.Springer, Berlin, pp 45–68

47. Inuiguchi M, Sakawa M, Kume Y (1994) The usefulness of possibilistic programming in production planning problems. Int J Prod Econ 33:49–52

48. Jamison KD (1998) Modeling Uncertainty Using Probabilistic Based Possibility Theory with Applications to Optimization. Ph D Thesis, University of Colorado Denver, Department of Mathematical Sciences. http://www-math.cudenver.edu/graduate/thesis/jamison.pdf

49. Jamison KD (2000) Possibilities as cumulative subjective probabilities and a norm on the space of congruence classes of fuzzy numbers motivated by an expected utility functional. Fuzzy Sets Syst 111:331–339

50. Jamison KD, Lodwick WA (2001) Fuzzy linear programming using penalty method. Fuzzy Sets Syst 119:97–110

51. Jamison KD, Lodwick WA (2002) The construction of consistent possibility and necessity measures. Fuzzy Sets Syst 132(1):1–10

52. Jamison KD, Lodwick WA (2004) Interval-valued probability measures. UCD/CCM Report No. 213, March 2004

53. Joubert JW, Luhandjula MK, Ncube O, le Roux G, de Wet F (2007) An optimization model for the management of South African game ranch. Agric Syst 92:223–239

54. Kacprzyk J, Orlovski SA (eds) (1987) Optimization Models Using Fuzzy Sets and Possibility Theory. D Reidel, Dordrecht

55. Kacprzyk J, Orlovski SA (1987) Fuzzy optimization and mathematical programming: A brief introduction and survey. In: Kacprzyk J, Orlovski SA (eds) Optimization Models Using Fuzzy Sets and Possibility Theory. D Reidel, Dordrecht, pp 50–72

56. Kasperski A, Zeilinski P (2007) Using gradual numbers for solving fuzzy-valued combinatorial optimization problems. In: Melin P, Castillo O, Aguilar LT, Kacprzyk J, Pedrycz W (eds) Foundations of Fuzzy Logic and Soft Computing: 12th International Fuzzy System Association World Congress, IFSA 2007, Cancun, June 2007, Proceedings. Springer, Berlin, pp 656–665

57. Kaufmann A, Gupta MM (1985) Introduction to Fuzzy Arithmetic – Theory and Applications. Van Nostrand Reinhold, New York

58. Kaymak U, Sousa JM (2003) Weighting of Constraints in Fuzzy Optimization. Constraints 8:61–78 (also in the 2001 Proceedings of IEEE Fuzzy Systems Conference)

59. Klir GJ, Yuan B (1995) Fuzzy Sets and Fuzzy Logic. Prentice Hall, Upper Saddle River

60. Lai Y, Hwang C (1992) Fuzzy Mathematical Programming. Springer, Berlin

61. Lin TY (2005) A function theoretic view of fuzzy sets: New extension principle. In: Filev D, Ying H (eds) Proceedings of NAFIPS05

62. Liu B (1999) Uncertainty Programming. Wiley, New York

63. Liu B (2000) Dependent-chance programming in fuzzy environs. Fuzzy Sets Syst 109:97–106

64. Liu B (2001) Fuzzy random chance-constrained programming. IEEE Trans Fuzzy Syst 9:713–720

65. Liu B (2002) Theory and Practice of Uncertainty Programming. Physica, Heidelberg

66. Liu B, Iwamura K (2001) Fuzzy programming with fuzzy decisions and fuzzy simulation-based genetic algorithm. Fuzzy Sets Syst 122:253–262

67. Lodwick WA (1990) Analysis of structure in fuzzy linear programs. Fuzzy Sets Syst 38:15–26

68. Lodwick WA (1990) A generalized convex stochastic dominance algorithm. IMA J Math Appl Bus Ind 2:225–246

69. Lodwick WA (1999) Constrained Interval Arithmetic. CCM Report 138, Feb. 1999. CCM, Denver

70. Lodwick WA (ed) (2004) Special Issue on Linkages Between Interval Analysis and Fuzzy Set Theory. Reliable Comput 10

71. Lodwick WA (2007) Interval and fuzzy analysis: A unified approach. In: Advances in Imaging and Electronic Physics, vol 148. Elsevier, San Diego, pp 75–192

72. Lodwick WA, Bachman KA (2005) Solving large-scale fuzzy and possibilistic optimization problems: Theory, algorithms and applications. Fuzzy Optim Decis Making 4(4):257–278. (also UCD/CCM Report No. 216, June 2004)

73. Lodwick WA, Inuiguchi M (eds) (2007) Special Issue on Optimization Under Fuzzy and Possibilistic Uncertainty. Fuzzy Sets Syst 158:17; 1 Sept 2007

74. Lodwick WA, Jamison KD (1997) A computational method for fuzzy optimization. In: Bilal A, Madan G (eds) Uncertainty Analysis in Engineering and Sciences: Fuzzy Logic, Statistics, and Neural Network Approach, chap 19. Kluwer, Norwell

75. Lodwick WA, Jamison KD (eds) (2003) Special Issue: Interfaces Fuzzy Set Theory Interval Anal 135:1; April 1, 2003

76. Lodwick WA, Jamison KD (2003) Estimating and Validating the Cumulative Distribution of a Function of Random Variables: Toward the Development of Distribution Arithmetic. Reliab Comput 9:127–141

77. Lodwick WA, Jamison KD (2005) Theory and semantics for fuzzy and possibilistic optimization, Proceedings of the 11th International Fuzzy System Association World Congress. IFSA 2005, Beijing, July 2005

78. Lodwick WA, Jamison KD (2006) Interval-valued probability in the analysis of problems that contain a mixture of fuzzy, possibilistic and interval uncertainty. In: Demirli K, Akgunduz A (eds) 2006 Conference of the North American Fuzzy Information Processing Society, 3–6 June 2006, Montréal, Canada, paper 327137

79. Lodwick WA, Jamison KD (2008) Interval-Valued Probability in the Analysis of Problems Containing a Mixture of Fuzzy, Possibilistic, Probabilistic and Interval Uncertainty. Fuzzy Sets and Systems 2008

80. Lodwick WA, Jamison KD (2007) The use of interval-valued probability measures in optimization under uncertainty for problems containing a mixture of fuzzy, possibilistic, and interval uncertainty. In: Melin P, Castillo O, Aguilar LT, Kacprzyk J, Pedrycz W (eds) Foundations of Fuzzy Logic and Soft Computing: 12 th International Fuzzy System Association World Congress, IFSA 2007, Cancun, Mexico, June 2007, Proceedings. Springer, Berlin, pp 361–370

81. Lodwick WA, Jamison KD (2007) Theoretical and semantic distinctions of fuzzy, possibilistic, and mixed fuzzy/possibilistic optimization. Fuzzy Sets Syst 158(17):1861–1872

82. Lodwick WA, McCourt S, Newman F, Humpheries S (1999) Optimization Methods for Radiation Therapy Plans. In: Borgers C, Natterer F (eds) IMA Series in Applied Mathematics – Computational Radiology and Imaging: Therapy and Diagnosis. Springer, New York, pp 229–250

83. Lodwick WA, Neumaier A, Newman F (2001) Optimization under uncertainty: methods and applications in radiation therapy. Proc 10th IEEE Int Conf Fuzzy Syst 2001 3:1219–1222

84. Luhandjula MK (1986) On possibilistic linear programming. Fuzzy Sets Syst 18:15–30

85. Luhandjula MK (1989) Fuzzy optimization: an appraisal. Fuzzy Sets Syst 30:257–282

86. Luhandjula MK (2004) Optimisation under hybrid uncertainty. Fuzzy Sets Syst 146:187–203

87. Luhandjula MK (2006) Fuzzy stochastic linear programming: Survey and future research directions. Eur J Oper Res 174:1353–1367

88. Luhandjula MK, Ichihashi H, Inuiguchi M (1992) Fuzzy and semi-infinite mathematical programming. Inf Sci 61:233–250

89. Markowitz H (1952) Portfolio selection. J Finance 7:77–91

90. Moore RE (1979) Methods and Applications of Interval Analysis. SIAM, Philadelphia

91. Negoita CV (1981) The current interest in fuzzy optimization. Fuzzy Sets Syst 6:261–269

92. Negoita CV, Ralescu DA (1975) Applications of Fuzzy Sets to Systems Analysis. Birkhäuser, Boston

93. Negoita CV, Sularia M (1976) On fuzzy mathematical pro-

gramming and tolerances in planning. Econ Comput Cybern Stud Res 3(31):3–14

94. Neumaier A (2003) Fuzzy modeling in terms of surprise. Fuzzy Sets Syst 135(1):21–38

95. Neumaier A (2004) Clouds, fuzzy sets and probability intervals. Reliab Comput 10:249–272. Springer

96. Neumaier A (2005) Structure of clouds. (submitted – downloadable http://www.mat.univie.ac.at/~neum/papers.html)

97. Ogryczak W, Ruszczynski A (1999) From stochastic dominance to mean-risk models: Semideviations as risk measures. Eur J Oper Res 116:33–50

98. Nguyen HT (1978) A note on the extension principle for fuzzy sets. J Math Anal Appl 64:369–380

99. Ralescu D (1977) Inexact solutions for large-scale control problems. In: Proceedings of the 1st International Congress on Mathematics at the Service of Man, Barcelona

100. Ramik J (1986) Extension principle in fuzzy optimization. Fuzzy Sets Syst 19:29–35

101. Ramik J, Rimanek J (1985) Inequality relation between fuzzy numbers and its use in fuzzy optimization. Fuzzy Sets Syst 16:123–138

102. Ramik J, Vlach M (2002) Fuzzy mathematical programming: A unified approach based on fuzzy relations. Fuzzy Optim Decis Making 1:335–346

103. Ramik J, Vlach M (2002) Generalized Concavity in Fuzzy Optimization and Decision Analysis. Kluwer, Boston

104. Riverol C, Pilipovik MV (2007) Optimization of the pyrolysis of ethane using fuzzy programming. Chem Eng J 133:133–137

105. Riverol C, Pilipovik MV, Carosi C (2007) Assessing the water requirements in refineries using possibilistic programming. Chem Eng Process 45:533–537

106. Rommelfanger HJ (1994) Some problems of fuzzy optimization with T-norm based extended addition. In: Delgado M, Kacprzyk J, Verdegay J-L, Vila MA (eds) Fuzzy Optimization: Recent Advances. Physica, Heidelberg, pp 158–168

107. Rommelfanger HJ (1996) Fuzzy linear programming and applications. Eur J Oper Res 92:512–527

108. Rommelfanger HJ (2004) The advantages of fuzzy optimization models in practical use. Fuzzy Optim Decis Making 3:293–309

109. Rommelfanger HJ, Slowinski R (1998) Fuzzy linear programming with single or multiple objective functions. In: Slowinski R (ed) Fuzzy Sets in Decision Analysis, Operations Research and Statistics. The Handbooks of Fuzzy Sets. Kluwer, Netherlands, pp 179–213

110. Roubens M (1990) Inequality constraints between fuzzy numbers and their use in mathematical programming. In: Slowinski R, Teghem J (eds) Stochastic versus Fuzzy Approaches to Multiobjective Mathematical Programming Under Uncertainty. Kluwer, Netherlands, pp 321–330

111. Russell B (1924) Vagueness. Aust J Philos 1:84–92

112. Sahinidis N (2004) Optimization under uncertainty: State-of-the-art and opportunities. Comput Chem Eng 28:971–983

113. Saito S, Ishii H (1998) Existence criteria for fuzzy optimization problems. In: Takahashi W, Tanaka T (eds) Proceedings of the International Conference on Nonlinear Analysis and Convex Analysis, Niigata, Japan, 28–31 July 1998. World Scientific Press, Singapore, pp 321–325

114. Shafer G (1976) Mathematical A Theory of Evidence. Princeton University Press, Princeton

115. Shafer G (1987) Belief functions and possibility measures. In: James Bezdek C (ed) Analysis of Fuzzy Information. Mathematics and Logic, vol 1. CRC Press, Boca Raton, pp 51–84

116. Sousa JM, Kaymak U (2002) Fuzzy Decision Making in Modeling and Control. World Scientific Press, Singapore

117. Steinbach M (2001) Markowitz revisited: Mean-variance models in financial portfolio analysis. SIAM Rev 43(1):31–85

118. Tanaka H, Asai K (1984) Fuzzy linear programming problems with fuzzy numbers. Fuzzy Sets Syst 13:1–10

119. Tanaka H, Okuda T, Asai K (1973) Fuzzy mathematical programming. Trans Soc Instrum Control Eng 9(5):607–613; (in Japanese)

120. Tanaka H, Okuda T, Asai K (1974) On fuzzy mathematical programming. J Cybern 3(4):37–46

121. Tanaka H, Ichihasi H, Asai K (1984) A formulation of fuzzy linear programming problems based on comparison of fuzzy numbers. Control Cybern 13(3):185–194

122. Tanaka H, Ichihasi H, Asai K (1985) Fuzzy decisions in linear programming with trapezoidal fuzzy parameters. In: Kacprzyk J, Yager R (eds) Management Decision Support Systems Using Fuzzy Sets and Possibility Theory. Springer, Heidelberg, pp 146–159

123. Tang J, Wang D, Fung R (2001) Formulation of general possibilistic linear programming problems for complex industrial systems. Fuzzy Sets Syst 119:41–48

124. Thipwiwatpotjani P (2007) An algorithm for solving optimization problems with interval-valued probability measures. CCM Report, No. 259 (December 2007). CCM, Denver

125. Untiedt E (2006) Fuzzy and possibilistic programming techniques in the radiation therapy problem: An implementation-based analysis. Masters Thesis, University of Colorado Denver, Department of Mathematical Sciences (July 5, 2006)

126. Untiedt E (2007) A robust model for mixed fuzzy and possibilistic programming. Project report for Math 7593: Advanced Linear Programming. Spring, Denver

127. Untiedt E (2007) Using gradual numbers to analyze non-monotonic functions of fuzzy intervals. CCM Report No 258 (December 2007). CCM, Denver

128. Untiedt E, Lodwick WA (2007) On selecting an algorithm for fuzzy optimization. In: Melin P, Castillo O, Aguilar LT, Kacprzyk J, Pedrycz W (eds) Foundations of Fuzzy Logic and Soft Computing: 12th International Fuzzy System Association World Congress, IFSA 2007, Cancun, Mexico, June 2007, Proceedings. Springer, Berlin, pp 371–380

129. Vasant PM, Barsoum NN, Bhattacharya A (2007) Possibilistic optimization in planning decision of construction industry. Int J Prod Econ (to appear)

130. Verdegay JL (1982) Fuzzy mathematical programming. In: Gupta MM, Sanchez E (eds) Fuzzy Information and Decision Processes. North-Holland, Amsterdam, pp 231–237

131. Vila MA, Delgado M, Verdegay JL (1989) A general model for fuzzy linear programming. Fuzzy Sets Syst 30:21–29

132. Wang R, Liang T (2005) Applying possibilistic linear programming to aggregate production planning. Int J Prod Econ 98:328–341

133. Wang S, Zhu S (2002) On fuzzy portfolio selection problems. Fuzzy Optim Decis Making 1:361–377

134. Wang Z, Klir GJ (1992) Fuzzy Measure Theory. Plenum Press, New York

135. Weichselberger K (2000) The theory of interval-probability as a unifying concept for uncertainty. Int J Approx Reason 24:149–170

136. Werners B (1988) Aggregation models in mathematical programming. In: Mitra G (ed) Mathematical Models for Decision Support, NATO ASI Series vol F48. Springer, Berlin
137. Werners B (1995) Fuzzy linear programming – Algorithms and applications. Addendum to the Proceedings of ISUMA-NAPIPS'95, College Park, Maryland, 17–20 Sept 1995, pp A7-A12
138. Whitmore GA, Findlay MC (1978) Stochastic Dominance: An Approach to DecisionDecision-Making Under Risk. Lexington Books, Lexington
139. Yager RR (1980) On choosing between fuzzy subsets. Kybernetes 9:151–154
140. Yager RR (1981) A procedure for ordering fuzzy subsets of the unit interval. Inf Sci 24:143–161
141. Yager RR (1986) A characterization of the extension principle. Fuzzy Sets Syst 18:205–217
142. Zadeh LA (1965) Fuzzy Sets. Inf Control 8:338–353
143. Zadeh LA (1968) Probability measures of fuzzy events. J Math Anal Appl 23:421–427
144. Zadeh LA (1975) The concept of a linguistic variable and its application to approximate reasoning. Inf Sci, Part I: 8:199–249; Part II: 8:301–357; Part III: 9:43–80
145. Zadeh LA (1978) Fuzzy sets as a basis for a theory of possibility. Fuzzy Sets Syst 1:3–28
146. Zeleny M (1994) Fuzziness, knowledge and optimization: New optimality concepts. In: Delgado M, Kacprzyk J, Verdegay J-L, M Vila A (eds) Fuzzy Optimization: Recent Advances. Physica, Heidelberg, pp 3–20
147. Zimmermann HJ (1974) Optimization in fuzzy environment. Paper presented at the International XXI TIMS and 46th ORSA Conference, Puerto Rico Meeting, San Juan, Porto Rico, Oct 1974
148. Zimmermann HJ (1976) Description and optimization of fuzzy systems. Int J General Syst 2(4):209–216
149. Zimmermann HJ (1978) Fuzzy programming and linear programming with several objective functions. Fuzzy Sets Syst 1:45–55
150. Zimmermann HJ (1983) Fuzzy mathematical programming. Comput Oper Res 10:291–298

Fuzzy Probability Theory

MICHAEL BEER
National University of Singapore, Kent Ridge, Singapore

Article Outline

Glossary

Fuzzy set and fuzzy vector Let $\underline{\mathbf{X}}$ represent a universal set and $\underline{x}$ be the elements of $\underline{\mathbf{X}}$, then

$$\tilde{\underline{A}} = \{(\underline{x}, \mu_A(\underline{x})) \mid \underline{x} \in \underline{\mathbf{X}}\}, \ \mu_A(\underline{x}) \geq 0 \quad \forall \underline{x} \in \underline{\mathbf{X}} \quad (1)$$

is referred to as fuzzy set $\tilde{\underline{A}}$ on $\underline{\mathbf{X}}$. $\mu_A(\underline{x})$ is the membership function (characteristic function) of the fuzzy set $\tilde{\underline{A}}$ and represents the degree with which the elements $\underline{x}$ belong to $\tilde{\underline{A}}$. If

$$\sup_{\underline{x} \in \underline{\mathbf{X}}}[\mu_A(\underline{x})] = 1 \ , \tag{2}$$

the membership function and the fuzzy set $\tilde{\underline{A}}$ are called normalized; see Fig. 1. In case of a limitation to the Euclidean space $\underline{\mathbf{X}} = \mathbb{R}^n$ and normalized fuzzy sets, the fuzzy set $\tilde{\underline{A}}$ is also referred to as fuzzy vector denoted by $\tilde{\underline{x}}$ with its membership function $\mu(\underline{x})$, or, in the one-dimensional case, as fuzzy variable $\tilde{x}$ with $\mu(x)$.

α-Level set and support The crisp sets

$$\underline{A}_{\alpha_k} = \{\underline{x} \in \underline{\mathbf{X}} \mid \mu_A(\underline{x}) \geq \alpha_k\} \tag{3}$$

extracted from the fuzzy set $\tilde{\underline{A}}$ for real numbers $\alpha_k \in (0, 1]$ are called α-level sets. These comply with the inclusion property

$$\underline{A}_{\alpha_k} \subseteq \underline{A}_{\alpha_i} \quad \forall \alpha_i, \alpha_k \in (0, 1] \text{ with } \alpha_i \leq \alpha_k \ . \tag{4}$$

The largest α-level set $\underline{A}_{\alpha_k \to +0}$ is called support $S(\tilde{\underline{A}})$; see Fig. 1.

σ-Algebra A family $\mathfrak{M}(\underline{\mathbf{X}})$ of sets $\underline{A}_i$ on the universal set $\underline{\mathbf{X}}$ is referred to as σ-algebra $\mathfrak{S}(\underline{\mathbf{X}})$ on $\underline{\mathbf{X}}$, if

$$\underline{\mathbf{X}} \in \mathfrak{S}(\underline{\mathbf{X}}) \ , \tag{5}$$

$$\underline{A}_i \in \mathfrak{S}(\underline{\mathbf{X}}) \Rightarrow \underline{A}_i^C \in \mathfrak{S}(\underline{\mathbf{X}}) \ , \tag{6}$$
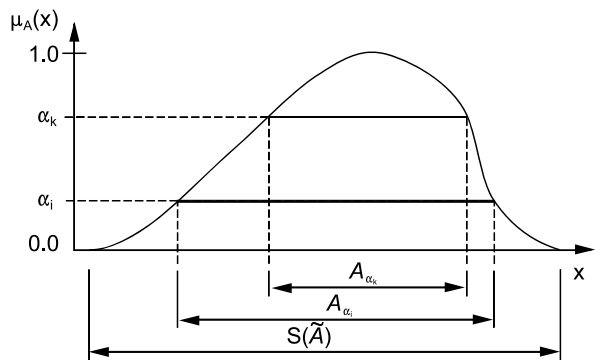
Fuzzy Probability Theory, Figure 1
Normalized fuzzy set with α-level sets and support

and if for every sequence of sets $\underline{A}_i$

$$\underline{A}_i \in \mathfrak{S}(\underline{\mathbf{X}}); \ i = 1, 2, \ldots \Rightarrow \bigcup_i^{\infty} \underline{A}_i \in \mathfrak{S}(\underline{\mathbf{X}}) \,. \qquad (7)$$

In this definition, $\underline{A}_i^C$ is the complementary set of $\underline{A}_i$ with respect to $\underline{\mathbf{X}}$, a family $\mathfrak{M}(\underline{\mathbf{X}})$ of sets $\underline{A}_i$ refers to subsets and systems of subsets of the power set $\mathfrak{P}(\underline{\mathbf{X}})$ on $\underline{\mathbf{X}}$, and the power set $\mathfrak{P}(\underline{\mathbf{X}})$ is the set of all subsets $\underline{A}_i$ of $\underline{\mathbf{X}}$.

Definition of the Subject

Fuzzy probability theory is an extension of probability theory to dealing with mixed probabilistic/non-probabilistic uncertainty. It provides a theoretical basis to model uncertainty which is only partly characterized by randomness and defies a pure probabilistic modeling with certainty due to a lack of trustworthiness or precision of the data or a lack of pertinent information. The fuzzy probabilistic model is settled between the probabilistic model and non-probabilistic uncertainty models. The significance of fuzzy probability theory lies in the treatment of the elements of a population not as crisp quantities but as set-valued quantities or granules in an uncertain fashion, which largely complies with reality in most everyday situations. Probabilistic and non-probabilistic uncertainty can so be transferred adequately and separately to the results of a subsequent analysis. This enables best case and worst case estimates in terms of probability taking account of variations within the inherent non-probabilistic uncertainty. The development of fuzzy probability theory was initiated by H. Kwakernaak with the introduction of fuzzy random variables in [47] in 1978. Subsequent developments have been reported in different directions and from different perspectives including differences in terminology. The usefulness of the theory has been underlined with various applications beyond mathematics and information science, in particular, in engineering. The application fields are not limited and may be extended increasingly, for example, to medicine, biology, psychology, economy, financial sciences, social sciences, and even to law.

Introduction

The probably most popular example of fuzzy probability is the evaluation of a survey on the subjective assessment of temperature. A group of test persons are asked – under equal conditions – to give a statement on the current temperature as realistically as possible. If the scatter of the statements is considered as random, the mean value of the statements provides a reasonable statistical point estimate for the actual temperature. The statements are, however, naturally given in an uncertain form. The test persons enunciate their perception in a form such as *about 25°C*, *possibly 27°C*, *between 24°C and 26°C*, or they even only come up with linguistic assessments such as *warm*, *very warm*, or *pleasant*. This uncertainty is non-probabilistic but has to be taken into account in the statistical evaluation. It is transferred to the estimated mean value, which is no longer obtained as a crisp number but as a value range or a set of values corresponding to the possibilities within the range of uncertainty of the statements. If the uncertain statements are initially quantified as fuzzy values, the mean value is obtained as a fuzzy value, too; and the probability of certain events is also computed as a fuzzy quantity – referred to as fuzzy probability. This example is a typical materialization of the following general real-world situation.

The numerical representation of a physical quantity with the aid of crisp numbers $x \in \mathbb{R}$ or sets thereof is frequently interfered by uncertainty regarding the trustworthiness of measured, or otherwise specified, values. The perceptions of physical quantities may appear, for example, as imprecise, diffuse, vague, dubious, or ambiguous. Underlying reasons for this phenomenon include the limited precision of any measurement (digital or analog), indirect measurements via auxiliary quantities in conjunction with a – more or less trustworthy – model to eventually determine the value wanted, measurements under weakly specified or arbitrarily changing boundary conditions, and the specification of values by experts in a linguistic manner. The type of the associated uncertainty is non-frequentative and improper for a subjective probabilistic modeling; hence, it is non-probabilistic. This uncertainty is unavoidable and may always be made evident by a respective choice of scale.

If a set of uncertain perceptions of a physical quantity is present in the form of a random sample, then the overall uncertainty possesses a mixed probabilistic/non-probabilistic character. Whilst the scatter of the realizations of the physical quantity possesses a probabilistic character (frequentative or subjective), each particular realization from the population may, additionally, exhibit non-probabilistic uncertainty. Consequently, a realistic modeling in those cases must involve both probabilistic and non-probabilistic uncertainty. This modeling without distorting or ignoring information is the mission of fuzzy probability theory. A pure probabilistic modeling would introduce unwarranted information in the form of a distribution function that cannot be justified and

would thus diminish the trustworthiness of the probabilistic results.

Mathematical Environment

Fuzzy probability is part of the framework of generalized information theory [38] and is covered by the umbrella of granular computing [50,62]. It represents a special case of imprecise probabilities [15,78] with ties to concepts of random sets [52]. This is underlined by Walley's summary of the semantics of imprecise probabilities with the term *indeterminacy*, which arises from ignorance about facts, events, or dependencies. Within the class of mathematical models covered by the term *imprecise probabilities*, see [15, 38,78], fuzzy probability theory has a relationship to concepts known as upper and lower probabilities [28], sets of probability measures [24], distribution envelops [7], interval probabilities [81], and p-box approach [23]. Also, similarities exist with respect to evidence theory (or Dempster–Shafer theory) [20,70] as a theory of infinitely monotone Choquet capacities [39,61]. The relationship to the latter is associated with the interpretation of the measures plausibility and belief, with the special cases of possibility and necessity, as upper and lower probabilities, respectively, [8].

Fuzzy probability shares the common feature of all imprecise probability models: the uncertainty of an event is characterized with a set of possible measure values in terms of probability, or with bounds on probability. Its distinctive feature is that set-valued information, and hence the probability of associated events, is described with the aid of uncertain sets according to fuzzy set theory [83,86]. This represents a marriage between fuzzy methods and probabilistics with fuzziness and randomness as special cases, which justifies the denotation as *fuzzy randomness*. Fuzzy probability theory enables a consideration of a fuzzy set of possible probabilistic models over the range of imprecision of the knowledge about the underlying randomness. The associated fuzzy probabilities provide weighted bounds on probability – the weights of which are obtained as the membership values of the fuzzy sets. Based on α-discretization [86] and the representation of fuzzy sets as sets of α-level sets, the relationship of fuzzy probability theory to various concepts of imprecise probabilities becomes obvious. For each α-level a common interval probability, crisp bounds on probability, or a classical set of probability measures, respectively, are obtained. The α-level sets of fuzzy events can be treated as random sets. Further, a relationship of these random sets to evidence theory can be constructed if a respective basic probability assignment is selected; see [21]. Consistency

with evidence theory is obtained if the focal sets are specified as fuzzy elementary events and if the basic probability assignment follows a discrete uniform distribution over the fuzzy elementary events.

The model fuzzy randomness with its two components – fuzzy methods and probabilistics – can utilize a distinction between aleatory and epistemic uncertainty with respect to the sources of uncertainty [29]. This is particularly important in view of practical applications. Irreducible uncertainty as a property of the system associated with fluctuations/variability may be summarized as aleatory uncertainty and described probabilistically, and reducible uncertainty as a property of the analysts, or its perception, associated with a lack of knowledge/precision may be understood as epistemic uncertainty and described with fuzzy sets. The model fuzzy randomness then combines, without mixing, both components in the form of a fuzzy set of possible probabilistic models over some particular range of imprecision. This distinction is retained throughout any subsequent analysis and reflected in the results.

The development of fuzzy probability was initiated with the introduction of fuzzy random variables by Kwakernaak [47,48] in 1978/79. Subsequent milestones were set by Puri and Ralescu [63], Kruse and Meyer [46], Wang and Zhang [80], and Krätschmer [43]. The developments show differences in terminology, concepts, and in the associated consideration of measurability; and the investigations are ongoing [11,12,14,17,35,36,37,45,73]. Krätschmer [43] showed that the different concepts can be unified to a certain extent. Generally, it can be noted that α-discretization is utilized as a helpful instrument. An overview with specific comments on the different developments is provided in [43] and [57]. Investigations were pursued on independent and dependent fuzzy random variables for which parameters were defined with particular focus on variance and covariance [22,32,40,41,60]. Fuzzy random processes were examined to reveal properties of limit theorems and martingales associated with fuzzy randomness [44,65]; see also [71,79] and for a survey [49]. Particular interest was devoted to the strong law of large numbers [10,33,34]. Further, the differentiation and the integration of fuzzy random quantities was investigated in [51,64].

Driven by the motivation for the establishment of fuzzy probability theory considerable effort was made in the modeling and statistical evaluation of imprecise data. Fundamental achievements were reported by Kruse and Meyer [46], Bandemer and Näther [2,5], and by Viertl [75]. Classical statistical methods were extended in order to take account of statistical fluctuations/variability and imprecision simultaneously, and the specific fea-

tures associated with the imprecision of the data were investigated. Ongoing research is reported, for example, in [58,74] in view of evaluating measurements, in [51,66] for decision making, and in [16,42,59] for regression analysis. Methods for evaluating imprecise data with the aid of generalized histograms are discussed in [9,77]. Also, the application of resampling methods is pursued; bootstrap concepts are utilized for statistical estimations [31] and hypothesis testing [26] based on imprecise data. Another method for hypothesis testing is proposed in [27], which employs fuzzy parameters in order to describe a fuzzy transition between rejection and acceptance. Bayesian methods have also been extended by the inclusion of fuzzy variables to take account of imprecise data; see [75] for basic considerations. A contribution to Bayesian statistics with imprecise prior distributions is presented in [76]. This leads to imprecise posterior distributions, imprecise predictive distributions, and may be used to deduce imprecise confidence intervals. A combination of the Bayesian theorem with kriging based on imprecise data is described in [3]. A Bayesian test of fuzzy hypotheses is discussed in [72], while in [67] the application of a fuzzy Bayesian method for decision making is presented.

In view of practical applications, probability distribution functions are defined for fuzzy random quantities [54,75,77,85] – despite some drawback [6]. These distribution functions can easily be formulated and used for further calculations, but they do not uniquely describe a fuzzy random quantity. This theoretical lack is, however, generally without an effect in practical applications so that stochastic simulations may be performed according to the distribution functions. Alternative simulation methods were proposed based on parametric [13] and non-parametric [6,55] descriptions of imprecision. The approach according to [55] enables a direct generation of fuzzy realizations based on a new concept for an incremental representation of fuzzy random quantities. This method is designed to simulate and predict fuzzy time series; it circumvents the problems of artificial uncertainty growth or bias of non-probabilistic uncertainty, which is frequently concerned with numerical simulations.

This variety of theoretical developments provides reasonable margins for the formulation of fuzzy probability theory but does not allow the definition of a unique concept. Choices have to be made within the elaborated margins depending on the envisaged application and environment. For the subsequent sections these choices are made in view of a broad spectrum of possible applications, for example, in civil/mechanical engineering [54]. These choices concern the following three main issues.

First, measurability has to be ensured according to a sound concept. According to [43], the concepts of measurable bounding functions [47,48], of measurable mappings of α-level sets [63], and of measurable fuzzy valued mappings [17,36] are available; or the unifying concept proposed in [43] itself, which utilizes a special topology on the space of fuzzy realizations, may be selected. From a practical point of view the concept of measurable bounding functions is most reasonable due to its analogy to traditional probability theory. On this basis, a fuzzy random quantity can be regarded as a fuzzy set of traditional, crisp, random quantities, each one carrying a certain membership degree. Each of these crisp random quantities is then measurable in the traditional fashion, and their membership degrees can be transferred to the respective measure values. The set of the obtained measure values including their membership degrees then represents a fuzzy probability.

Second, a concept for the integration of a fuzzy-valued function has to be selected from the different available approaches [86]. This is associated with the computation of the probability of a fuzzy event. An evaluation in a mean sense weighted by the membership function of the fuzzy event is suggested in [84], which leads to a scalar value for the probability. This evaluation is associated with the interpretation that an event may occur partially. An approach for calculating the probability of a fuzzy event as a fuzzy set is proposed in [82]; the resulting fuzzy probability then represents a set of measure values with associated membership degrees. This complies with the interpretation that the occurrence of an event is binary but it is not clearly indicated if the event has occurred or not. The imprecision is associated with the observation rather than with the event. The latter approach corresponds with the practical situation in many cases, provides useful information in form of the imprecision reflected in the measure values, and follows the selected concept of measurability. It is thus taken as a basis for further consideration.

Third, the meaning of the distance between fuzzy sets as realizations of a fuzzy random quantity must be defined, which is of particular importance for the definition of the variance and of further parameters of fuzzy random quantities. An approach that follows a set-theoretical point of view and leads to a crisp distance measure is presented in [40]. It is proposed to apply the Hausdorff metric to the α-level sets of a fuzzy quantity and to average the results over the membership scale. Consequently, parameters of a fuzzy random quantity which are associated with a distance between fuzzy realizations reflect the variability within the imprecision merely in an integrated form. For example, the variance of a fuzzy random variable is

obtained as a crisp value. In contrast to this, the application of standard algorithms for operations on fuzzy sets, such as the extension principle, [4,39,86] leads to fuzzy distances between fuzzy sets. Parameters, including variances, of fuzzy random quantities are then obtained as fuzzy sets of possible parameter values. This corresponds to the interpretation of fuzzy random quantities as fuzzy sets of traditional, crisp, random quantities. The latter approach is thus pursued further.

These three selections comply basically with the definitions in [46] and [47]; see also [57]. Among all the possible choices, this set-up ensures the most plausible settlement of fuzzy probability theory within the framework of imprecise probabilities [15,38,78] with ties to evidence theory and random set approaches [21,30]. Fuzzy probability is obtained in the form of weighted plausible bounds on probability. Moreover, in view of practical applications, the treatment of fuzzy random quantities as fuzzy sets of traditional random quantities enables the utilization of established probabilistic methods as kernel solutions in the environment of a fuzzy analysis. For example, sophisticated methods of Monte Carlo simulation [25,68,69] may be combined with a generally applicable fuzzy analysis based on an global optimization approach using α-discretization [53]. If some restricting conditions are complied with, numerically efficient methods from interval mathematics [1] may be employed for the α-level mappings instead of a global optimization approach; see [56]. The selected concept, eventually, enables

best-case and worst-case studies within the range of possible probabilistic models.

Fuzzy Random Quantities

With the above selections, the definitions from traditional probability theory can be adopted and extended to dealing with imprecise outcomes from a random experiment. Let Ω be the space of random elementary events $\omega \in \Omega$ and the universe on which the realizations are observed be the n-dimensional Euclidean space $\underline{\mathbf{X}} = \mathbb{R}^n$. Then, a membership scale μ is introduced perpendicular to the hyperplane $\Omega \times \underline{\mathbf{X}}$. This enables the specification of fuzzy sets on $\underline{\mathbf{X}}$ for given elementary events ω from Ω without interaction between Ω and μ. That is, randomness induced by Ω and fuzziness described by μ – only in x-direction – are not mixed with one another. Let $\mathfrak{F}(\underline{\mathbf{X}})$ be the set of all fuzzy quantities on $\underline{\mathbf{X}} = \mathbb{R}^n$; that is, $\mathfrak{F}(\underline{\mathbf{X}})$ denotes the collection of all fuzzy sets $\underline{\tilde{A}}$ on $\underline{\mathbf{X}} = \mathbb{R}^n$, with $\underline{\tilde{A}}$ according to Eq. (1). Then, the imprecise result of the mapping

$$\underline{\tilde{X}}: \Omega \to \mathfrak{F}(\underline{\mathbf{X}}) \tag{8}$$

is referred to as fuzzy random quantity $\underline{\tilde{X}}$. In contrast to real-valued random quantities, a fuzzy realization $\underline{\tilde{x}}(\omega) \in \mathfrak{F}(\underline{\mathbf{X}})$, or $\underline{\tilde{x}}(\omega) \subseteq \underline{\mathbf{X}}$, is now assigned to each elementary event $\omega \in \Omega$; see Fig. 2. These fuzzy realizations may be understood as a numerical representation of granules. Generally, a fuzzy random quantity can be discrete
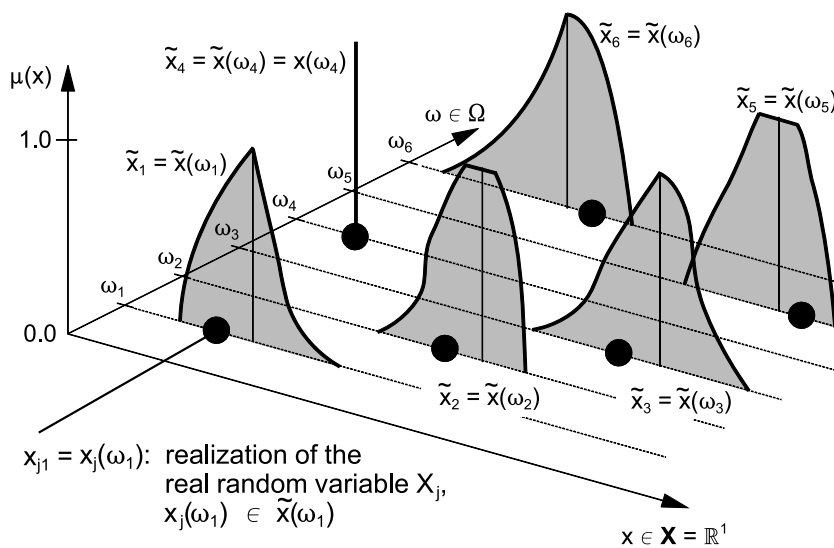
Fuzzy Probability Theory, Figure 2
Fuzzy random variable

or continuous with respect to both fuzziness and randomness. The further consideration refers to the continuous case, from which the discrete case may be derived.

Without a limitation in generality, the fuzzy realizations may be restricted to normalized fuzzy quantities, thus representing fuzzy vectors. Further restrictions can be defined in view of a convenient numerical treatment if the application field allows for. This concerns, for example, a restriction to connected and compact α-level sets $\underline{A}_\alpha = \underline{x}_\alpha$ of the fuzzy realizations, a restriction to convex fuzzy sets as fuzzy realizations (a fuzzy set $\tilde{A}$ is convex if all its α-level sets $\underline{A}_\alpha$ are convex sets), or a restriction to fuzzy realizations with only one element $\underline{x}_i$ carrying the membership $\mu(\underline{x}_i) = 1$ as in [47].

For the treatment of the fuzzy realizations, a respective algorithm for operations on fuzzy sets has to be selected. Following the literature and the above selection, the standard basis is employed with the min-operator as a special case of a t-norm and the associated max-operator as a special case of a t-co-norm [18,19,86]. This leads to the min-max operator and the extension principle [4,39,86].

With the interpretation of a fuzzy random quantity as a fuzzy set of real-valued random quantities, according to the above selection, the following relationship to traditional probability theory is obtained. Let $\underline{x}_{ji}$ be a realization of a real-valued random quantity $\underline{X}_j$ and $\tilde{x}_i$ be a fuzzy realization of a fuzzy random quantity $\tilde{X}$ with $\underline{x}_{ji}$ and $\tilde{x}_i$ be assigned to the same elementary event ω_i. If $\underline{x}_{ji} \in \tilde{x}_i$, then $\underline{x}_{ji}$ is called contained in $\tilde{x}_i$. If, for all elementary events $\omega_i \in \Omega$, $i = 1, 2, \ldots$, the $\underline{x}_{ji}$ are contained in the $\tilde{x}_i$, the set of the $\underline{x}_{ji}$, $i = 1, 2, \ldots$, then constitutes an original $\underline{X}_j$ of the fuzzy random quantity $\tilde{X}$; see Fig. 2. The original $\underline{X}_j$ is referred to as completely contained in $\tilde{X}$, $\underline{X}_j \in \tilde{X}$. Each real-valued random quantity $\underline{X}$ that is completely contained in $\tilde{X}$ is an original $\underline{X}_j$ of $\tilde{X}$ and carries the membership degree

$$\mu(\underline{X}_j) = \max[\alpha \,|\, \underline{x}_{ji} \in \underline{x}_{i\alpha} \,\forall i.] \,. \tag{9}$$

That is, in the Ω-direction, each original $\underline{X}_j$ must be consistent with the fuzziness of $\tilde{X}$. Consequently, the fuzzy random quantity $\tilde{X}$ can be represented as the fuzzy set of all originals $\underline{X}_j$ contained in $\tilde{X}$,

$$\tilde{X} = \left\{ (\underline{X}_j, \mu(\underline{X}_j)) \,|\, \underline{x}_{ji} \in \tilde{x}_i \,\forall i. \right\} \,. \tag{10}$$

Each fuzzy random quantity $\tilde{X}$ contains at least one real-valued random quantity $\underline{X}$ as an original $\underline{X}_j$ of $\tilde{X}$. Each fuzzy random quantity $\tilde{X}$ that possesses precisely one original is thus a real-valued random quantity $\underline{X}$. That is, real-valued random quantities are a special case of fuzzy random quantities. This enables a simultaneous treatment of

real-valued random quantities and fuzzy random quantities within the same theoretical environment and with the same numerical algorithms. Or, vice versa, it enables the utilization of theoretical results and established numerical algorithms from traditional probability theory within the framework of fuzzy probability theory.

If α-discretization is applied to the fuzzy random quantity $\tilde{X}$, random α-level sets $\underline{X}_\alpha$ are obtained,

$$\underline{X}_\alpha = \{ \underline{X} = \underline{X}_j \,|\, \mu(\underline{X}_j) \geq \alpha \} \,. \tag{11}$$

Their realizations are α-level sets $\underline{x}_{i\alpha}$ of the respective fuzzy realizations $\tilde{x}_i$ of the fuzzy random quantity $\tilde{X}$. A fuzzy random quantity can thus, alternatively, be represented by the set of its α-level sets,

$$\tilde{X} = \left\{ (\underline{X}_\alpha, \mu(\underline{X}_\alpha)) \,|\, \mu(\underline{X}_\alpha) = \alpha \,\forall \alpha \in (0, 1] \right\} \,. \tag{12}$$

In the one-dimensional case and with the restriction to connected and compact α-level sets of the realizations, the random α-level sets X_α of the fuzzy random variable $\tilde{X}$ become closed random intervals $[X_{\alpha l}, X_{\alpha r}]$.

Fuzzy Probability

Fuzzy probability is derived as a fuzzy set of probability measures for events the occurrence of which depends on the behavior of a fuzzy random quantity. These events are referred to as fuzzy random events with the following characteristics.

Let $\tilde{X}$ be a fuzzy random quantity according to Eq. (8) with the realizations $\tilde{x}$ and $\mathfrak{S}(\mathbf{X})$ be a σ-algebra of sets $\underline{A}_i$ defined on $\mathbf{X}$. Then, the event

$$\tilde{E}_i : \tilde{X} \text{ hits } \underline{A}_i \tag{13}$$

is referred to as fuzzy random event, which occurs if a fuzzy realization $\tilde{x}$ of the fuzzy random quantity $\tilde{X}$ hits the set $\underline{A}_i$. The associated probability of occurrence of $\tilde{E}_i$ is referred to as fuzzy probability $\tilde{P}(\underline{A}_i)$. It is obtained as the fuzzy set of the probabilities of occurrence of the events

$$E_{ij} : \underline{X}_j \in \underline{A}_i \tag{14}$$

associated with all originals $\underline{X}_j$ of the fuzzy random quantity $\tilde{X}$ with their membership values $\mu(\underline{X}_j)$. Specifically,

$$\begin{aligned} \tilde{P}(\underline{A}_i) = \{ &(P(\underline{X}_j \in \underline{A}_i), \mu(P(\underline{X}_j \in \underline{A}_i))) \\ &|\, \underline{X}_j \in \tilde{X}, \mu(P(\underline{X}_j \in \underline{A}_i)) = \mu(\underline{X}_j) \,\forall j \} \,. \end{aligned} \tag{15}$$

Each of the involved probabilities $P(\underline{X}_j \in \underline{A}_i)$ is a traditional, real-valued probability associated with the traditional probability space $[\mathbf{X}, \mathfrak{S}, P]$ and complying with all
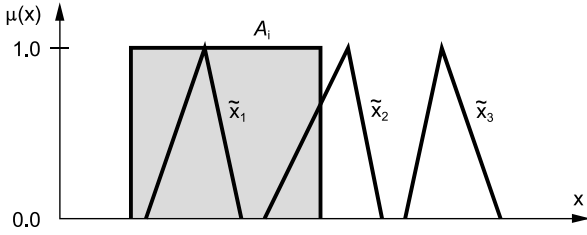
Fuzzy Probability Theory, Figure 3
Fuzzy event $\tilde{\underline{x}}_k$ hits $\underline{A}_i$ in the one-dimensional case

established theorems and properties of traditional probability. For a formal closure of fuzzy probability theory, the membership scale μ is incorporated in the probability space to constitute the fuzzy probability space denoted by $[\underline{X}, \mathfrak{S}, P, \mu]$ or $[\underline{X}, \mathfrak{S}, \tilde{P}]$.

The evaluation of the fuzzy random event Eq. (13) hinges on the question whether a fuzzy realization $\tilde{\underline{x}}_k$ hits the set $\underline{A}_i$ or not. Due to the fuzziness of the $\tilde{\underline{x}}$, these events appear as fuzzy events $\tilde{E}_{ik}$: $\tilde{\underline{x}}_k$ *hits* $\underline{A}_i$ with the following three options for occurrence; see Fig. 3:

- The fuzzy realization $\tilde{\underline{x}}_k$ lies completely inside the set $\underline{A}_i$, the event $\tilde{E}_{ik}$ has occurred.
- The fuzzy realization $\tilde{\underline{x}}_k$ lies only partially inside $\underline{A}_i$, the event $\tilde{E}_{ik}$ may have occurred or not occurred.
- The fuzzy realization $\tilde{\underline{x}}_k$ lies completely outside the set $\underline{A}_i$, the event $\tilde{E}_{ik}$ has not occurred.

The fuzzy probability $\tilde{P}(\underline{A}_i)$ takes account of all three options within the range of fuzziness. The fuzzy random quantity $\tilde{\underline{X}}$ is discretized into a set of random α-level sets $\underline{X}_\alpha$ according to Eq. (11), and the events $\tilde{E}_i$ and $\tilde{E}_{ik}$, re-

spectively, are evaluated α-level by α-level. In this evaluation, the event $E_{ik\alpha}$: $\underline{x}_{k\alpha}$ *hits* $\underline{A}_i$ admits the following two "extreme" interpretations of occurrence:

- $E_{ik\alpha l}$: "$\underline{x}_{k\alpha}$ is contained in $\underline{A}_i$: $\underline{x}_{k\alpha} \subseteq \underline{A}_i$", and
- $E_{ik\alpha r}$: "$\underline{x}_{k\alpha}$ and $\underline{A}_i$ possess at least one element in common: $\underline{x}_{k\alpha} \cap \underline{A}_i \neq \emptyset$".

Consequently, the events $E_{ik\alpha l}$ are associated with the smallest probability

$$P_{\alpha l}(\underline{A}_i) = P(\underline{X}_\alpha \subseteq \underline{A}_i), \qquad (16)$$

and the events $E_{ik\alpha r}$ correspond to the largest probability

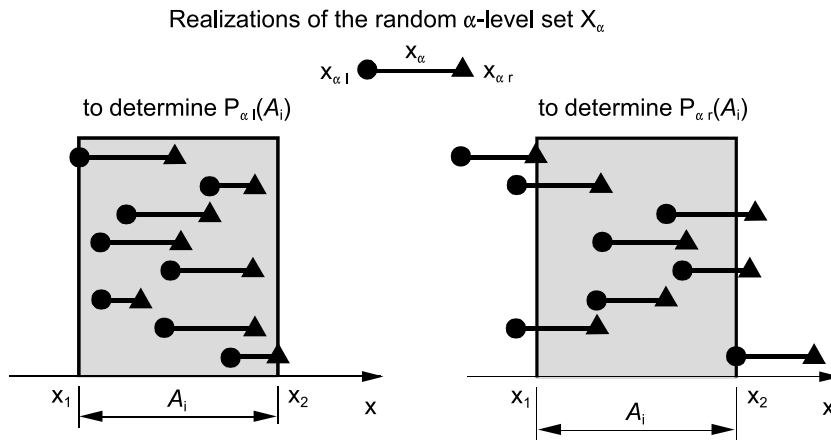$$P_{\alpha r}(\underline{A}_i) = P(\underline{X}_\alpha \cap \underline{A}_i \neq \emptyset). \qquad (17)$$

The probabilities $P_{\alpha l}(\underline{A}_i)$ and $P_{\alpha r}(\underline{A}_i)$ are bounds on the probability $\tilde{P}(\underline{A}_i)$ on the respective α-level associated with the random α-level set $\underline{X}_\alpha$ of the fuzzy random quantity $\tilde{\underline{X}}$; see Fig. 4. As all elements of $\underline{X}_\alpha$ are originals $\underline{X}_j$ of $\tilde{\underline{X}}$, the probability that an $\underline{X}_j \in \underline{X}_\alpha$ hits $\underline{A}_i$ is bounded according to

$$P_{\alpha l}(\underline{A}_i) \leq P(\underline{X}_j \in \underline{A}_i) \leq P_{\alpha r}(\underline{A}_i) \quad \forall \underline{X}_j \in \underline{X}_\alpha . \quad (18)$$

This enables a computation of the probability bounds directly from the real-valued probabilities $P(\underline{X}_j \in \underline{A}_i)$ associated with the originals $\underline{X}_j$,

$$P_{\alpha l}(\underline{A}_i) = \min_{\underline{X}_j \in \underline{X}_\alpha} P(\underline{X}_j \in \underline{A}_i), \qquad (19)$$

$$P_{\alpha r}(\underline{A}_i) = \max_{\underline{X}_j \in \underline{X}_\alpha} P(\underline{X}_j \in \underline{A}_i). \qquad (20)$$

Fuzzy Probability Theory, Figure 4
Events for determining $P_{\alpha l}(\underline{A}_i)$ and $P_{\alpha r}(\underline{A}_i)$ in the one-dimensional case

If the fuzzy random quantity $\underline{\tilde{X}}$ represents a fuzzy set of continuous real-valued random quantities and if the membership functions of all fuzzy realizations $\underline{\tilde{x}}_i$ of $\underline{\tilde{X}}$ are at least segmentally continuous, then the probability bounds $P_{\alpha l}(\underline{A}_i)$ and $P_{\alpha r}(\underline{A}_i)$ determine closed connected intervals $[P_{\alpha l}(\underline{A}_i), P_{\alpha r}(\underline{A}_i)]$. In this case, the fuzzy probability $\tilde{P}(\underline{A}_i)$ is obtained as a continuous and convex fuzzy set, which may be specified uniquely with the aid of α-discretization,

$$\tilde{P}(\underline{A}_i) = \big\{(P_\alpha(\underline{A}_i), \mu(P_\alpha(\underline{A}_i)))|P_\alpha(\underline{A}_i)$$
$$= [P_{\alpha l}(\underline{A}_i), P_{\alpha r}(\underline{A}_i)], \mu(P_\alpha(\underline{A}_i)) = \alpha$$
$$\forall \alpha \in (0, 1]\big\}. \quad (21)$$

The properties of the fuzzy probability $\tilde{P}(\underline{A}_i)$ result from the properties of the traditional probability measure in conjunction with fuzzy set theory. For example, a complementary relationship may be derived for $\tilde{P}(\underline{A}_i)$ as follows. The equivalence

$$(\underline{X}_\alpha \subseteq \underline{A}_i) \Leftrightarrow (\underline{X}_\alpha \cap \underline{A}_i^C = \emptyset) \quad (22)$$

with $\underline{A}_i^C$ being the complementary set of $\underline{A}_i$ with respect to the universe $\underline{X}$, leads to

$$P(\underline{X}_\alpha \subseteq \underline{A}_i) = P(\underline{X}_\alpha \cap \underline{A}_i^C = \emptyset) \quad (23)$$

for each α-level. If the event $\underline{X}_\alpha \cap \underline{A}_i^C = \emptyset$ is expressed in terms of its complementary event $\underline{X}_\alpha \cap \underline{A}_i^C \neq \emptyset$, Eq. (23) can be rewritten as

$$P(\underline{X}_\alpha \subseteq \underline{A}_i) = 1 - P(\underline{X}_\alpha \cap \underline{A}_i^C \neq \emptyset). \quad (24)$$

This leads to the relationships

$$P_{\alpha l}(\underline{A}_i) = 1 - P_{\alpha r}(\underline{A}_i^C), \quad (25)$$

$$P_{\alpha r}(\underline{A}_i) = 1 - P_{\alpha l}(\underline{A}_i^C), \quad (26)$$

and

$$\tilde{P}(\underline{A}_i) = 1 - \tilde{P}(\underline{A}_i^C). \quad (27)$$

In the special case that the set $\underline{A}_i$ contains only one element $\underline{A}_i = \underline{x}_i$, the fuzzy probability $\tilde{P}(\underline{A}_i)$ changes to $\tilde{P}(\underline{x}_i)$. The event $\underline{X}_\alpha \cap \underline{A}_i \neq \emptyset$ is then replaced by $\underline{x}_i \in \underline{X}_\alpha$, and $\underline{X}_\alpha \subseteq \underline{A}_i$ becomes $\underline{X}_\alpha = \underline{x}_i$. The probability $P_{\alpha l}(\underline{x}_i) = P(\underline{X}_\alpha = \underline{x}_i)$ may take values greater than zero only if a realization of $\underline{X}_\alpha$ exists that possesses exactly one element $\underline{X}_\alpha = \underline{t}$ with $\underline{t} = \underline{x}_i$ and if this element $\underline{t}$ represents a realization of a discrete original of $\underline{X}_\alpha$. Otherwise, $P_{\alpha l}(\underline{x}_i) = 0$, and the fuzziness of $\tilde{P}(\underline{x}_i)$ is exclusively specified by $P_{\alpha r}(\underline{x}_i)$.

In the one-dimensional case with

$$A_i = \{x \mid x \in \mathbf{X}; x_1 \leq x \leq x_2\} \quad (28)$$

the fuzzy probability $\tilde{P}(\underline{A}_i)$ can be represented in a simplified manner. If the random α-level sets X_α are restricted to be closed random intervals $[X_{\alpha l}, X_{\alpha r}]$, the associated fuzzy random variable $\tilde{X}$ can be completely described by means of the bounding real-valued random quantities $X_{\alpha l}$ and $X_{\alpha r}$; see Fig. 4. $X_{\alpha l}$ and $X_{\alpha r}$ represent specific originals X_j of $\tilde{X}$. This enables the specification of the probability bounds for each α-level according to

$$P_{\alpha l}(A_i) = \max \big[0, P(X_{\alpha r} = t_r \mid x_2, t_r \in \mathbf{X}, t_r \leq x_2)$$
$$- P(X_{\alpha l} = t_l \mid x_1, t_l \in \mathbf{X}, t_l < x_1)\big], \quad (29)$$

and

$$P_{\alpha r}(A_i) = P(X_{\alpha l} = t_l \mid x_2, t_l \in \mathbf{X}; t_l \leq x_2)$$
$$- P(X_{\alpha r} = t_r \mid x_1, t_r \in \mathbf{X}; t_r < x_1). \quad (30)$$

From the above framework, the special case of real-valued random quantities $\underline{X}$, may be reobtained as a fuzzy random quantity $\underline{\tilde{X}}$ that contains precisely one original $\underline{X}_j = \underline{X}_1$,

$$\underline{X}_1 = (\underline{X}_{j=1}, \mu(\underline{X}_{j=1}) = 1); \quad (31)$$

see Eq. (10). Then, all $\underline{X}_\alpha$ contain only the sole original $\underline{X}_1$, and both $\underline{X}_\alpha \cap \underline{A}_i \neq \emptyset$ and $\underline{X}_\alpha \subseteq \underline{A}_i$ reduce to $\underline{X}_1 \in \underline{A}_i$. That is, the event $\underline{X}_1$ *hits* $\underline{A}_i$ does no longer provide options for interpretation reflected as fuzziness,

$$P_{\alpha l}(\underline{A}_i) = P_{\alpha r}(\underline{A}_i) = P(\underline{A}_i) = P(\underline{X}_1 \in \underline{A}_i). \quad (32)$$

In the above special one-dimension case, Eqs. (29) and (30), the setting $X = X_{\alpha l} = X_{\alpha r}$ leads to

$$P_{\alpha l}(A_i) = P_{\alpha r}(A_i)$$
$$= P(X_1 = t \mid x_1, x_2, t \in \mathbf{X}; x_1 \leq t_l \leq x_2), \quad (33)$$

Further properties and computation rules for fuzzy probabilities may be derived from traditional probability theory in conjunction with fuzzy set theory.

Representation of Fuzzy Random Quantities

The fuzzy probability $\tilde{P}(\underline{A}_i)$ may be computed for each arbitrary set $\underline{A}_i \in \mathfrak{S}(\underline{X})$. If – as a special family of sets $\mathfrak{S}(\underline{X})$ – the Borel σ-algebra $\mathfrak{S}_0(\mathbb{R}^n)$ of the $\mathbb{R}^n$ is selected, the concept of the probability distribution function may

be applied to fuzzy random quantities. That is, the system $\mathfrak{S}_0(\mathbb{R}^n)$ of the open sets

$$\underline{A}_{i0} = \{\underline{t} = (t_1, \dots, t_k, \dots, t_n) | \underline{x} = \underline{x}_i; \underline{x}, \underline{t} \in \mathbb{R}^n;$$
$$t_k < x_k; k = 1, \dots, n\} \quad (34)$$

on $\underline{X} = \mathbb{R}^n$ is considered; $\mathfrak{S}_0(\mathbb{R}^n)$ is a Boolean set algebra.

The concept of fuzzy probability according to Sect. "Fuzzy Probability" applied to the sets from Eq. (34) leads to fuzzy probability distribution functions; see Fig. 5. The fuzzy probability distribution function $\tilde{F}(\underline{x})$ of the fuzzy random quantity $\underline{\tilde{X}}$ on $\underline{X} = \mathbb{R}^n$ is obtained as the set of the fuzzy probabilities $\tilde{P}(\underline{A}_{i0})$ with $\underline{A}_{i0}$ according to Eq. (34) for all $\underline{x}_i \in \underline{X}$,

$$\tilde{F}(\underline{x}) = \{\tilde{P}(\underline{A}_{i0}) \,\forall \underline{x}_i \in \underline{X}\} . \quad (35)$$

It is a fuzzy function. Bounds for the functional values $\tilde{F}(\underline{x})$ are specified for each α-level depending on $\underline{x} = \underline{x}_i$ in Eq. (34) and in compliance with Eqs. (19) and (20),

$$F_{\alpha l}\big(\underline{x} = (x_1, \dots, x_n)\big)$$
$$= 1 - \max_{\underline{X}_j \in X_\alpha} P\big(\underline{X}_j = \underline{t} = (t_1, \dots, t_n) | \underline{x}, \underline{t} \in \underline{X} = \mathbb{R}^n,$$
$$\exists t_k \geq x_k, 1 \leq k \leq n\big) , \quad (36)$$

$$F_{\alpha r}(\underline{x} = (x_1, \dots, x_n))$$
$$= \max_{\underline{X}_j \in X_\alpha} P\big(\underline{X}_j = \underline{t} = (t_1, \dots, t_n) | \underline{x}, \underline{t} \in \underline{X} = \mathbb{R}^n,$$
$$t_k < x_k, k = 1, \dots, n\big) . \quad (37)$$

For the determination of $F_{\alpha l}(\underline{x})$ the relationship in Eq. (25) is used. If $F_{\alpha l}(\underline{x})$ and $F_{\alpha r}(\underline{x})$ form closed connected intervals $[F_{\alpha l}(\underline{x}), F_{\alpha r}(\underline{x})]$ – see Sect. "Fuzzy Probability" for the conditions – the functional values $\tilde{F}(\underline{x})$ are determined based on Eq. (21),

$$\tilde{F}(\underline{x}) = \{(F_\alpha(\underline{x}), \mu(F_\alpha(\underline{x}))) | F_\alpha(\underline{x}) = [F_{\alpha l}(\underline{x}), F_{\alpha r}(\underline{x})],$$
$$\mu(F_\alpha(\underline{x})) = \alpha \,\forall \alpha \in (0, 1]\} , \quad (38)$$

In this case, the functional values of the fuzzy probability distribution function $\tilde{F}(\underline{x})$ are continuous and convex fuzzy sets.

In correspondence with Eq. (15), the fuzzy probability distribution function $\tilde{F}(\underline{x})$ of $\underline{\tilde{X}}$ represents the fuzzy set of the probability distribution functions $F_j(\underline{x})$ of all originals $\underline{X}_j$ of $\underline{\tilde{X}}$ with the membership values $\mu(F_j(\underline{x}))$,

$$\tilde{F}(\underline{x}) = \{(F_j(\underline{x}), \mu(F_j(\underline{x}))) | \underline{X}_j \in \underline{\tilde{X}}, \,\mu(F_j(\underline{x}))$$
$$= \mu(\underline{X}_j) \,\forall j\} . \quad (39)$$

Each original $\underline{X}_j$ determines precisely one trajectory $F_j(\underline{x})$ within the bunch $\tilde{F}(\underline{x})$ of weighted functions $F_j(\underline{x}) \in \tilde{F}(\underline{x})$.

In the one-dimensional case and with the restriction to closed random intervals $[X_{\alpha l}, X_{\alpha r}]$ for each α-level, the fuzzy probability distribution function $\tilde{F}(x)$ is determined by

$$F_{\alpha l}(x) = P(X_{\alpha r} = t_r \,|\, x, t_r \in X, \, t_r < x) , \quad (40)$$

$$F_{\alpha r}(x) = P(X_{\alpha l} = t_l \,|\, x, t_l \in X, \, t_l < x) . \quad (41)$$

In correspondence with traditional probability theory, fuzzy probability density functions $\tilde{f}(\underline{t})$ or $\tilde{f}(\underline{x})$ are defined in association with the $\tilde{F}(\underline{x})$; see Fig. 5. The $\tilde{f}(\underline{t})$ or $\tilde{f}(\underline{x})$ are fuzzy functions which – in the continuous case with respect to randomness – are integrable for each original $\underline{X}_j$ of $\underline{\tilde{X}}$ and satisfy the relationship

$$F_j(\underline{x}) = \int_{t_1=-\infty}^{t_1=x_1} \dots \int_{t_k=-\infty}^{t_k=x_k} \dots \int_{t_n=-\infty}^{t_n=x_n} f_j(\underline{t}) \,\mathrm{d}\underline{t}, \quad (42)$$

with $\underline{t} = (t_1, \dots, t_n) \in \underline{X}$. For each original $\underline{X}_j$ the integration of the associated trajectory $f_j(\underline{x}) \in \tilde{f}(\underline{x})$ leads to the respective trajectory $F_j(\underline{x}) \in \tilde{F}(\underline{x})$.

For the description of a fuzzy random quantity $\underline{\tilde{X}}$, parameters in the form of fuzzy quantities $\tilde{p}_t(\underline{\tilde{X}})$ may be used. These fuzzy parameters may represent any type of
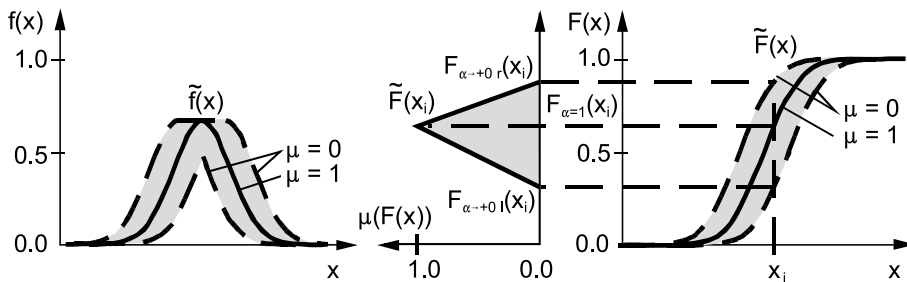
Fuzzy Probability Theory, Figure 5
Fuzzy probability density function $\tilde{f}(\underline{x})$ and fuzzy probability distribution function $\tilde{F}(\underline{x})$ of a continuous fuzzy random variable $\underline{\tilde{X}}$

parameters known from real-valued random quantities, such as moments, weighting factors for different distribution types in a compound distribution, or functional parameters of the distribution functions. The fuzzy parameter $\tilde{p}_t(\underline{\tilde{X}})$ of the fuzzy random quantity $\underline{\tilde{X}}$ is the fuzzy set of the parameter values $p_t(\underline{X}_j)$ of all originals $\underline{X}_j$ with the membership values $\mu(p_t(\underline{X}_j))$,

$$\tilde{p}_t(\underline{\tilde{X}}) = \{(p_t(\underline{X}_j), \mu(p_t(\underline{X}_j))) \mid \underline{X}_j \in \underline{\tilde{X}},$$
$$\mu(p_t(\underline{X}_j)) = \mu(\underline{X}_j) \, \forall j\} \, . \quad (43)$$

For each α-level, bounds are given for the fuzzy parameter $\tilde{p}_t(\underline{\tilde{X}})$ by

$$p_{t,\alpha l}(\underline{\tilde{X}}) = \min_{\underline{X}_j \in \underline{X}_\alpha} [p_t(\underline{X}_j)] \, , \quad (44)$$

$$p_{t,\alpha r}(\underline{\tilde{X}}) = \max_{\underline{X}_j \in \underline{X}_\alpha} [p_t(\underline{X}_j)] \, . \quad (45)$$

If the fuzzy random quantity $\underline{\tilde{X}}$ represents a fuzzy set of continuous real-valued random quantities, if all fuzzy realizations $\underline{\tilde{x}}_i$ of $\underline{\tilde{X}}$ are connected sets, and if the parameter p_t is defined on a continuous scale, then the fuzzy parameter $\tilde{p}_t(\underline{\tilde{X}})$ is determined by its α-level sets

$$p_{t,\alpha}(\underline{\tilde{X}}) = [p_{t,\alpha l}(\underline{\tilde{X}}), p_{t,\alpha r}(\underline{\tilde{X}})] \, , \quad (46)$$

$$\tilde{p}_t(\underline{\tilde{X}}) = \{(p_{t,\alpha}(\underline{\tilde{X}}), \mu(p_{t,\alpha}(\underline{\tilde{X}}))) \mid$$
$$\mu(p_{t,\alpha}(\underline{\tilde{X}})) = \alpha \, \forall \, \alpha \in (0, 1]\} \, , \quad (47)$$

and represents a continuous and convex fuzzy set.

If a fuzzy random quantity $\underline{\tilde{X}}$ is described by more than one fuzzy parameter $\tilde{p}_t(\underline{\tilde{X}})$, interactive dependencies are generally present between the different fuzzy parameters. If this interaction is neglected, a fuzzy random quantity $\underline{\tilde{X}}_{\text{hull}}$ is obtained, which covers the actual fuzzy random quantity $\underline{\tilde{X}}$ completely. That is, for all realizations of $\underline{\tilde{X}}_{\text{hull}}$ and $\underline{\tilde{X}}$ the following holds,

$$\underline{\tilde{x}}_{i\,\text{hull}} \supseteq \underline{\tilde{x}}_i \, \forall i \, . \quad (48)$$

Fuzzy parameters and fuzzy probability distribution functions do not enable a unique reproduction of fuzzy realizations based on the above description. But they are sufficient to compute fuzzy probabilities correctly for any events defined according to Eq. (13).

The presented concept of fuzzy probability can be extended to fuzzy random functions and processes.

Future Directions

Fuzzy probability theory provides a powerful key to solving a broad variety of practical problems that defy an appropriate treatment with traditional probabilistics due to imprecision of the information for model specification. Fuzzy probabilities reflect aleatory uncertainty and epistemic uncertainty of the underlying problem simultaneously and separately and provide extended information and decision aids. These features can be utilized in all application fields of traditional probability theory and beyond. Respective developments can be observed, primarily, in information science and, increasingly, in engineering. Potential for further extensive fruitful applications exists, for example, in psychology, economy, financial sciences, medicine, biology, social sciences, and even in law. In all cases, fuzzy probability theory is not considered as a replacement for traditional probabilistics but as a beneficial supplement for an appropriate model specification according to the available information in each particular case.

The focus of further developments is seen on both theory and applications. Future theoretical developments may pursue a measure theoretical clarification of the embedding of fuzzy probability theory in the framework of imprecise probabilities under the umbrella of generalized information theory. This is associated with the ambition to unify the variety of available fuzzy probabilistic concepts and to eventually formulate a consistent generalized fuzzy probability theory. Another important issue for future research is the mathematical description and treatment of dependencies within the fuzziness of fuzzy random quantities such as non-probabilistic interaction between fuzzy realizations, between fuzzy parameters, and between fuzzy probabilities of certain events. In parallel to theoretical modeling, further effort is worthwhile towards a consistent concept for the statistical evaluation of imprecise data including the analysis of probabilistic and non-probabilistic dependencies of the data.

In view of applications, the further development of fuzzy probabilistic simulation methods is of central importance. This concerns both theory and numerical algorithms for the direct generation of fuzzy random quantities – in a parametric and in a non-parametric fashion. Representations and computational procedures for fuzzy random quantities must be developed with focus on a high numerical efficiency to enable a solution of real-world problems. For a spread into practice it is further essential to elaborate options and potentials for an interpretation and evaluation of fuzzy probabilistic results such as fuzzy mean values or fuzzy failure probabilities. The most promising potentials for a utilization are seen in worst-case investigations in terms of probability, in a sensitivity analysis with respect to non-probabilistic uncertainty, and in decision-making based on mixed probabilistic/non-probabilistic information.

In summary, fuzzy probability theory and its further developments significantly contribute to an improved uncertainty modeling according to reality.

Bibliography

Primary Literature

1. Alefeld G, Herzberger J (1983) Introduction to interval computations. Academic Press, New York
2. Bandemer H (1992) Modelling uncertain data. Akademie-Verlag, Berlin
3. Bandemer H, Gebhardt A (2000) Bayesian fuzzy kriging. Fuzzy Sets Syst 112:405–418
4. Bandemer H, Gottwald S (1995) Fuzzy sets, fuzzy logic fuzzy methods with applications. Wiley, Chichester
5. Bandemer H, Näther W (1992) Fuzzy data analysis. Kluwer, Dordrecht
6. Beer M (2007) Model-free sampling. Struct Saf 29:49–65
7. Berleant D, Zhang J (2004) Representation and problem solving with distribution envelope determination (denv). Reliab Eng Syst Saf 85(1–3):153–168
8. Bernardini A, Tonon F (2004) Aggregation of evidence from random and fuzzy sets. Special Issue of ZAMM. Z Angew Math Mech 84(10–11):700–709
9. Bodjanova S (2000) A generalized histogram. Fuzzy Sets Syst 116:155–166
10. Colubi A, Domínguez-Menchero JS, López-Díaz M, Gil MA (1999) A generalized strong law of large numbers. Probab Theor Relat Fields 14:401–417
11. Colubi A, Domínguez-Menchero JS, López-Díaz M, Ralescu DA (2001) On the formalization of fuzzy random variables. Inf Sci 133:3–6
12. Colubi A, Domínguez-Menchero JS, López-Díaz M, Ralescu DA (2002) A de[0, 1]-representation of random upper semicontinuous functions. Proc Am Math Soc 130:3237–3242
13. Colubi A, Fernández-García C, Gil MA (2002) Simulation of random fuzzy variables: an empirical approach to statistical/probabilistic studies with fuzzy experimental data. IEEE Trans Fuzzy Syst 10:384–390
14. Couso I, Sanchez L (2008) Higher order models for fuzzy random variables. Fuzzy Sets Syst 159:237–258
15. de Cooman G (2002) The society for imprecise probability: theories and applications. http://www.sipta.org
16. Diamond P (1990) Least squares fitting of compact set-valued data. J Math Anal Appl 147:351–362
17. Diamond P, Kloeden PE (1994) Metric spaces of fuzzy sets: theory and applications. World Scientific, Singapore
18. Dubois D, Prade H (1980) Fuzzy sets and systems theory and applications. Academic Press, New York
19. Dubois D, Prade H (1985) A review of fuzzy set aggregation connectives. Inf Sci 36:85–121
20. Dubois D, Prade H (1986) Possibility theory. Plenum Press, New York
21. Fellin W, Lessmann H, Oberguggenberger M, Vieider R (eds) (2005) Analyzing uncertainty in civil engineering. Springer, Berlin
22. Feng Y, Hu L, Shu H (2001) The variance and covariance of fuzzy random variables and their applications. Fuzzy Sets Syst 120(3):487–497
23. Ferson S, Hajagos JG (2004) Arithmetic with uncertain numbers: rigorous and (often) best possible answers. Reliab Eng Syst Saf 85(1–3):135–152
24. Fetz T, Oberguggenberger M (2004) Propagation of uncertainty through multivariate functions in the framework of sets of probability measures. Reliab Eng Syst Saf 85(1–3):73–87
25. Ghanem RG, Spanos PD (1991) Stochastic finite elements: a spectral approach. Springer, New York; Revised edition 2003, Dover Publications, Mineola
26. González-Rodríguez G, Montenegro M, Colubi A, Ángeles Gil M (2006) Bootstrap techniques and fuzzy random variables: Synergy in hypothesis testing with fuzzy data. Fuzzy Sets Syst 157(19):2608–2613
27. Grzegorzewski P (2000) Testing statistical hypotheses with vague data. Fuzzy Sets Syst 112:501–510
28. Hall JW, Lawry J (2004) Generation, combination and extension of random set approximations to coherent lower and upper probabilities. Reliab Eng Syst Saf 85(1–3):89–101
29. Helton JC, Johnson JD, Oberkampf WL (2004) An exploration of alternative approaches to the representation of uncertainty in model predictions. Reliab Eng Syst Saf 85(1–3):39–71
30. Helton JC, Oberkampf WL (eds) (2004) Special issue on alternative representations of epistemic uncertainty. Reliab Eng Syst Saf 85(1–3):1–369
31. Hung W-L (2001) Bootstrap method for some estimators based on fuzzy data. Fuzzy Sets Syst 119:337–341
32. Hwang C-M, Yao J-S (1996) Independent fuzzy random variables and their application. Fuzzy Sets Syst 82:335–350
33. Jang L-C, Kwon J-S (1998) A uniform strong law of large numbers for partial sum processes of fuzzy random variables indexed by sets. Fuzzy Sets Syst 99:97–103
34. Joo SY, Kim YK (2001) Kolmogorovs strong law of large numbers for fuzzy random variables. Fuzzy Sets Syst 120:499–503
35. Kim YK (2002) Measurability for fuzzy valued functions. Fuzzy Sets Syst 129:105–109
36. Klement EP, Puri ML, Ralescu DA (1986) Limit theorems for fuzzy random variables. Proc Royal Soc A Math Phys Eng Sci 407:171–182
37. Klement EP (1991) Fuzzy random variables. Ann Univ Sci Budapest Sect Comp 12:143–149
38. Klir GJ (2006) Uncertainty and information: foundations of generalized information theory. Wiley-Interscience, Hoboken
39. Klir GJ, Folger TA (1988) Fuzzy sets, uncertainty, and information. Prentice Hall, Englewood Cliffs
40. Körner R (1997) Linear models with random fuzzy variables. Phd thesis, Bergakademie Freiberg, Fakultät für Mathematik und Informatik
41. Körner R (1997) On the variance of fuzzy random variables. Fuzzy Sets Syst 92:83–93
42. Körner R, Näther W (1998) Linear regression with random fuzzy variables: extended classical estimates, best linear estimates, least squares estimates. Inf Sci 109:95–118
43. Krätschmer V (2001) A unified approach to fuzzy random variables. Fuzzy Sets Syst 123:1–9
44. Krätschmer V (2002) Limit theorems for fuzzy-random variables. Fuzzy Sets Syst 126:253–263
45. Krätschmer V (2004) Probability theory in fuzzy sample space. Metrika 60:167–189
46. Kruse R, Meyer KD (1987) Statistics with vague data. Reidel, Dordrecht

47. Kwakernaak H (1978) Fuzzy random variables I. definitions and theorems. Inf Sci 15:1–19

48. Kwakernaak H (1979) Fuzzy random variables II. algorithms and examples for the discrete case. Inf Sci 17:253–278

49. Li S, Ogura Y, Kreinovich V (2002) Limit theorems and applications of set valued and fuzzy valued random variables. Kluwer, Dordrecht

50. Lin TY, Yao YY, Zadeh LA (eds) (2002) Data mining, rough sets and granular computing. Physica, Germany

51. López-Díaz M, Gil MA (1998) Reversing the order of integration in iterated expectations of fuzzy random variables, and statistical applications. J Stat Plan Inference 74:11–29

52. Matheron G (1975) Random sets and integral geometry. Wiley, New York

53. Möller B, Graf W, Beer M (2000) Fuzzy structural analysis using alpha-level optimization. Comput Mech 26:547–565

54. Möller B, Beer M (2004) Fuzzy randomness – uncertainty in civil engineering and computational mechanics. Springer, Berlin

55. Möller B, Reuter U (2007) Uncertainty forecasting in engineering. Springer, Berlin

56. Muhanna RL, Mullen RL, Zhang H (2007) Interval finite element as a basis for generalized models of uncertainty in engineering mechanics. J Reliab Comput 13(2):173–194

57. Gil MA, López-Díaz M, Ralescu DA (2006) Overview on the development of fuzzy random variables. Fuzzy Sets Syst 157(19):2546–2557

58. Näther W, Körner R (2002) Statistical modelling, analysis and management of fuzzy data, chapter on the variance of random fuzzy variables. Physica, Heidelberg, pp 25–42

59. Näther W (2006) Regression with fuzzy random data. Comput Stat Data Analysis 51:235–252

60. Näther W, Wünsche A (2007) On the conditional variance of fuzzy random variables. Metrika 65:109–122

61. Oberkampf WL, Helton JC, Sentz K (2001) Mathematical representation of uncertainty. In: AIAA non-deterministic approaches forum, number AIAA 2001–1645. AIAA, Seattle

62. Pedrycz W, Skowron A, Kreinovich V (eds) (2008) Handbook of granular computing. Wiley, New York

63. Puri ML, Ralescu D (1986) Fuzzy random variables. J Math Anal Appl 114:409–422

64. Puri ML, Ralescu DA (1983) Differentials of fuzzy functions. J Math Anal Appl 91:552–558

65. Puri ML, Ralescu DA (1991) Convergence theorem for fuzzy martingales. J Math Anal Appl 160:107–122

66. Rodríguez-Muñiz L, López-Díaz M, Gil MA (2005) Solving influence diagrams with fuzzy chance and value nodes. Eur J Oper Res 167:444–460

67. Samarasooriya VNS, Varshney PK (2000) A fuzzy modeling approach to decision fusion under uncertainty. Fuzzy Sets Syst 114:59–69

68. Schenk CA, Schuëller GI (2005) Uncertainty assessment of large finite element systems. Springer, Berlin

69. Schuëller GI, Spanos PD (eds) (2001) Proc int conf on monte carlo simulation MCS 2000. Swets and Zeitlinger, Monaco

70. Shafer G (1976) A mathematical theory of evidence. Princeton University Press, Princeton

71. Song Q, Leland RP, Chissom BS (1997) Fuzzy stochastic fuzzy time series and its models. Fuzzy Sets Syst 88:333–341

72. Taheri SM, Behboodian J (2001) A bayesian approach to fuzzy hypotheses testing. Fuzzy Sets Syst 123:39–48

73. Terán P (2006) On borel measurability and large deviations for fuzzy random variables. Fuzzy Sets Syst 157(19):2558–2568

74. Terán P (2007) Probabilistic foundations for measurement modelling with fuzzy random variables. Fuzzy Sets Syst 158(9):973–986

75. Viertl R (1996) Statistical methods for non-precise data. CRC Press, Boca Raton

76. Viertl R, Hareter D (2004) Generalized Bayes' theorem for non-precise a-priori distribution. Metrika 59:263–273

77. Viertl R, Trutschnig W (2006) Fuzzy histograms and fuzzy probability distributions. In: Proceedings of the 11th Conference on Information Processing and Management of Uncertainty in Knowledge-based Systems, Editions EDK, Paris, CD-ROM

78. Walley P (1991) Statistical reasoning with imprecise probabilities. Chapman & Hall, London

79. Wang G, Qiao Z (1994) Convergence of sequences of fuzzy random variables and its application. Fuzzy Sets Syst 63:187–199

80. Wang G, Zhang Y (1992) The theory of fuzzy stochastic processes. Fuzzy Sets Syst 51:161–178

81. Weichselberger K (2000) The theory of interval-probability as a unifying concept for uncertainty. Int J Approx Reason 24(2–3):149–170

82. Yager RR (1984) A representation of the probability of fuzzy subsets. Fuzzy Sets Syst 13:273–283

83. Zadeh LA (1965) Fuzzy sets. Inf Control 8:338–353

84. Zadeh LA (1968) Probability measures of fuzzy events. J Math Anal Appl 23:421–427

85. Zhong Q, Yue Z, Guangyuan W (1994) On fuzzy random linear programming. Fuzzy Sets Syst 65(1):31–49

86. Zimmermann HJ (1992) Fuzzy set theory and its applications. Kluwer, Boston

Books and Reviews

Ayyub BM (1998) Uncertainty modeling and analysis in civil engineering. CRC Press, Boston

Blockley DI (1980) The nature of structural design and safety. Ellis Horwood, Chichester

Buckley JJ (2003) Fuzzy Probabilities. Physica/Springer, Heidelberg

Cai K-Y (1996) Introduction to fuzzy reliability. Kluwer, Boston

Chou KC, Yuan J (1993) Fuzzy-bayesian approach to reliability of existing structures. ASCE J Struct Eng 119(11):3276–3290

Elishakoff I (1999) Whys and hows in uncertainty modelling probability, fuzziness and anti-optimization. Springer, New York

Feng Y (2000) Decomposition theorems for fuzzy supermartingales and submartingales. Fuzzy Sets Syst 116:225–235

Gil MA, López-Díaz M (1996) Fundamentals and bayesian analyses of decision problems with fuzzy-valued utilities. Int J Approx Reason 15:203–224

Gil MA, Montenegro M, González-Rodríguez G, Colubi A, Casals MR (2006) Bootstrap approach to the multi-sample test of means with imprecise data. Comput Stat Data Analysis 51:148–162

Grzegorzewski P (2001) Fuzzy sets b defuzzification and randomization. Fuzzy Sets Syst 118:437–446

Grzegorzewski P (2004) Distances between intuitionistic fuzzy sets and/or interval-valued fuzzy sets based on the Hausdorff metric. Fuzzy Sets Syst 148(2):319–328

Hareter D (2004) Time series analysis with non-precise data. In: Wojtkiewicz S, Red-Horse J, Ghanem R (eds) 9th ASCE specialty conference on probabilistic mechanics and structural reliability. Sandia National Laboratories, Albuquerque

Helton JC, Cooke RM, McKay MD, Saltelli A (eds) (2006) Special issue: The fourth international conference on sensitivity analysis of model output – SAMO 2004. Reliab Eng Syst Saf 91:(10–11):1105–1474

Hirota K (1992) An introduction to fuzzy logic applications in intelligent systems. In: Kluwer International Series in Engineering and Computer Science, Chapter probabilistic sets: probabilistic extensions of fuzzy sets, vol 165. Kluwer, Boston, pp 335–354

Klement EP, Puri ML, Ralescu DA (1984) Cybernetics and Systems Research 2, Chapter law of large numbers and central limit theorems for fuzzy random variables. Elsevier, North-Holland, pp 525–529

Kutterer H (2004) Statistical hypothesis tests in case of imprecise data, V Hotine-Marussi Symposium on Mathematical Geodesy. Springer, Berlin, pp 49–56

Li S, Ogura Y (2003) A convergence theorem of fuzzy-valued martingales in the extended hausdorff metric h(inf). Fuzzy Sets Syst 135:391–399

Li S, Ogura Y, Nguyen HT (2001) Gaussian processes and martingales for fuzzy valued random variables with continuous parameter. Inf Sci 133:7–21

Li S, Ogura Y, Proske FN, Puri ML (2003) Central limit theorems for generalized set-valued random variables. J Math Anal Appl 285:250–263

Liu B (2002) Theory and practice of uncertainty programming. Physica, Heidelberg

Liu Y, Qiao Z, Wang G (1997) Fuzzy random reliability of structures based on fuzzy random variables. Fuzzy Sets Syst 86:345–355

Möller B, Graf W, Beer M (2003) Safety assessment of structures in view of fuzzy randomness. Comput Struct 81:1567–1582

Möller B, Liebscher M, Schweizerhof K, Mattern S, Blankenhorn G (2008) Structural collapse simulation under consideration of uncertainty – improvement of numerical efficiency. Comput Struct 86(19–20):1875–1884

Montenegro M, Casals MR, Lubiano MA, Gil MA (2001) Two-sample hypothesis tests of means of a fuzzy random variable. Inf Sci 133:89–100

Montenegro M, Colubi A, Casals MR, Gil MA (2004) Asymptotic and bootstrap techniques for testing the expected value of a fuzzy random variable. Metrika 59:31–49

Montenegro M, González-Rodríguez G, Gil MA, Colubi A, Casals MR (2004) Soft methodology and random information systems, chapter introduction to ANOVA with fuzzy random variables. Springer, Berlin, pp 487–494

Muhanna RL, Mullen RL (eds) (2004) Proceedings of the NSF workshop on reliable engineering computing. Center for Reliable Engineering Computing. Georgia Tech Savannah, Georgia

Muhanna RL, Mullen RL (eds) (2006) NSF workshop on reliable engineering computing. center for reliable engineering computing. Georgia Tech Savannah, Georgia

Negoita VN, Ralescu DA (1987) Simulation, knowledge-based computing and fuzzy-statistics. Van Nostrand, Reinhold, New York

Oberguggenberger M, Schuëller GI, Marti K (eds) (2004) Special issue on application of fuzzy sets and fuzzy logic to engineering problems. ZAMM: Z Angew Math Mech 84(10–11):661–776

Okuda T, Tanaka H, Asai K (1978) A formulation of fuzzy decision problems with fuzzy information using probability measures of fuzzy events. Inf Control 38:135–147

Proske FN, Puri ML (2002) Strong law of large numbers for banach space valued fuzzy random variables. J Theor Probab 15:543–551

Reddy RK, Haldar A (1992) Analysis and management of uncertainty: theory and applications, chapter a random-fuzzy reliability analysis of engineering systems. North-Holland, Amsterdam, pp 319–329

Reddy RK, Haldar A (1992) A random-fuzzy analysis of existing structures. Fuzzy Sets Syst 48:201–210

Ross TJ (2004) Fuzzy logic with engineering applications, 2nd edn. Wiley, Philadelphia

Ross TJ, Booker JM, Parkinson WJ (eds) (2002) Fuzzy logic and probability applications – bridging the gap. SIAM & ASA, Philadelphia

Stojakovic M (1994) Fuzzy random variables, expectation, and martingales. J Math Anal Appl 184:594–606

Terán P (2004) Cones and decomposition of sub- and supermartingales. Fuzzy Sets Syst 147:465–474

Tonon F, Bernardini A (1998) A random set approach to the optimization of uncertain structures. Comput Struct 68(6):583–600

Weichselberger K (2000) The theory of interval-probability as a unifying concept for uncertainty. Int J Approx Reason 24(2–3):149–170

Yukari Y, Masao M (1999) Interval and paired probabilities for treating uncertain events. IEICE Trans Inf Syst E82–D(5):955–961

Zadeh L (1985) Is probability theory sufficient for dealing with uncertainty in ai: A negative view. In: Proceedings of the 1st Annual Conference on Uncertainty in Artificial Intelligence (UAI-85). Elsevier Science, New York, pp 103–116

Zhang Y, Wang G, Su F (1996) The general theory for response analysis of fuzzy stochastic dynamical systems. Fuzzy Sets Syst 83:369–405

Fuzzy Sets Theory, Foundations of

Janusz Kacprzyk
Systems Research Institute,
Polish Academy of Sciences,
Warsaw, Poland

Article Outline

F

Glossary

Fuzzy set A mathematical tool that can formally characterize an imprecise concept. Whereas a conventional set to which elements can either belong or not, elements in a fuzzy set can belong to some extent, from zero, which stands for a full nonbelongingness) to one, which stands for a full belongingness, through all intermediate values.

Fuzzy relation A mathematical tool that can formally characterize that which is imprecisely specified, notably by using natural language, relations between variables, for instance, *similar*, *much greater than*, *almost equal*, etc.

Extension principle Makes it possible to extend relations, algorithms, etc. defined for variables that take on nonfuzzy (e. g. real) values to those that take on fuzzy values.

Linguistic variable, fuzzy conditional statement, compositional rule of inference Make it possible to use variables, which take on linguistic (instead of numeric) values to represent relations between such variables, by using fuzzy conditional statements and use them in inference by using the compositional rule of inference.

Fuzzy event and its probability Make it possible to formally define events which are imprecisely specified, like "high temperature" and calculate their probabilities, for instance the probability of a "high temperature tomorrow".

Fuzzy logic Provides formal means for the representation of, and inference based on imprecisely specified premises and rules of inference; can be understood in different ways, basically as fuzzy logic in a narrow sense, being some type of multivalued logic, and fuzzy logic in a broad sense, being a way to formalize inference based on imprecisely specified premises and rules of inference.

Definition of the Subject

We provide a brief exposition of basic elements of Zadeh's [95] fuzzy sets theory. We discuss basic properties, operations on fuzzy sets, fuzzy relations and their compositions, linguistic variables, the extension principle, fuzzy arithmetic, fuzzy events and their probabilities, fuzzy logic, fuzzy dynamic systems, etc. We also outline Bellman and Zadeh's [8] general approach to decision making in a fuzzy environment which is a point of departure for virtually all fuzzy decision making, optimization, control, etc. models.

Introduction

This paper is meant to briefly expose a novice reader to basic elements of theory of fuzzy sets and fuzzy systems viewed for our purposes as an effective and efficient means and calculus to deal with imprecision in the definition of data, information and knowledge, and to provide tools and techniques for dealing with imprecision therein. Our exposition will be as formal as necessary, of more intuitive and constructive a character, so that fuzzy tools and techniques can be useful for the multidisciplinary audience of this encyclopedia. For the readers requiring or interested in a deeper exposition of fuzzy sets and related concepts, we will recommend many relevant references, mainly books. However, as the number of books and volumes on this topic and its applications in a variety of fields is huge, we will recommend some of them only, mostly those better known ones. For the newest literature entries the readers should consult the most recent catalogs of major scientific publishers who have books and edited volumes on fuzzy sets/logic and their applications.

Our discussion will proceed, on the other hand, in the *pure* fuzzy setting, and we will not discuss possibility theory (which is related to fuzzy sets theory). The reader interested in possibility theory is referred to, e. g., Dubois and Prade [29,30] or their article in this encyclopedia.

We will consecutively discuss the idea of a fuzzy set, basic properties of fuzzy sets, operations on fuzzy sets, some extensions of the basic concept of a fuzzy set, fuzzy relations and their compositions, linguistic variables, fuzzy conditional statements, and the compositional rule of inference, the extension principle, fuzzy arithmetic, fuzzy events and their probabilities, fuzzy logic, fuzzy dynamic systems, etc. We also outline Bellman and Zadeh's [8] general approach to decision making in a fuzzy environment which is a point of departure for virtually all fuzzy decision making, optimization, control, etc. models.

Fuzzy Sets – Basic Definitions and Properties

Fuzzy sets theory, introduced by Zadeh in 1965 [95], is a simple yet very powerful, effective and efficient means to represent and handle imprecise information (of vagueness type) exemplified by *tall* buildings, *large* numbers, etc. We will present fuzzy sets theory as some *calculus of imprecision*, not as a new set theory in the mathematical sense.

The Idea of a Fuzzy Set

From our point of view, the main purpose of a (conventional) set in mathematics is to formally characterize some concept (or property). For instance, the concept of "in-

teger numbers which are greater than or equal three and less than or equal ten" may be uniquely represented just by showing all integer numbers that satisfy this condition; that is, given by the following set: $\{x \in I : 3 \le x \le 10\} = \{3, 4, 5, 6, 7, 8, 9, 10\}$ where I is the set of integers. Notice that we need to specify first a *universe of discourse* (universe, universal set, referential, reference set, etc.) that contains all those elements which are relevant for the particular concept as, e. g., the set of integers I in our example.

A conventional set, say A, may be equated with its *characteristic function* defined as

$$\varphi_A : X \longrightarrow \{0, 1\} \tag{1}$$

which associates with each element x of a universe of discourse $X = \{x\}$ a number $\varphi(x) \in \{0, 1\}$ such that: $\varphi_A(x) = 0$ means that $x \in X$ does not belong to the set A, and $\varphi_A(x) = 1$ means that x belongs to the set A.

Therefore, for the set verbally defined as integer numbers which are greater than or equal three and less than or equal ten, its equivalent set $A = \{3, 4, 5, 6, 7, 8, 9, 10\}$, listing all the respective integer numbers, may be represented by its characteristic function

$$\varphi_A(x) = \begin{cases} 1 & \text{for } x \in \{3, 4, 5, 6, 7, 8, 9, 10\} \\ 0 & \text{otherwise} . \end{cases}$$

Notice that in a conventional set there is a clear-cut differentiation between elements belonging to the set and not, i. e. the transition from the belongingness to nonbelongingness is clear-cut and abrupt.

However, it is easy to notice that a serious difficulty arises when we try to formalize by means of a set vague concepts which are commonly encountered in everyday discourse and widely used by humans as, e. g., the statement "integer numbers which are *more or less* equal to six." Evidently, the (conventional) set cannot be used to adequately characterize such an imprecise concept because an abrupt and clear-cut differentiation between the elements belonging and not belonging to the set is artificial here.

This has led Zadeh [95] to the idea of a *fuzzy set* which is a class of objects with unsharp boundaries, i. e. in which the transition from the belongingness to nonbelongingness is not abrupt; thus, elements of a fuzzy set may belong to it to *partial degrees*, from the full belongingness to the full nonbelongingness through all intermediate values. Notice that this is presumably the most natural and simple way to formally define the imprecision of meaning.

We should therefore start again with a *universe of discourse* (universe, universal set, referential, reference set, etc.) containing all elements relevant for the (imprecise)

concept we wish to formally represent. Then, the characteristic function $\varphi : X \longrightarrow \{0, 1\}$ is replaced by a *membership function* defined as

$$\mu_A : X \longrightarrow [0, 1] \tag{2}$$

such that $\mu_A(x) \in [0, 1]$ is the degree to which an element $x \in X$ belongs to the fuzzy set A: From $\mu_A(x) = 0$ for the full nonbelongingness to $\mu_A(x) = 1$ for the full belongingness, through all intermediate $(0 < \mu_A(x) < 1)$ values.

Now, if we consider as an example the concept of integer numbers which are *more or less six*. Then $x = 6$ certainly belongs to this set so that $\mu_A(6) = 1$, the numbers five and seven belong to this set *almost surely* so that $\mu_A(5)$ and $\mu_A(7)$ are very close to one, and the more a number differs from six, the less its $\mu_A(.)$. Finally, the numbers below one and above ten do not belong to this set, so that their $\mu_A(.) = 0$. This may be sketched as in Fig. 1 though we should bear in mind that although in our example the membership function is evidently defined for the integer numbers (x's) only, it is depicted in a continuous form to be more illustrative.

In practice the membership function is usually assumed to be piecewise linear as shown in Fig. 2 (for the same fuzzy set as in Fig. 1, i. e. the fuzzy set integer numbers which are *more or less* six). To specify the membership function we then need four numbers only: a, b, c, and d as, e. g., $a = 2$, $b = 5$, $c = 7$, and $d = 10$ in Fig. 2.

Notice that the particular form of a membership function is *subjective* as opposed to an *objective* form of a characteristic function. However, this may be viewed as quite natural as the underlying concepts are subjective indeed as, e. g., the set of integer numbers *more or less* six depend on an individual opinion. Unfortunately, this inherent subjectivity of the membership function may lead to some problems in many formal models in which users would rather have a limit to the scope of subjectivity. We will comment on this issue later on.
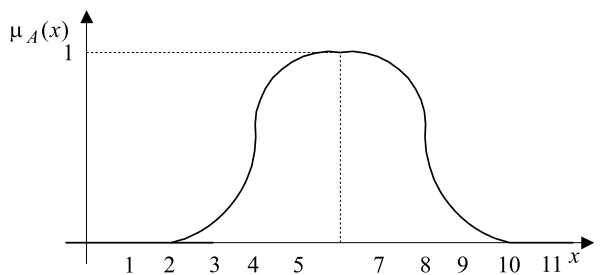
Fuzzy Sets Theory, Foundations of, Figure 1
Membership function of a fuzzy set, integer numbers which are *more or less* six
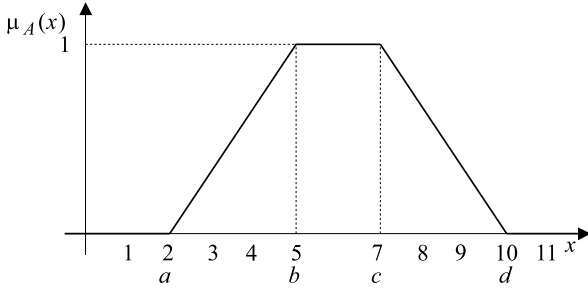
Fuzzy Sets Theory, Foundations of, Figure 2
Membership function of a fuzzy set, integer numbers which are
more or less **six**

We will now define formally a fuzzy set in a form that is very often used.

A *fuzzy set A* in a universe of discourse $X = \{x\}$, written A in X, is defined as a set of pairs

$$A = \{(\mu_A(x), x)\} \tag{3}$$

where $\mu_A: X \longrightarrow [0, 1]$ is the *membership function* of A and $\mu_A(x) \in [0, 1]$ is the *grade of membership* (or a *membership grade*) of an element $x \in X$ in a fuzzy set A.

Needless to say that our definition of a fuzzy set (3) is clearly equivalent to the definition of the membership function (2) because a function may be represented by a set of pairs *argument–value of the function for this argument*. For our purposes, however, the definition (3) is more *set-theoretic-like* which will often be more convenient.

So, in this paper we will practically equate fuzzy sets with their membership functions saying, e. g., a fuzzy set, $\mu_A(x)$, and also very often we will equate fuzzy sets with their labels saying, e. g., a fuzzy set, *large numbers*, with the understanding that the label *large numbers* is equivalent to the fuzzy set mentioned, written A = *large numbers*. However, we will use the notation $\mu_A(x)$ for the membership function of a fuzzy set A in X, and not an abbreviated notation $A(x)$ as in some more technical papers, to be consistent with our *more-set-theoretic-like* convention.

For practical reasons, it is very often assumed (also in this paper) that all the universes of discourse are finite as, e. g., $X = \{x_1, \ldots, x_n\}$. In such a case the pair $\{(\mu_A(x), x)\}$ will be denoted by $\mu_A(x)/x$ which is called a *fuzzy singleton*.

Then, the fuzzy set A in X will be written as

$$A = \{(\mu_A(x), x)\} = \{\mu_A(x)/x\}$$
$$= \mu_A(x_1)/x_1 + \ldots + \mu_A(x_n)/x_n = \sum_{i=1}^{n} \mu_A(x_i)/x_i , \tag{4}$$

where $+$ and $\sum$ are meant in the set-theoretic sense. By convention, the pairs $\mu_A(x)/x$ with $\mu_A(x) = 0$ are omitted here.

A conventional (nonfuzzy) set may obviously be written in the fuzzy sets notation introduced above, for instance the (non-fuzzy) set, integer numbers greater than or equal three and less than or equal ten, may be written as

$$A = 1/3 + 1/4 + 1/5 + 1/6 + 1/7 + 1/8 + 1/9 + 1/10 .$$

The family of all fuzzy sets defined in X is denoted by $\mathcal{A}$; it includes evidently also the empty fuzzy set to be defined by (9), i. e. $A = \emptyset$ such that $\mu_A(x) = 0$, for each $x \in X$, and the whole universe of discourse X written as $X = 1/x_1 + \ldots + 1/x_n$.

The concept of a fuzzy set as defined above has been the point of departure for the *theory of fuzzy sets* (or *fuzzy sets theory*) which will be briefly sketched below. We will again follow a more intuitive and less formal presentation, which is better suited for this encyclopedia.

Some Extensions of the Concept of Zadeh's Fuzzy Set

The concept of Zadeh's [95] fuzzy set introduced in the previous section is the by far the simplest and most natural way to *fuzzify* the concept of a (conventional) set, and clearly provides what we mean to represent and handle imprecision. However, its underlying elements are the most straightforward possible. This concerns above all the membership function. Therefore, it is quite natural that some extensions have been presented to this basic concept. We will just briefly mention some of them.

First, it is quite easy to notice that though the definition of a fuzzy set by the membership function of the type $\mu_A: X \longrightarrow [0, 1]$ is the simplest and most straightforward one, allowing for a gradual transition from the belongingness and nonbelongingness, it can readily be extended. The same role is namely played by a generalized definition by a membership function of the type

$$\mu_A: X \longrightarrow L , \tag{5}$$

where L is some (partially) ordered set as, e. g., a lattice.

This obvious, but powerful extension was introduced by Goguen [37] as an *L-fuzzy set*, where l stands for a lattice. Notice that by using a lattice as the set of values of the membership function we can accommodate situations i which we can encounter elements of the universe of discourse which are not comparable.

Another quite and obvious extension, already mentioned but not developed by Zadeh [95,101] is the concept of a *type 2 fuzzy set*. The rationale behind this concept is obvious. One can easily imagine that the values of

grades of membership of the particular elements of a universe of discourse are fuzzy sets themselves. And further, these fuzzy sets may have grades of membership which are type 2 fuzzy sets, which leads to type 3 fuzzy sets, and one can continue arriving at type n fuzzy sets.

The next, natural extension is that instead of assuming that the degrees of membership are real numbers from the unit interval, one can go a step further and replace these real numbers from $[0, 1]$ by intervals with endpoints belonging to the unit interval. This leads to *interval valued fuzzy sets* which are attributed to Dubois and Gorzałczany (cf. Klir and Yuan [53]). Notice that by using intervals as values of degrees of membership we significantly increase our ability to represent imprecision.

A more radical extension to the concept of Zade's fuzzy set is the so-called *intuitionistic fuzzy set* introduced by Atanassov [1,2].

An intuitionistic fuzzy set A' in a universe of discourse X is defined as

$$A' = \{\langle x, \mu_{A'}(x), \nu_A(x)\rangle | x \in X\} \tag{6}$$

where:

- The degree of membership is

  $$\mu_{A'}: X \to [0, 1],$$

- the degree of non-membership is

  $$\nu_{A'}: X \to [0, 1],$$

- and the condition holds

  $$0 \le \mu_{A'}(x) + \nu_{A'}(x) \le 1; \quad \text{for each } x \in X.$$

Obviously, each (conventional) fuzzy set A in X corresponds to the following intuitionistic fuzzy set A' in X:

$$A = \{\langle x, \mu_{A'}(x), 1 - \mu_{A'}(x)\rangle | x \in X\} \tag{7}$$

For each intuitionistic fuzzy set A' in X, we call

$$\pi_{A'}(x) = 1 - \mu_{A'}(x) - \nu_{A'}(x); \quad \text{for each } x \in X \tag{8}$$

the intuitionistic fuzzy index (or a hesitation margin) of x in A'. The intuitionistic fuzzy index expresses a lack of knowledge of whether an element $x \in X$ belongs to an intuitionistic fuzzy set A' or not.

Notice that the concept of an intuitionistic fuzzy set is a substantial departure from the concept of a (conventional) fuzzy set as it assumes that the degrees of membership and non-membership do not sum up to one, as it is the case in virtually all traditional set theories and their

extensions. For more information, we refer the reader to Atanassov's [3] book.

We will not use these extensions in this short introductory article, and the interested readers are referred to the source literature cited.

Basic Definition and Properties Related to Fuzzy Sets

We will now provide a brief account of basic definitions and properties related to fuzzy sets. We illustrate them with simple examples.

A fuzzy set A is said to be *empty*, written $A = \emptyset$, if and only if

$$\mu_A(x) = 0, \quad \text{for each } x \in X \tag{9}$$

and since we omit the pairs $0/x$, an empty fuzzy set is really void in the notation (4) as there are no singletons in the right-hand side.

Two fuzzy sets A and B defined in the same universe of discourse X are said to be *equal*, written $A = B$, if and only if

$$\mu_A(x) = \mu_B(x), \quad \text{for each } x \in X \tag{10}$$

Example 1 Suppose that $X = \{1, 2, 3\}$ and

$$A = 0.1/1 + 0.5/2 + 1/3$$
$$B = 0.2/1 + 0.5/2 + 1/3$$
$$C = 0.1/1 + 0.5/2 + 1/3$$

then $A = C$ but $A \ne C$ and $B \ne C$.

It is easy to see that this classic definition of the equality of two fuzzy sets by (10) is rigid and clear-cut, contradicting in a sense our intuitive feeling that the equality of fuzzy sets should be *softer*, and not abrupt, i. e. should rather be to some degree, from zero to one. We will show below one of possible definitions of such an equality to a degree.

Two fuzzy sets A and B defined in X are said to be *equal to a degree* $e(A, B) \in [0, 1]$, written $A =_e B$, and the degree of equality $e(A, B)$ may be defined in many ways exemplified by those given below (cf. Bandler and Kohout [6]).

First, to simplify, we denote:

Case 1: $A = B$ in the sense of (10);

Case 2: $A \ne B$ in the sense of (10) and $T = \{x \in X: \mu_A(x) \ne \mu_B(x)\}$;

Case 3: $A \ne B$ in the sense of (10) and there exists an $x \in X$ such that

$$\mu_A(x) = 0 \quad \text{and} \quad \mu_B(x) \ne 0$$
$$\text{or} \quad \mu_A(x) \ne 0 \quad \text{and} \quad \mu_B(x) = 0.$$

Case 4: $A \neq B$ in the sense of (10) and there exists an $x \in X$ such that

$$\mu_A(x) = 0 \quad \text{and} \quad \mu_B(x) = 1$$
$$\text{or} \quad \mu_A(x) = 1 \quad \text{and} \quad \mu_B(x) = 0 .$$

Now, the following degrees of equality of two fuzzy sets, A and B, may be defined:

$$e_1(A, B) = \begin{cases} 1 & \text{for case 1} \\ \bigwedge_{x \in T}[\mu_A(x) \wedge \mu_B(x)] & \text{for case 2} \\ 0 & \text{for case 3} \end{cases} \quad (11)$$

$$e_2(A, B) = \begin{cases} 1 & \text{for case 1} \\ \bigwedge_{x \in T}[\mu_A(x)/\mu_B(x) \\ -\mu_B(x)/\mu_A(x)] & \text{for case 2} \\ 0 & \text{for case 3} \end{cases} \quad (12)$$

$$e_3(A, B) = \begin{cases} 1 & \text{for case 1} \\ 1 - \max_{x \in X} | \mu_A(x) \\ -\mu_B(x) | & \text{for case 2} \\ 0 & \text{for case 4} \end{cases} \quad (13)$$

$$e_4(A, B) = \begin{cases} 1 & \text{for case 1} \\ \max_{x \in X}\{[(1 - \mu_A(x)] \\ \wedge[\mu_A(x) \vee (1 - \mu_B(x))]\} & \text{for case 2} \\ 0 & \text{for case 4} . \end{cases} \quad (14)$$

Now we will proceed to the second basic concept of the containment between two fuzzy sets.

A fuzzy set A defined in X is said to be *contained in* or, alternatively, is said to be a *subset of* a fuzzy set B in X, written $A \subseteq B$, if and only if

$$\mu_A(x) \leq \mu_B(x) , \quad \text{for each } x \in X . \quad (15)$$

Example 2 Suppose that $X = \{1, 2, 3\}$ and

$$A = 0.1/1 + 0.5/2 + 1/4$$
$$B = 0.1/1 + 0.4/2 + 0.9/3$$
$$C = 0.1/1 + 0.6/2 + 1/3 ,$$

then only $B \subseteq A$.

This traditional definition of containment is clearly rigid and clear-cut, and hence there have been proposed many other *softer* definitions in which a *degree of containment*, $c(A, B) \in [0, 1]$, has been employed. Once again, Bandler and Kohout's [6] definitions can be mentioned here, and these basically follow the line of reasoning analogous to that behind the degree of equality, i.e. (11)–

(14). The above definitions of the degree of equality and containment are popular but not the only possible ones, some remarks can be found in the books by Dubois and Prade [28]!, Klir and Folger [51], Klir and Yuan [53].

Let us proceed now to some further relevant foundational notions.

A fuzzy set A defined in X is said to be *normal* if and only if

$$\max_{x \in X} \mu_A(x) = 1 \quad (16)$$

i. e. when the membership function takes on the value of one for at least one argument. Otherwise, the fuzzy set is said to be *subnormal*.

Example 3 If $X = \{1, 2, 3\}$, $A = 0.1/1 + 0.5/2 + 1/3$ and $B = 0.1/1 + 0.6/2 + 0.9/3$, then A is normal and B is subnormal.

Normally, it is desirable to work with normal fuzzy sets since they may provide for some sort of *context-free* comparability, or a common ground or denominator. However, in many instances we obtain in the course of algorithms or procedures subnormal fuzzy sets. They are then often normalized although, unfortunately, the normalization is not a straightforward solution and should be applied with care after some consideration.

We have now some important concepts of nonfuzzy sets associated with a fuzzy set.

The *support* of a fuzzy set A in X, written supp$A$, is the following (nonfuzzy) set

$$\text{supp}A = \{x \in X: \mu_A(x) > 0\} \quad (17)$$

and, evidently, $\emptyset \subseteq \text{supp}A \subseteq X$.

Example 4 If $X = \{1, 2, \ldots, 7\}$ and $A = 0.1/3 + 0.5/4 + 0.8/5 + 1/6$, then supp$A = \{3, 4, 5, 6\} \subset \{1, 2, \ldots, 7\}$.

The *α-cut*, or *α-level set*, of a fuzzy set A in X, written A_α, is defined as the following (nonfuzzy) set

$$A_\alpha = \{x \in X: \mu_A(x) \geq \alpha\}, \quad \text{for each } \alpha \in (0, 1] \quad (18)$$

and if $\geq$ in (18) is replaced by $>$, then we have the *strong α-cut*, or *strong α-level set*, of a fuzzy set A in X. In principle, we will use the α-cuts given by (18) if not otherwise specified.

Example 5 If $X = \{1, 2, 3, 4\}$ and $A = 0.1/1 + 0.5/2 + 0.8/3 + 1/4$, then we obtain the following α-cuts

$$A_{0.1} = \{1, 2, 3, 4\} \quad A_{0.5} = \{2, 3, 4\}$$
$$A_{0.8} = \{3, 4\} \quad A_1 = \{4\} .$$

The α-cuts have many interesting and relevant properties, and among them one can mention the following one

$$\alpha_1 \leq \alpha_2 \iff A_{\alpha_1} \subseteq A_{\alpha_2} . \tag{19}$$

The α-cuts play an extremely relevant role in both formal analysis and applications as they make it possible to uniquely replace a fuzzy set by a sequence of nonfuzzy sets. We will widely use them in the sequel, and the interested reader is referred for details and properties to any book on fuzzy sets theory as, e. g., Dubois and Prade [28], Klir and Folger [51] or Klir and Yuan [53].

The following theorem, called the *representation theorem* (cf. Negoita and Ralescu [67]), is very relevant both in theoretical analysis and applications.

Theorem 1 *Each fuzzy set A in X can be represented as*

$$A = \sum_{\alpha \in (0,1]} \alpha A_{\alpha} , \tag{20}$$

where A_{α} is an α-cut of A defined as (19), Σ is in the set-theoretic sense, and αA_{α} denotes the fuzzy set whose degrees of membership are

$$\mu_{\alpha A_{\alpha}}(x) = \begin{cases} \alpha & for \quad x \in A_{\alpha} \\ 0 & otherwise . \end{cases} \tag{21}$$

The expression (20) is also called the *resolution identity*.

Example 6 Let $X = \{1, 2, \ldots, 10\}$, and $A = 0.1/2 + 0.3/3 + 0.6/4 + 0.8/5 + 1/6 + 0.7/7 + 0.4/8 + 0.2/9$.
Then:

$$\begin{aligned} A = \sum_{\alpha \in (0,1]} \alpha A_{\alpha} &= 0.1(1/2 + 1/3 + 1/4 + 1/5 + 1/6 \\ &+ 1/7 + 1/8 + 1/9) + 0.3(1/3 + 1/4 + 1/5 \\ &+ 1/6 + 1/7 + 1/8 + 1/9) + 0.6(1/4 + 1/5 \\ &+ 1/6 + 1/7) + 0.7(1/5 + 1/6 + 1/7) \\ &+ 0.8(1/5 + 1/6) + 1(1/6) = 0.1/2 + 0.3/3 \\ &+ 0.6/4 + 0.8/5 + 1/6 + 0.7/7 + 0.4/8 + 0.2/9 . \end{aligned}$$

Notice that the very essence of the presentation theorem is that each fuzzy sets can be uniquely represented by a set of its α-cuts.

An important issue, both in theory and application, is to be able to define the *cardinality* of a fuzzy set, i. e. to define how many elements it contains. Unfortunately, this is a difficult problem, and the definitions proposed have been criticized. We will discuss below two of them which are presumably the most widely used.

A *nonfuzzy cardinality* of a fuzzy set $A = \mu_A(x_1)/x_1 + \ldots + \mu_A(x_n)/x_n$, the so-called *sigma-count*, denoted $\Sigma\mathrm{Count}(A)$, is defined as (cf. Zadeh [100,101])

$$\Sigma\mathrm{Count}(A) = \sum_{i=1}^{n} \mu_A(x_i) . \tag{22}$$

Example 7 If $A = 1/x_1 + 0.8/x_2 + 0.6/x_3 + 0.2/x_4 + 0/x_5$, then

$$\Sigma\mathrm{Count}(A) = 1 + 0.8 + 0.6 + 0.2 = 2.6 .$$

The $\Sigma\mathrm{Count}$ is very simple, and is hence widely used. However, an immediate objection may be that the set is fuzzy but its cardinality is not. A solution in this respect, a *fuzzy cardinality*, was proposed by Zadeh [101], and it is shown below. Unfortunately, it is more complicated than a nonfuzzy cardinality defined by (22).

Let A be a fuzzy set defined in X, and A_{α}, for each $\alpha \in (0, 1]$, its α-cuts defined by (19). First, Zadeh [101] introduces the $FG\mathrm{Count}(A)$ as the fuzzy integer defined as

$$FG\mathrm{Count}(A) = \{1/0\} \sum_{\alpha \in (0,1]} \alpha/\mathrm{card}(A_{\alpha}) \tag{23}$$

where $\sum$ is in the set-theoretic [cf. (4)], $\mathrm{card}(A_{\alpha})$ is the usual number of elements in A_{α}, and $1/0$ means the integer number 0.

Equivalently, if A is defined in X such that $\mathrm{card}(X) = n$, then for each non-negative integer $i = 0, 1, \ldots, n$, we denote

$$FG\mathrm{Count}(A)_i = \sum_{\alpha \in (0,1]} \{\alpha : \mathrm{card}(A_{\alpha}) \geq i\} . \tag{24}$$

Semantically, $FG\mathrm{Count}(A)_i$ is the truth of the proposition *A contains at least i elements*.

Next, Zadeh [101] introduces the $FL\mathrm{Count}(A)$ which is defined as the truth of the proposition, "*A* contains at most *i* elements", i. e.

$$FL\mathrm{Count}(A) = \neg[FG\mathrm{Count}(A)] - 1 , \tag{25}$$

where $\neg[.]$ is the complement to be defined by (32), and $-$ is the subtraction in the sense of fuzzy numbers (64).

Similarly as in (24), if A is defined in $X = \{1, 2, \ldots, n\}$, then we can denote

$$FL\mathrm{Count}(A)_i = \sup_{\alpha \in (0,1]} \{\alpha : \mathrm{card}(A_{\alpha}) \geq n - i\} . \tag{26}$$

Notice that

$$FG\mathrm{Count}(A)_i = 1 - FL\mathrm{Count}(A)_{i+1} ,$$
$$\text{for} \quad i = 1, 2, \ldots, n \tag{27}$$

Finally, Zadeh [101] introduces the $FECount(A)$ as

$$FECount(A) = FGCount(A) \cap FLCount(A) \qquad (28)$$

or, similarly as above,

$$FECount(A)_i = FGCount(A)_i \wedge FLCount(A)_i \, ,$$
$$i = 1, 2, \ldots, n \quad (29)$$

where $\cap$ and $\wedge$ denote the intersection of two fuzzy sets and the minimum operations, respectively, as in (34).

Example 8   For the same fuzzy set as in Example 7, i. e. $A = 1/x_1 + 0.8/x_2 + 0.6/x_3 + 0.2/x_4 + 0/x_5$, we obtain

$$FGCount(A) = 1/0 + 1/1 + 0.8/2 + 0.6/3 + 0.2/4$$
$$+ \, 0/5$$
$$FLCount(A) = 0/0 + 0.2/1 + 0.4/2 + 0.6/3$$
$$+ \, 0.2/4 + 0/5$$
$$FECount(A) = 0/0 + 0.2/1 + 0.4/2 + 0.6/3$$
$$+ \, 0.2/4 + 0/5 \, .$$

The above classical definitions of the cardinality of a fuzzy set are widely employed, in particular the nonfuzzy cardinality $\Sigma Count$. However, the problem of how to define the cardinality of a fuzzy set is conceptually difficult, and the best source are here Wygalak's [89,90] books.

An important issue, which is widely used in applications, is a *distance* between two fuzzy sets. In practice, normalized distances are clearly more interesting. In the literature [42], and in this book too, the following two basic definitions are used.

Suppose that we have two fuzzy sets, A and B, both defined in $X = \{x_1, \ldots, x_n\}$. Then, we have the following two basic (normalized) distances:

- The *normalized linear* (Hamming) *distance* between A and B in X defined as

$$l(A, B) = \frac{1}{n} \sum_{i=1}^{n} | \mu_A(x_i) - \mu_B(x_i) | \, . \qquad (30)$$

- The *normalized quadratic* (Euclidean) *distance* between A and B in X defined as

$$q(A, B) = \sqrt{\frac{1}{n} \sum_{i=1}^{n} [\mu_A(x_i) - \mu_B(x_i)]^2} \qquad (31)$$

Example 9   If $X = \{1, 2, \ldots, 7\}$, $A = 0.7/1 + 0.2/2 + 0.6/4 + 0.5/5 + 1/6$ and $B = 0.2/1 + 0.6/4 + 0.8/5 + 1/7$, then:

$$l(A, B) = 0.37 \qquad q(A, B) = 0.49 \, .$$

Now we will proceed to the basic set-theoretic and algebraic operations on fuzzy sets. They are clearly crucial for both theoretical analysis and applications.

Basic Operations on Fuzzy Sets

Similarly as in the conventional (nonfuzzy) set theory, the basic operations in fuzzy set theory are also the complement, intersection and union which will be defined below.

The *complement* of a fuzzy set A in X, written $\neg A$, is defined as

$$\mu_{\neg A}(x) = 1 - \mu_A(x) \, , \quad \text{for each } x \in X \qquad (32)$$

and the complement corresponds to the negation, *not*.

Example 10   If $X = \{1, 2, 3\}$ and $A = 0.1/1 + 0.7/2 + 1/3$, then $\neg A = 0.9/1 + 0.3/2$.

The idea of the complement can be portrayed as in Fig. 3.

This definition is the most widely used due to its simplicity and some important mathematical properties. Sometimes different definitions can be justified and useful in specific contexts as, for instance when $X = [0, 1]$, we have the following

$$\mu_{\neg A}(x) = \mu_A(1 - x) \, , \quad \text{for each } x \in [0, 1] \, . \qquad (33)$$

The *intersection* of two fuzzy sets A and B in X, written $A \cap B$, is defined as

$$\mu_{A \cap B}(x) = \mu_A(x) \wedge \mu_B(x) \, , \quad \text{for each } x \in X \qquad (34)$$

where $\wedge$ is the minimum operation, i. e. $a \wedge b = \min(a, b)$; the intersection of two fuzzy sets corresponds to the connective *and*

Example 11   If $X = \{1, 2, 3, 4\}$, and $A = 0.2/1 + 0.5/2 + 0.8/3 + 1/4$ and $B = 1/1 + 0.8/2 + 0.5/3 + 0.2/4$, then we obtain by (34) $A \cap B = 0.2/1 + 0.5/2 + 0.5/3 + 0.2/4$.
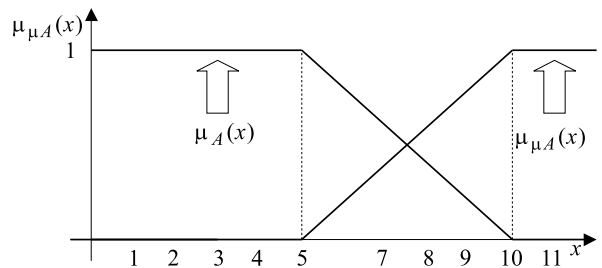
Fuzzy Sets Theory, Foundations of, Figure 3
The complement of a fuzzy set

Fuzzy Sets Theory, Foundations of, Figure 4
Membership function of a fuzzy set, *integer numbers which are more or less six*

Fuzzy Sets Theory, Foundations of, Figure 5
The union of two fuzzy sets

The intersection can be illustrated as in Fig. 4 where $\mu_{A \cap B}(x)$ is shown in bold line.

The *union* of two fuzzy sets A and B in X, written $A + B$, is defined as

$$\mu_{A+B}(x) = \mu_A(x) \vee \mu_B(x), \quad \text{for each } x \in X \quad (35)$$
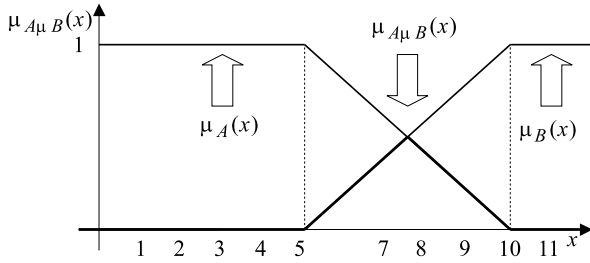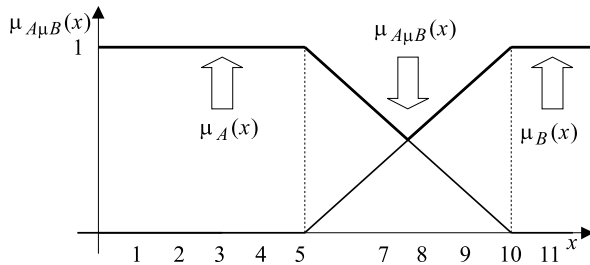
where $\vee$ is the maximum operation, i. e. $a \vee b = \max(a, b)$; the union of two fuzzy sets corresponds to the connective "or".

Example 12 If $X = \{1, 2, 3, 4\}$, and $A = 0.2/1 + 0.5/2 + 0.8/3 + 1/4$ and $B = 1/1 + 0.8/2 + 0.5/3 + 0.2/4$, then $A + B = 1/1 + 0.8/2 + 0.8/3 + 1/4$.

The union can be portrayed as in Fig. 5 in which $\mu_{A+B}(x)$ is shown in bold line.

The above definitions of the basic operations are widely used, and justified. However, other definitions are often employed too. In particular, for the intersection and union the *t*-norms and *s*-norms are popular.

A *t-norm* is defined as

$$t: [0, 1] \times [0, 1] \longrightarrow [0, 1] \quad (36)$$

such that, for each $a, b, c \in [0, 1]$:

1. It has 1 as the unit element, i. e. $t(a, 1) = a$,
2. It is monotone, i. e. $a \leq b \Longrightarrow t(a, c) \leq t(b, c)$,

3. It is commutative, i. e. $t(a, b) = t(b, a)$, and
4. It is associative, i. e. $t[a, t(b, c)] = t[t(a, b), c]$.

Some more relevant examples of *t*-norms are:

- The minimum (which is the most widely used)

$$t(a, b) = a \wedge b = \min(a, b), \quad (37)$$

- the algebraic product

$$t(a, b) = a \cdot b, \quad (38)$$

- the Łukasiewicz *t*-norm

$$t(a, b) = \max(0, a + b - 1) \quad (39)$$

and notice that we have written above both $t(a, b)$ and atb.

An *s-norm* (or a *t-conform*) is defined as

$$s: [0, 1,] \times [0, 1] \longrightarrow [0, 1] \quad (40)$$

such that, for each $a, b, c \in [0, 1]$:

1. It has 0 as the unit element, i. e. $s(a, 0) = a$,
2. it is monotone, i. e. $a \leq b \Longrightarrow s(a, c) \leq s(b, c)$,
3. it is commutative, i. e. $s(a, b) = s(b, a)$, and
4. it is associative, i. e. $s[a, s(b, c)] = s[s(a, b), c]$.

Some more relevant examples of *s*-norms are:

- The maximum (which is the most widely used)

$$s(a, b) = a \vee b = \max(a, b), \quad (41)$$

- the probabilistic product

$$s(a, b) = a + b - ab, \quad (42)$$

- the Łukasiewicz *s*-norm

$$s(a, b) = \min(a + b, 1). \quad (43)$$

Notice that a *t*-norms is *dual* to an *s*-norms in that $s(a, b) = 1 - t(1 - a, 1 - b)$.

The *t*-norms and *s*-norms are very important for fuzzy sets theory, and among a multitude of works dealing with them, the most complete reference is provided by Klement, Mesiar and Pap's [49] book.

Notice that while defining the basic concepts and operations on fuzzy sets we have always referred to their linguistic sense. This is very important because without such a linguistic connection it would have been very difficult to provide semantics of both the very concept of a fuzzy set and the operations. This is very relevant and has great implications for the development of fuzzy sets related tools and techniques as we will see later.

As to some other operations on fuzzy sets that may be of use in this book, one should also mention the following ones.

The *product of a scalar $a \in R$ and a fuzzy set A* in X, written aA, is defined as

$$\mu_{aA}(x) = a\mu_A(x), \quad \text{for each } x \in X \tag{44}$$

where, by necessity, $0 \le a \le 1/\mu_A(x)$, for each $x \in X$.

The *kth power of a fuzzy set A* in X, written A^k, is defined as

$$\mu_{A^k}(x) = [\mu_A(x)]^k, \quad \text{for each } x \in X, \tag{45}$$

where $k \in R$ and, evidently, $0 \le [\mu_A(x)]^k \le 1$.

An important issue is the *adequacy* of the operations on fuzzy sets, i. e. whether they do reflect the real human perception of their essence, i. e. whether they really reflect the semantics of "not," "and," "or," etc. Diverse approaches have been used to find and justify a particular definition. These approaches may be classified as:

- *Intuitive* as the original Zadeh's [95,97] works in which it is shown by a rational argument that the operations defined are proper,
- *axiomatic* whose line of reasoning is to assume some set of plausible conditions to be fulfilled, and then to show using some analytic tools that definitions assumed are the only possible ones; this may be exemplified by Bellman and Giertz [7],
- *experimental* whose essence is to devise some psychological tests for a group of certain individuals, and then use the responses to find which operation is best justified; this may be exemplified by Zimmermann and Zysno [111].

Now we will present the concept of a fuzzy relation which is, as its nonfuzzy counterpart, crucial for the theory and applications.

Fuzzy Relations

The concept of a relation is crucial for virtually all areas of mathematics and its applications, and the same holds true for fuzzy sets theory and its applications.

A *fuzzy relation R* between two (nonfuzzy) sets $X = \{x\}$ and $Y = \{y\}$ is defined as a fuzzy set in the Cartesian product $X \times Y$, i. e.

$$R = \{(\mu_R(x, y), (x, y))\} = \{\mu_R(x, y)/(x, y)\},$$
$$\text{for each } (x, y) \in X \times Y \tag{46}$$

where $\mu_R(x, y): X \times Y \longrightarrow [0, 1]$ is the membership function of the fuzzy relation R, and $\mu_R(x, y) \in [0, 1]$

gives the degree to which the elements $x \in X$ and $y \in Y$ are between each other in relation R.

The above fuzzy relation is defined in the Cartesian product of (nonfuzzy!) two sets, X and Y, and is called a binary fuzzy relation. In general, a fuzzy relation may be defined in the Cartesian product of k sets, $X_1 \times \ldots \times X_k$, and is then called a *k-ary fuzzy relation*. In this perspective, a fuzzy set is an unary fuzzy relation.

Example 13 If $X = \{\text{horse, donkey}\}$ and $Y = \{\text{mule, cow}\}$, then the fuzzy relation R labeled *similarity* may be exemplified by

$$R = similarity = 0.8/(\text{horse, mule}) + 0.4/(\text{horse, cow})$$
$$+ 0.9/(\text{donkey, mule})$$
$$+ 0.2/(\text{donkey, cow})$$

to be read as: The horse and the mule are similar (with respect to *our own* subjective aspects!) to degree 0.8, i. e. to a very high extent, the horse and the cow are similar to degree 0.4, i. e. to quite a low extent, etc.

It may easily be seen that the concept of a fuzzy relation makes it possible to express an imprecise, or imprecisely specified, relationship between elements of some sets, as opposed to a precise and abrupt one in the case of a nonfuzzy relation in which any two elements can either be or not in relation.

A fuzzy relation R in $X \times Y$ for X and Y of a sufficiently low dimensionality may be conveniently represented in matrix form exemplified, for the fuzzy relation $R = similarity$ in Example 13, by

$R = similarity =$		$y = $ mule	cow
	$x = $ horse	0.8	0.4
	donkey	0.9	0.2

Since a fuzzy relation is a fuzzy set, all the definitions, properties, operations, etc. on fuzzy sets presented in Sects. "The Idea of a Fuzzy Set" – "Basic Operations on Fuzzy Sets" hold as well, and we will concentrate below on those more specific ones.

The *max-min composition* of two fuzzy relations R in $X \times Y$ and S in $Y \times Z$, written $R \circ_{\max-\min} S$ is defined as a fuzzy relation in $X \times Z$ such that

$$\mu_{R \circ_{\max-\min} S}(x, y) = \max_{y \in Y} [\mu_R(x, y) \wedge \mu_S(y, z)],$$
$$\text{for each } x \in X, \ z \in Z \tag{47}$$

and since this type of composition will be used throughout this paper, if not otherwise specified, then it will be briefly denoted as $R \circ S$.

Example 14   If $X = \{1, 2\}$, $Y = \{1, 2, 3\}$ and $Z = \{1, 2, 3, 4\}$, and the fuzzy relations R and S are as below. Its resulting max-min composition, $R \circ S$, is then:

$R \circ S =$	$x = 1$	$y = 1$	2	3
	$x = 1$	0.3	0.8	1
	2	0.9	0.7	0.4

$\circ$

	$z = 1$	2	3	4
$y = 1$	0.7	0.6	0.4	0.1
2	0.4	1	0.7	0.2
3	0.5	0.9	0.6	0.8

$=$	$z = 1$	2	3	4
$x = 1$	0.5	0.9	0.7	0.8
2	0.7	0.7	0.7	0.4

This max-min composition of fuzzy relations is the original Zadeh's definition (cf. Zadeh [97]), and is certainly the most widely used. However, if we notice that the two basic operations involved in the definition of their composition, i. e. *min* ($\wedge$) and *max* ($\vee$) are just specific examples of the *t*-norm and *s*-norm (*t*-conorm) discussed in Sect. "Basic Operations on Fuzzy Sets", then one can well define a much more general type of composition given below.

The *s − t-norm composition* of two fuzzy relations R in $X \times Y$ and S in $Y \times Z$, written $R \circ_{s-t} S$, is defined as a fuzzy relation in $X \times Z$ such that

$$\mu_{R \circ_{s-t} S}(x, z) = s_{y \in Y}[\mu_R(x, y) t \mu_S(y, z)] \,,$$
$$\text{for each } x \in X \,, \ z \in Z \,. \quad (48)$$

Fuzzy relations, similar to their nonfuzzy counterparts, play a crucial role in virtually all aspects of the theory and application of fuzzy sets, notably in rule based fuzzy modeling discussed later in this paper. An important issue is related to so-called fuzzy relational equations. We will not deal with this due to lack of space, and we refer the interested reader to, for instance, the recent book by Peeva and Kyosev [72].

Finally, let us mention two concepts concerning the fuzzy sets that are related to fuzzy relations.

The *Cartesian product* of two fuzzy sets A in X and B in Y, written $A \times B$, is defined as a fuzzy set in $X \times Y$ such that

$$\mu_{A \times B}(x, y) = [\mu_A(x) \wedge \mu_B(y)] \,,$$
$$\text{for each } x \in X \,, \ y \in Y \,. \quad (49)$$

A fuzzy relation R in $X \times Y \times \ldots \times Z$ is said to be *decomposable* if and only if it can be represented as

$$\mu_R(x, y, \ldots, z) = \mu_{R_x}(x) \wedge \mu_{R_y}(y) \wedge \ldots \wedge \mu_{R_z}(z) \,,$$
$$\text{for each } x \in X \,, \ y \in Y \,, \ldots \,, \ z \in Z \,, \quad (50)$$

where $\mu_{R_x}(x), \mu_{R_y}(y), \ldots, \mu_{R_z}(z)$ are projections of the fuzzy relation $\mu_R(x, y)$ on $X, Y, \ldots, Z$, respectively, defined as

$$\mu_{R_x}(x) = \sup_{\{y, \ldots, z\} \in Y \times Z} \mu_R(x, y, \ldots, z) \,,$$
$$\text{for each } x \in X \quad (51)$$

Fuzzy relations play a major role in fuzzy sets theory and its applications, and will also be relevant for some of our next considerations.

Linguistic Variable, Fuzzy Conditional Statement, and Compositional Rule of Inference

We have already mentioned that while defining various concepts and properties of fuzzy sets it is expedient to use natural language because linguistic terms and descriptions best provide semantics. Now we will briefly present the essence of Zadeh's [97] *linguistic approach* in which this fact has been exploited to an even greater extent. This approach has inspired and triggered many new areas of research, notably fuzzy control which has resulted in so many real-world applications in diverse areas and has been a decisive factor in the so-called *fuzzy boom* which was started in the mid-1980s by the launching of a multitude of domestic appliances and professional products exemplified by fuzzy logic-controlled washing machines, cameras, automobile automatic transmissions, cranes, subway trains, etc. These applications have been decisive for a wide acceptance of fuzzy logic as a powerful and potentially useful tool.

Basically, the rationale behind Zadeh's [97] linguistic approach to the analysis of complex systems and decision (and control) processes is that the basic element is a *linguistic variable* exemplified by "temperature" which takes on as their values not conventional numerical values as, e. g., 150°C, but linguistic values as, e. g., "high," "low," etc. that are in turn equated semantically with some fuzzy sets. Notice that such linguistic values are common in human discourse as natural language is the only fully natural human means of communication. Clearly, one can then form more complex linguistic expressions as, e. g., "not very low and not very high," "more or less medium," etc. by using some connectives (e. g., and, or, . . .), modifiers (e. g., more or less, very, . . .), etc. (see also Sect. "Fuzzy Logic – Basic Issues"), and employing a syntactic analysis.

To represent a relationship between linguistic variables, *fuzzy conditional statements* are employed. For instance, if we have two linguistic variables, a primary one L and a secondary one K, such that the value of L is a fuzzy set A in X, and the value of K is a fuzzy set B in Y, then

a relationship between L and K, in terms of their values A and B, respectively, may be written as

$$\text{IF}\quad L = A\quad \text{THEN}\quad K = B \qquad (52)$$

or, shortly

$$\text{IF}\quad A\quad \text{THEN}\quad B. \qquad (53)$$

This fuzzy conditional statement is now assumed to be equivalent to

$$\text{IF}\quad A\quad \text{THEN}\quad B = A \times B \qquad (54)$$

i. e. to the Cartesian product (49) of the two fuzzy sets A and B which is in turn a fuzzy relation in $X \times Y$.

Notice that this is what Mamdani [61] used in his controller, and which is often called Mamdani's implication (though it is not an implication!). This definition is the simplest one, and we can devise more sophisticated ones by using, e. g., various definitions of implication (80)–(85).

It is easy to see that the fuzzy conditional statement (53) may be extended to account for multiple values of A and B obtaining, if we use (54),

$$\text{IF}\quad A_1\quad \text{THEN}\quad B_1\quad \text{ELSE}\quad \dots\quad \text{ELSE}$$
$$\text{IF}\quad A_n\quad \text{THEN}\quad B_n = A_1 \times B_1 + \dots + A_n \times B_n, \qquad (55)$$

where A_i's are fuzzy sets in X and B_i's are fuzzy sets in Y, $i = 1, \dots, n$.

Notice that in (53) and (55) we only specify what happens if the primary variable takes on some value. It is often, however, also relevant to explore what happens if that value is not taken. In such a case (53) becomes

$$\text{IF}\quad A\quad \text{THEN}\quad B\quad \text{ELSE}\quad C \qquad (56)$$

and it is represented as

$$\text{IF}\quad A\quad \text{THEN}\quad B\quad \text{ELSE}\quad C = A \times B + \neg A \times C. \qquad (57)$$

Evidently, one can generalize the above fuzzy conditional statements to involve more than one primary variable. For details, we refer the reader to, e. g., Zadeh [97].

We therefore have some tool to represent a relation between a primary and secondary variable that is represented by some fuzzy relation. An immediate question is then:

> If the primary variable takes on some fuzzy value, say A', and we have a (fuzzy) relation, IF A THEN B, then what will be the implied (inferred) value of the secondary variable B'?

This may be represented by the inference scheme

$$\begin{array}{l} A' \\ \text{IF}\quad A\quad \text{THEN}\quad B \\ \hline \qquad\qquad B' = ? \end{array} \qquad (58)$$

and, what is the very essence, the fuzzy values A' and A need not be the same (notice that this prohibits the use of conventional logical inference tools).

The answer to the question (58) is provided by the *compositional rule of inference* which states that if R in $X \times Y$ is a fuzzy relation representing a dependence between a primary and secondary variable, represented by a fuzzy conditional statement, and the primary variable takes on a fuzzy value A' in X, then the implied fuzzy value of the secondary variable B' in Y is given by the (max-min) composition (47) of A' and R, i. e.

$$\mu_{B'}(y) = \max_{x \in X}[\mu_A(x) \wedge \mu_R(x, y)], \quad \text{for each}\ \ y \in Y \quad (59)$$

and notice that A' is here considered to be a unary fuzzy relation as mentioned in Sect. "Fuzzy Relations".

Example 15 Suppose that $X = \{x\} = \{1, 2, 3\}$ and $Y = \{y\} = \{1, 2, 3, 4\}$, and the fuzzy conditional statement representing the dependence between L and K is

$$\text{IF}\quad L = low\quad \text{THEN}\quad K = high = (low) \circ (high)$$

where

$$low = 1/1 + 0.7/2 + 0.3/3$$
$$high = 0.2/1 + 0.5/2 + 0.8/3 + 1/4$$

and is equivalent to the following fuzzy relation

$$R = (low) \circ (high) =$$

	$y = 1$	2	3	4
$x = 1$	0.2	0.5	0.8	1
2	0.2	0.5	0.7	0.7
3	0.2	0.3	0.3	0.3

If now $L = medium = 0.5/1 + 1/2 + 0.5/3$, then the K induced is given by

$$K = (medium) \circ R = \max_{x \in \{1,2,3\}}[\mu_L(x) \wedge \mu_R(x, y)]$$

$$= 0.2/1 + 0.5/2 + 0.7/3 + 0.7/4$$

The fuzzy conditional statements may be used to represent simple dependencies and relations between linguistic variables. For more complicated dependencies and relations, fuzzy algorithms may be used (cf. Zadeh [97]) which will not be dealt with here.

A further extension in the spirit of the linguistic approach presented in this section is Zadeh's *computing with words* or, more generally, *computing with words and perceptions*. We will not deal with this and will refer the reader to a comprehensive coverage of main theoretical issues, and many applications, related to this approach which is given in Zadeh and Kacprzyk [105,106].

The Extension Principle

Now we will briefly present the essence of Zadeh's classic *extension principle* (cf. Zadeh [97]) which is one of the most important and powerful tools in fuzzy sets theory. The extension principle addresses the following fundamental issue:

If there is some relationship (e. g., a function) between *conventional* (nonfuzzy) entities (e. g., variables taking on nonfuzzy values), then what is its equivalent relationship between fuzzy entities (e. g., variables taking on fuzzy values)?

The extension principle makes it therefore possible, for instance, to extend some known conventional models, algorithms, etc. involving nonfuzzy variables to the case of fuzzy variables.

Let $A_1, \ldots, A_n$ be fuzzy sets in $X_1 = \{x_1\}$, $\ldots, X_n = \{x_n\}$, respectively, and

$$f \colon X_1 \times \ldots \times X_n \longrightarrow Y \qquad (60)$$

be some (nonfuzzy) function such that $y = f(x_1, \ldots, x_n)$.

Then, according to the *extension principle*, the fuzzy set B in $Y = \{y\}$ induced by the fuzzy sets $A_1, \ldots, A_n$ via the function f (60) is

$$\mu_B(y) = \max_{(x_1,\ldots,x_n)\in X_1\times\ldots\times X_n \,:\, y=f(x_1,\ldots,x_n)} \bigwedge_{i=1}^{n} \mu_{A_i}(x_i) . \qquad (61)$$

Example 16 Suppose that: $X_1 = \{1, 2, 3\}$, $X_2 = \{1, 2, 3, 4\}$, f represents the addition, i. e. $y = x_1 + x_2$, $A_1 = 0.1/1 + 0.6/2 + 1/3$ and $A_2 = 0.6/1 + 1/2 + 0.5/3 + 0.1/4$, then

$$B = A_1 + A_2$$
$$= 0.1/2 + 0.6/3 + 0.6/4 + 1/5 + 0.5/6 + 0.1/7$$

and notice that $+$ is used here in both the arithmetic (the sum of real and fuzzy numbers – cf. Sect. "Fuzzy Numbers") and set-theoretic sense which should not lead to confusion.

Equivalently, the extension principle (61) may also be written in terms of the α-cuts (18). Namely, suppose for simplicity that $f \colon X \longrightarrow Y$, $X = \{x\}$, $Y = \{y\}$, and A_α, for each $\alpha \in (0, 1]$, are α-cuts of A. Then, the fuzzy set B in Y, induced by A via the extension principle is given as

$$B = f(A) = f\left(\sum_{\alpha\in(0,1]} \alpha \cdot A_\alpha\right) = \sum_{\alpha} \alpha f(A_\alpha) \qquad (62)$$

which is clearly implied by the representation theorem (Theorem 1).

Notice that the extension principle plays an extraordinary role in the sense that we have a multitude of nonfuzzy tools like algorithms, procedures, etc. which have been widely used to solve various problems. However, they need precise information, notably real or integer numbers, real intervals, etc. We have neither fuzzy computers nor fuzzy tools of the type mentioned above so that we are not in a position to directly use imprecise information to solve our problems. However, by virtu of the extension principle we can extend our known nonfuzzy tools and techniques (their related algorithms and procedures) to their fuzzy counterparts.

Fuzzy Numbers

The same fundamental role as played by nonfuzzy (real, integer, …) numbers in conventional models, the fuzzy numbers play in fuzzy models.

A *fuzzy number* is defined as a fuzzy set in R, the real line. Usually, but not always, it is assumed to be a normal and convex fuzzy set. For example, the membership function of a fuzzy number that is *more or less six* may be as shown in Fig. 6, i. e. as a bell-shaped function.

For our purposes, operations on fuzzy numbers are the most relevant. It is easy to notice that their definitions may readily be obtained by applying the extension principle (61).

Suppose therefore that A and B are two fuzzy numbers in $R = \{x\}$ characterized by their membership functions
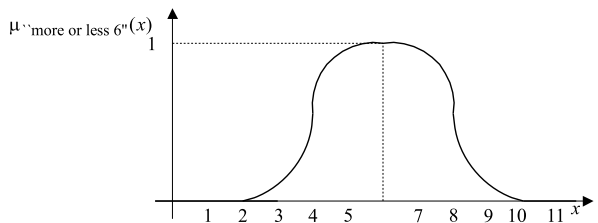
Fuzzy Sets Theory, Foundations of, Figure 6
The membership function of a fuzzy number *more or less six*

$\mu_A(x)$ and $\mu_B(x)$, respectively. Then the extension principle yields the following definitions of the four basic *arithmetic operations* on fuzzy numbers:

- *Addition*

$$\mu_{A+B}(z) = \max_{x+y=z} [\mu_A(x) \wedge \mu_B(y)], \quad \text{for each } z \in R,$$

(63)

- *Subtraction*

$$\mu_{A-B}(z) = \max_{x-y=z} [\mu_A(x) \wedge \mu_B(y)], \quad \text{for each } z \in R,$$

(64)

- *Multiplication*

$$\mu_{A \cdot B}(z) = \max_{x \cdot y=z} [\mu_A(x) \wedge \mu_B(y)], \quad \text{for each } z \in R,$$

(65)

- *Division*

$$\mu_{A/B}(z) = \max_{x/y=z, y \neq 0} [\mu_A(x) \wedge \mu_B(y)], \quad \text{for each } z \in R.$$

(66)

In some applications, the following one-argument operations on fuzzy numbers may also be of use:

- The *opposite* of a fuzzy number

$$\mu_{-A}(x) = \mu_A(-x), \quad \text{for each } x \in R,$$

(67)

- the *inverse* of a fuzzy number

$$\mu_{A^{-1}}(x) = \mu_A \left(\frac{1}{x} \right), \quad \text{for each } x \in R \setminus \{0\}.$$

(68)

In practice, however, such a general definition of fuzzy numbers and operations on them is seldom used. Normally a further simplification is made, namely the fuzzy numbers are assumed to be *triangular* and eventually *trapezoid* fuzzy numbers whose membership functions are sketched in Fig. 7.

For triangular and trapezoid fuzzy numbers, for the four basic operations, i. e. the addition, subtraction, multiplication and division, special formulas can be devised whose calculation is simpler than that of (63)–(66). Moreover, a crucial problem of comparison of two fuzzy numbers may also be simplified. For details, we will refer the reader to a rich literature exemplified by Dubois and Prade [28], Kaufmann and Gupta [48], Klir and Folger [51], Klir and Yuan [53], Mareš [62], Hanss [41], etc.
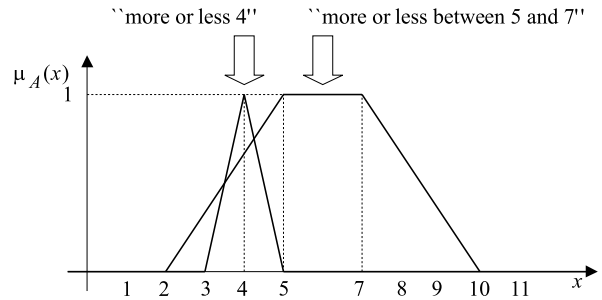
Fuzzy Sets Theory, Foundations of, Figure 7
Triangular (*about* four) and trapezoid (*more or less* between six and seven) fuzzy numbers

Fuzzy Events and Their Probabilities

Fuzziness and randomness are meant, in our perspective, as two aspects of *imperfect information*. Fuzziness is meant to concern entities and relations which are not crisply defined, with gradual transition between the elements belonging and not belonging to a class. Randomness concerns situations in which the event is well defined but it occurrence is uncertain.

However, in practice and in everyday discourse there is an abundance of situations in which there jointly occur fuzziness and randomness, for instance, when we ask about the probability of *cold weather* tomorrow or of a *high inflation* in the next year, we have an imprecise (fuzzy) event – cold weather and high inflation, respectively. To be able to formally deal with such problems, we need a concept of a fuzzy event, and that of a probability of a fuzzy event.

The first approach in this respect is due to Zadeh [96]. Its point of departure is the concept of a *fuzzy event* which is simply a fuzzy set A in $X = \{x\} = \{x_1, \dots, x_n\}$ whose membership function is Borel measurable. We assume that the probabilities of the (nonfuzzy) elementary events $x_1, \dots, x_n \in X$ are known and equal to $p(x_1), \dots, p(x_n) \in [0, 1]$, respectively, with $p(x_1) + \dots + p(x_n) = 1$.

As to some more important concepts related to fuzzy events, the following may be stated.

Two fuzzy events A and B in X are *independent* if and only if

$$p(AB) = p(A)p(B).$$

(69)

The *conditional probability* of a fuzzy event A in X with respect to a fuzzy event B in X is denoted $p(A \mid B)$ and defined as

$$p(A \mid B) = \frac{p(AB)}{p(B)}, \quad p(B) > 0$$

(70)

and if the fuzzy events A and A are independent, then

$$p(A \mid B) = p(A) \tag{71}$$

Notice that both of the above concepts are analogous to their nonfuzzy counterparts.

The (*nonfuzzy*) *probability of a fuzzy event A* in $X = \{x_1, \ldots, x_n\}$ is denoted $p(A)$ and defined by Zadeh [96] as

$$p(A) = \sum_{i=1}^{n} \mu_A(x_i) p(x_i) \tag{72}$$

i. e. as the expected value of the membership function of A, $\mu_A(x)$.

Example 17 Suppose that $X = \{1, 2, \ldots, 5\}$, $p(x_1) = 0.1$, $p(x_2) = 0.1$, $p(x_3) = 0.1$, $p(x_4) = 0.3$, $p(x_5) = 0.4$, and $A = 0.1/2 + 0.5/3 + 0.7/4 + 0.9/5$.

Then

$$p(A) = 0.1 \times 0.1 + 0.1 \times 0.5 + 0.3 \times 0.7 + 0.4 \times 0.9 = 0.73.$$

Notice that Zadeh's [96] (nonfuzzy) probability of a fuzzy event (72) satisfies:

1. $p(\emptyset) = 0$,
2. $p(\neg A) = 1 - p(A)$,
3. $p(A + B) = p(A) + p(B) - p(A \cap B)$,
4. and

$$p\left(\sum_{i=1}^{r} A_i\right) = \sum_{i=1}^{r} p(A_i) - \sum_{j=1}^{r} \sum_{k=1, k<j}^{r} p(A_j \cap A_k)$$
$$+ \sum_{j=1}^{r} \sum_{k=1, k<j}^{r} \sum_{l=1, l<k}^{r} p(A_j \cap A_k \cap A_l)$$
$$+ \ldots + (-1)^{r+1} p(A_1 \cap A_2 \cap \ldots \cap A_r)$$

so it does make sense to term the expression (72) a "probability."

The above Zadeh' [96] classic definition of a (non-fuzzy) probability of a fuzzy event is by far the most popular and most widely used. However, though the event is fuzzy, its probability is nonfuzzy, i. e. is a real number from the unit interval. This may be viewed counter-intuitive but provides simplicity. For some approaches to a fuzzy probability of a fuzzy event, see, e. g., Klir and Folger [51] or Klir and Yuan [53].

Defuzzification of Fuzzy Sets

In many applications we arrive at a fuzzy result. However, in it is a crisp (non-fuzzy) result that should be applied.

A notable example is fuzzy control (cf. Driankov, Hellendoorn and Reinfrank [27] or Kacprzyk [44].

Suppose that we have a fuzzy set A defined in $X = \{x_1, x_2, \ldots, x_n\}$, i. e. $A = \mu_A(x_1)/x_1 + \mu_A(x_2)/x_2 + \ldots + \mu_A(x_n)/x_n$. We need to find a crisp number $a \in [x_1, x_n]$ which *best* represents A. Notice that we assume here that A is defined in a finite universe of discourse, but its corresponding defuzzified number a need not be in general any of the finite values of X but should be between the lowest and highest elements of X (evidently, this requires some ordering of x_i's but this is clearly satisfied as x_i's are in virtually all practical cases just real numbers).

The most commonly used defuzzification procedure is certainly the *center-of-area*, also called the *center-of-gravity*, method whose essence is

$$a = \frac{\sum_{i=1}^{n} x_i \mu_A(x_i)}{\sum_{i=1}^{n} \mu_A(x_i)}. \tag{73}$$

The above defuzzification (73) is however often too complex if our analysis involves, e. g., some optimization (cf. Kacprzyk [44]). In such a case one needs to resort to an even simpler defuzzification method which simply assumes that the defuzzified value of a fuzzy value is $x_i \in X = \{x_1, \ldots, x_n\}$ for which $\mu_A(x)$ takes on its maximum values, i. e.

$$\mu_A(a) = \max_{x_i \in X} \mu_A(x) \tag{74}$$

with an obvious extension that if the A determined in (74) is not unique, then we take, say, the mean value of such equivalent a's.

In Sect. "Bellman and Zadeh's General Approach to Decision Making Under Fuzziness" we will provide a justification of the above maximizing-value-type defuzzification procedure in the framework of decision making.

There is a whole array of other defuzzification procedures, and the reader is referred to, e. g., Driankov, Hellendorn and Reinfrank [27], Kacprzyk [44], Klir and Yuan [53] or Yager and Filev [92].

Fuzzy Logic – Basic Issues

The concept of a *fuzzy logic* is not uniquely understood. Basically, it may be meant in (at least) the three following ways:

- As a foundation of reasoning based on ambiguous, vague and imprecise statements (cf. Goguen [38]),
- as a foundation of reasoning based on ambiguous, vague and imprecise statements in which fuzzy set-theoretic tools are used (see, e. g., Zadeh [98]; or Zadeh and Kacprzyk [104]), and

- as a multivalued logic with truth values in the unit interval in which the logical operations of negation, union, intersection, implication, equivalence, etc. are chosen in a special way, and have some fuzzy interpretation (cf. Hájek [40] or Nová, Perfilieva and Močkoř [71]).

It is easy to notice that the meaning of fuzzy logic in the first and second sense is similar, though the generality of the former is clearly higher, while its meaning in the third way is different.

For the purposes of this paper, we will assume the third view on fuzzy logic, and will present a very limited survey of basic issues. The interested reader is referred for more detail on various aspects of fuzzy logic to Zadeh and Kacprzyk's [104] volume which is practically the only up-to-date and exhaustive treatise on fuzzy logic available today.

Notice what some authors mean by fuzzy logic is the whole theory of fuzzy sets and related topics. We will not follow this line of reasoning, and view fuzzy sets theory as more *set-theoretic* while fuzzy logic as more *logical*.

Suppose that we have a statement (predicate) u is P denoted, for brevity, as P and exemplified by temperature (u) is high (P), where u is a variable taking on its values in a universe of discourse $U = \{u\}$, and P is an imprecise term equated with a fuzzy set in U, $P = \{\mu_P(u)/u\}$.

For a specified value $u \in U$ the truth of u is P (or of P) is denoted $\tau(P)$ and meant to be $\tau(u$ is $P) = \tau(P) = \mu_P(u)$, for each $u \in U$.

The following general definitions of basic logical operations (in terms of their respective truth values) are usually employed:

- The *negation* of P, i. e. *not P*, denoted by $\neg P$:

$$\tau(\neg P) = 1 - \tau(P), \tag{75}$$

- the *intersection* of P and Q, i. e. *P and Q*, denoted by $P \cap Q$:

$$\tau(P \cap Q) = t[\tau(P), \tau(Q)] \tag{76}$$

where $t: [0,1] \times [0,1] \longrightarrow [0,1]$ is a t-norm (36); the original Zadeh's definition is

$$\tau(P \cap Q) = \tau(P) \wedge \tau(Q) = \min[\tau(P), \tau(Q)], \tag{77}$$

- the *union* of P and Q, i. e. *P or Q*, denoted by $P \cup Q$:

$$\tau(P \cup Q) = s[\tau(P), \tau(Q)] \tag{78}$$

where $s: [0,1] \times [0,1] \longrightarrow [0,1]$ is an s-norm (40); the original Zadeh's definition is

$$\tau(P \cup Q) = \tau(P) \vee \tau(Q) = \max[\tau(P), \tau(Q)], \tag{79}$$

- the implication, i. e. if P then Q, denoted by $P \Longrightarrow Q$, which may be defined as, e. g.:
  1. The *Łukasiewicz implication*

$$\tau(P \Longrightarrow Q) = \min\{1 - \tau(P) + \tau(Q), 1\}, \tag{80}$$

  2. the *Gödel implication*

$$\tau(P \Longrightarrow Q) = \begin{cases} 1 & \text{if } \tau(P) \leq \tau(Q) \\ \tau(Q) & \text{otherwise}, \end{cases} \tag{81}$$

  3. the *Goguen implication*

$$\tau(P \Longrightarrow Q) = \begin{cases} 1 & \text{if } \tau(P) = 0 \\ \min\left\{1, \frac{\tau(Q)}{\tau(P)}\right\} & \text{otherwise}, \end{cases} \tag{82}$$

  4. the *Kleene–Dienes implication*

$$\tau(P \Longrightarrow Q) = \max\{1 - \tau(P), \tau(Q)\}, \tag{83}$$

  5. the *Zadeh implication*

$$\tau(P \Longrightarrow Q) = \max\{1 - \tau(P), \min\{\tau(P), \tau(Q)\}\}, \tag{84}$$

  6. the *Reichenbach implication*

$$\tau(P \Longrightarrow Q) = 1 - \tau(P) + \tau(P) \cdot \tau(Q), \tag{85}$$

- The *equivalence*, i. e. *P is equivalent to Q* or if *P then Q* and if *Q then P*, denoted by $P \Longleftrightarrow Q$:

$$\tau(P \Longleftrightarrow Q) = \tau[(P \Longrightarrow Q) \cap (Q \Longrightarrow P)] \tag{86}$$

and we can assume an appropriate definition of the intersection $\cap$ (34), and the implication $\Longrightarrow$ (80)–(85).

Among some other important aspects of fuzzy logic, one should mention the use of (fuzzy) linguistic quantifiers, exemplified by *most*, *almost all*, *a few*, etc. which are common in everyday discourse but cannot be handled by conventional logic in which only the two quantifiers are in principle employed, i. e. the universal quantifier *for all* and the existential quantifier *for at least one*.

A *linguistically quantified statement* is exemplified by *most experts are convinced* and may be generally written as

$$Q y\text{'s are } F, \tag{87}$$

where Q is a linguistic quantifier (e. g., most), $Y = \{y\}$ is a set of objects (e. g., experts) and F is a property (e. g., convinced).

Importance B may also be added to the linguistically quantified statement (87) yielding

$$QBy's \text{ are } F \tag{88}$$

exemplified by *most (Q) of the important (B) experts (y's) are convinced (F)*.

For our purposes, the problem is to find the (degree of) truth of such linguistically quantified statements (87) and (88), denoted $\tau(Qy's \text{ are } F)$ in the former case and $\tau(QBy's \text{ are } F)$ in the latter case, knowing the truth of the statements, y is F, denoted $\tau(y \text{ is } F)$, for all $y \in Y$. Evidently, all these degrees of truth (truths) will be meant as real numbers from the unit interval.

Fortunately enough these truth values may be found by some fuzzy logic calculi [cf. Kacprzyk [44], Yager [91] or Zadeh [101]. For lack of space we are unable to discuss these issues here, and refer the reader to, e. g., the source papers or Kacprzyk's [44] book.

Fuzzy linguistic quantifiers are very relevant both for theory and applications (ranging from decision analysis, social choice, optimization and control to database queries). The fuzzy linguistic quantifiers have a relation to the so-called ordered weighted averaging (OWA) operators introduced by Yager in 1982, and we refer the reader for a comprehensive coverage of their theory and applications to Yager and Kacprzyk's [93] volume.

Bellman and Zadeh's General Approach to Decision Making Under Fuzziness

Fuzzy logic, the essence of which has been presented in previous sections, has found applications in a multitude of areas of science and technology, and a full coverage is beyond the scope of our exposition.

Since *decision making* is by far the most well known, omnipresent problem, we will just sketch the application of fuzzy sets theory to the broadly perceived decision making.

The purpose of this section is to provide the reader with a brief introduction to Bellman and Zadeh's [8] general approach to decision making under fuzziness, originally termed *decision making in a fuzzy environment*, a simple yet extremely powerful framework within which virtually all fuzzy models related to decision making, optimization and control have been dealt with.

In Bellman and Zadeh's [8] setting the imprecision (fuzziness) of the environment within which the decision making (control, …) process proceeds is modeled by the introduction of the so-called *fuzzy environment* which consists of fuzzy goals, fuzzy constraints, and fuzzy decision.

Suppose that we have some set of possible *options* (or alternatives, variants, choices, decisions, …), $X = \{x\}$, which contains all the possible (relevant, feasible, …) values, courses of action, etc.

The *fuzzy goal* is now defined as a fuzzy set in the set G in the set of options X, characterized by its membership function $\mu_G \colon X \longrightarrow [0, 1]$ such that $\mu_G(x) \in [0, 1]$ specifies the grade of membership of a particular option $x \in X$ in the fuzzy goal G.

The *fuzzy constraint* is similarly defined as a fuzzy set C in the set of options X, characterized by its membership function $\mu_C \colon X \longrightarrow [0, 1]$ such that $\mu_C(x) \in [0, 1]$ specifies the grade of membership of a particular option $x \in X$ in the fuzzy constraint C.

The fuzzy goal and fuzzy constraint are illustrated in Example 18.

As to the interpretation of fuzzy goals and/or constraints (notice that their definitions do indicate an intrinsic analogy!), in some, mostly earlier, works as, e. g., in Bellman and Zadeh [8], the following view on the essence of the fuzzy goal is advocated. Suppose that $f \colon X \longrightarrow R$ is a conventional performance (objective) function which associates with each option $x \in X$ a real number $f(x) \in R$, and which is bounded, i. e. $f(x) \leq M < \infty$, for each $x \in X$, where $M = \max_{x \in X} f(x)$.

Then the membership function of the fuzzy goal G can be defined as a normalized performance function f, i. e.

$$\mu_G(x) = \frac{f(x)}{M} = \frac{f(x)}{\max_{x \in X} f(x)}, \quad \text{for each } x \in X. \tag{89}$$

A fuzzy goal may however be also viewed from a different perspective that is often convenient, i. e. in terms of Simon's *satisfaction levels*, in particular in view of the representation of the fuzzy goal's membership function in a piecewise linear form as usually assumed, also here. Then the piecewise linear membership function of a fuzzy goal in Fig. 8 should be understood as follows: If the value of x attained is at least $\overline{x}_G(= 8)$, which is the *satisfaction level* of x, i. e. for $x \geq \overline{x}_G$, then $\mu_G(x) = 1$ which means that we are fully satisfied with the x attained. On the other hand, if the x attained does not exceed $\underline{x}_G(= 5)$, which is the lowest possible value of x, then $\mu_G(x) = 0$ which means that we are fully dissatisfied with such a value of x or, in other words, this value is impossible. For the intermediate values, $\underline{x}_G < x < \overline{x}_G$, we have $0 < \mu_G(x) < 1$ which means that our satisfaction as to a particular value of x is intermediate. The interpretation of a fuzzy constraint is analogous.

It is now easy to see that the above interpretation provides a *common denominator* for the fuzzy goal and fuzzy constraint. They may be treated in an analogous
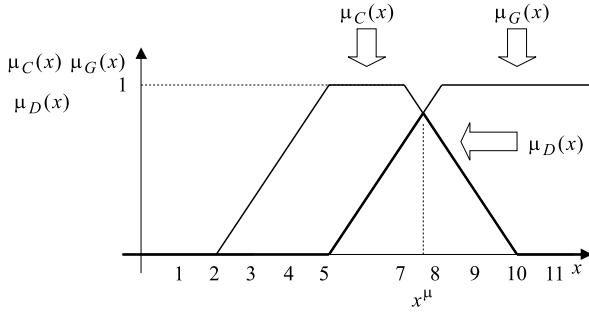
Fuzzy Sets Theory, Foundations of, Figure 8
Fuzzy goal, fuzzy constraint, fuzzy decision, and the optimal (maximizing) decision

way which is one of merits of Bellman and Zadeh's [8] approach.

The above suggests the following general formulation of the decision making problem in a fuzzy environment

"Attain G and satisfy C "   (90)

which should be meant as to determine a decision (an option or a set of options) which simultaneously fulfills the fuzzy goal and fuzzy constraint, and which belongs to the available (or, maybe, relevant, feasible, …) ones.

The fuzziness of the fuzzy goal and fuzzy constraint implies the above decision, a fuzzy decision, to be a fuzzy set defined in the set of options which results from the intersection (34) of the fuzzy goal and fuzzy constraint. Formally, if G is a fuzzy goal and C is a fuzzy constraint, both defined as fuzzy sets in the set of options $X = \{x\}$, the *fuzzy decision D* is a fuzzy set defined in X given as

$$\mu_D(x) = \mu_G(x) \wedge \mu_C(x), \quad \text{for each } x \in X \quad (91)$$

where $\wedge$ is the minimum operation, i. e. $a \wedge b = \min(a, b)$.

The fuzzy decision (91) is most widely used, but $\wedge$ may clearly be replaced by another operation as, e. g., a t -norm (36).

Example 18 Suppose that G is *x should be much larger than five*, and C is *x should be about 5*, as in Fig. 8.

The membership function of the fuzzy decision is given in heavy line and interpreted as follows. The set of possible options is the interval [5, 10] because $\mu_D(x) > 0$ for $5 \le x \le 10$. The value of $\mu_D(x) \in [0, 1]$ is meant as the degree of satisfaction from the choice of a particular $x \in X$, from zero for full dissatisfaction (impossibility of x) to one for full satisfaction, though all intermediate values, and the higher the value of $\mu_D(x)$, the higher the satisfaction from x .

Notice that in Fig. 8, $\mu_D(x) < 1$ which means that there is no option which fully satisfies both the fuzzy goal and fuzzy constraint. In other words, there is a discrepancy or conflict between the fuzzy goal and constraint.

In practice, however, we need to find a nonfuzzy solution to be implemented. The above interpretation of the fuzzy decision immediately suggests that the best (nonfuzzy) choice in this case would be the one corresponding to the highest value of $\mu_D(x)$.

The *maximizing decision* is therefore defined as an $x^* \in X$ such that

$$\mu_D(x^*) = \max_{x \in X} \mu_D(x) \quad (92)$$

and an example may be found in Fig. 8 where $x^* = 7.5$.

Notice that the above is clearly equivalent to the defuzzification of the fuzzy decision (cf. Sect. "Defuzzification of Fuzzy Sets"), and other defuzzification procedures may also be used in principle. However, (92) is often the only practical choice (cf. Kacprzyk [44]), and will be assumed here.

In non-trivial real problems there are multiple fuzzy goals and fuzzy constraints, and they may be handled within the above framework in quite a straightforward manner.

Suppose a more general situation than the fuzzy constraint C is defined as a fuzzy set in $X = \{x\}$, and the fuzzy goal G is defined as a fuzzy set in $Y = \{y\}$. Moreover, suppose that a function $f: x \longrightarrow Y, y = f(x)$, is known. Typically, X and Y may be sets of options and outcomes, notably causes and effects.

Now, the *induced fuzzy goal G'* in X generated by the given fuzzy goal G in Y is defined as

$$\mu_{G'}(x) = \mu_G[f(x)], \quad \text{for each } x \in X. \quad (93)$$

Example 19 Let $X = \{1, 2, 3, 4\}$, $Y = \{2, 3, \dots, 10\}$, and $y = 2x + 1$. If now

$$G = 0.1/2 + 0.2/3 + 0.4/4 + 0.5/5 + 0.6/6 + 0.7/7$$
$$+ 0.8/8 + 1/9 + 1/10$$

then

$$G' = \mu_G(3)/1 + \mu_G(5)/2 + \mu_G(7)/70.2/1 + 0.5/2$$
$$+ 0.7/3 + 1/4.$$

The *fuzzy decision* is now defined analogously as (91), i. e.

$$\mu_D(x) = \mu_{G'}(x) \wedge \mu_C(x), \quad \text{for each } x \in X. \quad (94)$$

Clearly, for $n > 1$ fuzzy goals $G_1, \ldots, G_n$ defined in Y, $m > 1$ fuzzy constraints $C_1, \ldots, C_m$ defined in X, and a function $f: X \longrightarrow Y$, $y = f(x)$, we analogously have

$$\mu_D(x) = \mu_{G_1'}(x) \wedge \ldots \wedge \mu_{G_n'}(x) \wedge \mu_{C_1}(x) \wedge \ldots \wedge \mu_{C_n}(x),$$
$$\text{for each } x \in X . \quad (95)$$

The *maximizing decision* is defined as (92), i. e.

$$\mu_D(x^*) = \max_{x \in X} \mu_D(x) .$$

The basic conceptual fuzzy decision making model can be used in many specific areas, notably in fuzzy optimization which will be covered elsewhere in this volume.

The models of decision making under fuzziness developed above can also be extended to the case of multiple criteria, multiple decision makers, and multiple stage cases. We will present the last extension, to fuzzy multistage decision making (control) case which makes it possible to account for dynamics.

Multistage Decision Making (Control) Under Fuzziness

In this case it is convenient to use control-related notation and terminology. In particular, decisions will be referred to as *controls*, the discrete time moments at which decisions are to be made – as *control stages*, and the input-output (or cause-effect) relationship – as a *system under control*.

The essence of multistage control in a fuzzy environment may be portrayed as in Fig. 9.

First, suppose that the control space is $U = \{u\} = \{c_1, \ldots, c_m\}$ and the state space is $X = \{x\} = \{s_1, \ldots, s_n\}$. Initially we are in some initial state $x_0 \in X$. We apply a control $u_0 \in U$ subjected to a fuzzy constraint $\mu_{C^0}(u_0)$. We attain a state $x_1 \in X$ via a known cause-effect relationship (i. e. S); a fuzzy goal $\mu_{G^1}(x_1)$ is imposed on x_1.

Next, we apply a control u_1 subjected to a fuzzy constraint $\mu_{C^1}(u_1)$, and attain a fuzzy state x_2 on which a fuzzy goal $\mu_{G^2}(x_2)$ is imposed, etc.

Suppose for simplicity that the system under control is deterministic and its temporal evolution is governed by a *state transition equation*

$$f: X \times U \longrightarrow X , \quad (96)$$

such that

$$x_{t+1} = f(x_t, u_t) , \quad t = 0, 1, \ldots \quad (97)$$

where $x_t, x_{t+1} \in X = \{s_1, \ldots, s_n\}$ are the states at control stages t and $t + 1$, respectively, and $u_t \in U = \{c_1, \ldots, c_m\}$ is the control at t.

At each t, $u_t \in U$ is subjected to a *fuzzy constraint* $\mu_{C^t}(u_t)$, and on the state attained $x_{t+1} \in X$ a *fuzzy goal* is imposed; $t = 0, 1, \ldots$. The *initial state* is $x_0 \in X$ and is assumed to be known, and given in advance. The *termination time* (planning, or control, horizon), i. e. is the maximum number of control stages, is denoted by $N \in \{1, 2, \ldots\}$, and may be finite or infinite.

The performance (goodness) of the multistage control process under fuzziness is evaluated by the fuzzy decision

$$\mu_D(u_0, \ldots, u_{N-1} \mid x_0) = \mu_{C^0}(u_0) \wedge \mu_{G^1}(x_1) \wedge \ldots$$
$$\wedge \mu_{C^{N-1}}(u_{N-1}) \wedge \mu_{G^N}(x_N) . \quad (98)$$

In most cases, however, a slightly simplified form of the fuzzy decision (98) is used, namely it is assumed all the subsequent fuzzy controls, $u_0, u_1, \ldots, u_{N-1}$, are subjected to the fuzzy constraints, $\mu_{C^0}(u_0), \mu_{C^1}(u_1), \ldots, \mu_{C^{N-1}}(u_{N-1})$, while the fuzzy goal is just imposed on the final state x_N, $\mu_{G^N}(x_N)$. In such a case the fuzzy decision becomes

$$\mu_D(u_0, \ldots, u_{N-1} \mid x_0) = \mu_{C^0}(u_0) \wedge \ldots$$
$$\wedge \mu_{C^{N-1}}(u_{N-1}) \wedge \mu_{G^N}(x_N) . \quad (99)$$
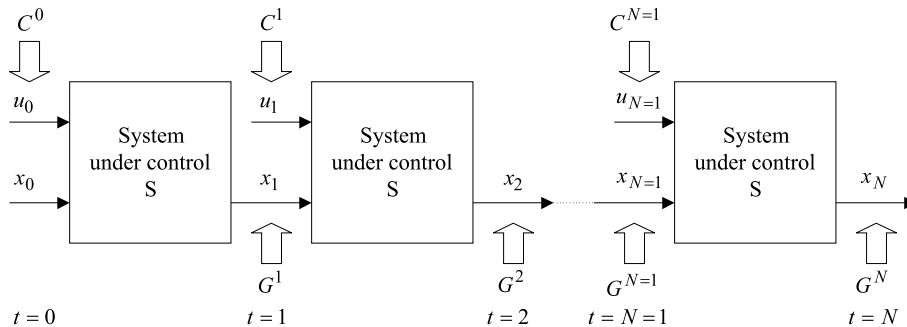
Fuzzy Sets Theory, Foundations of, Figure 9
Essence of the multistage control in a fuzzy environment (under fuzziness)

The multistage control problem in a fuzzy environment is now formulated as to find an optimal sequence of controls $u_0^*, \ldots, u_{N-1}^*$, $u_t^* \in U$, $t = 0, 1, \ldots, N-1$, such that:

$$\mu_D(u_0^*, \ldots, u_{N-1}^* \mid x_0) = \max_{u_0, \ldots, u_{N-1} \in U} \mu_D(u_0, \ldots, u_{N-1} \mid x_0) . \quad (100)$$

Usually it is more convenient to express the solution, i. e. the controls to be applied, as a *control policy* $a_t \colon X \longrightarrow U$, such that $u_t = a_t(x_t)$, $t = 0, 1, \ldots$, i. e. the control to be applied at t is expressed as a function of the state at t.

The above basic formulation of multistage control in a fuzzy environment may readily be extended with respect to:

- The type of the termination time (fixed and specified, implicitly specified, fuzzy, and infinite), and
- the type of the system under control (deterministic, stochastic, and fuzzy).

For a detailed analysis of resulting problems and their solutions (by employing dynamic programming, branch-and-bound, genetic algorithms, etc.) we refer the reader to Kacprzyk's [42,44] books.

Concluding Remarks

We provided a brief survey of basic elements of Zadeh's [95] fuzzy sets theory, mainly of basic properties of fuzzy sets, operations on fuzzy sets, fuzzy relations and their compositions, linguistic variables, the extension principle, fuzzy arithmetic, fuzzy events and their probabilities, fuzzy logic, and Bellman and Zadeh's [8] general approach to decision making in a fuzzy environment. Various aspects of fuzzy sets theory will be expanded in other papers in this part of the volume.

Bibliography

1. Atanassov KT (1983) Intuitionistic fuzzy sets. VII ITKR session. Central Sci.-Techn. Library of Bulg. Acad. of Sci., Sofia pp 1697/84. (in Bulgarian)
2. Atanassov KT (1986) Intuitionistic fuzzy sets. Fuzzy Sets Syst 20:87–96
3. Atanassov KT (1999) Intuitionistic fuzzy sets: Theory and applications. Springer, Heidelberg
4. Bandemer H, Gottwald S (1995) Fuzzy sets, fuzzy logic, fuzzy methods, with applications. Wiley, Chichester
5. Bandemer H, Näther W (1992) Fuzzy data analysis. Kluwer, Dordrecht
6. Bandler W, Kohout LJ (1980) Fuzzy power sets for fuzzy implication operators. Fuzzy Sets Syst 4:13–30
7. Bellman RE, Giertz M (1973) On the analytic formalism of the theory of fuzzy sets. Inform Sci 5:149–157
8. Bellman RE, Zadeh LA (1970) Decision making in a fuzzy environment. Manag Sci 17:141–164
9. Belohlávek R, Vychodil V (2005) Fuzzy equational logic. Springer, Heidelberg
10. Bezdek JC (1981) Pattern recognition with fuzzy objective function algorithms. Plenum, New York
11. Black M (1937) Vagueness: An exercise in logical analysis. Philos Sci 4:427–455
12. Black M (1963) Reasoning with loose concepts. Dialogue 2: 1–12
13. Black M (1970) Margins of precision. Cornell University Press, Ithaca
14. Buckley JJ (2004) Fuzzy statistics. Springer, Heidelberg
15. Buckley JJ (2005) Fuzzy probabilities. New approach and applications, 2nd edn. Springer, Heidelberg
16. Buckley JJ (2005) Simulating fuzzy systems. Springer, Heidelberg
17. Buckley JJ (2006) Fuzzy probability and statistics. Springer, Heidelberg
18. Buckley JJ, Eslami E (2002) An introduction to fuzzy logic and fuzzy sets. Springer, Heidelberg
19. Calvo T, Mayor G, Mesiar R (2002) Aggregation operators. New trends and applications. Springer, Heidelerg
20. Carlsson C, Fullér R (2002) Fuzzy reasoning in decision making and optimization. Springer, Heidelberg
21. Castillo O, Melin P (2008) Type-2 fuzzy logic: Theory and applications. Springer, Heidelberg
22. Cox E (1994) The fuzzy system handbook. A practitioner's guide to building, using, and maintaining fuzzy systems. Academic, New York
23. Cross V, Sudkamp T (2002) Similarity and compatibility in fuzzy set theory. Assessment and applications. Springer, Heidelberg
24. Delgado M, Kacprzyk J, Verdegay JL, Vila MA (eds) (1994) Fuzzy optimization: Recent advances. Physica, Heidelberg
25. Dompere KK (2004) Cost-benefit analysis and the theory of fuzzy decisions. Fuzzy value theory. Springer, Heidelberg
26. Dompere KK (2004) Cost-benefit analysis and the theory of fuzzy decisions. Identification and measurement theory. Springer, Heidelberg
27. Driankov D, Hellendorn H, Reinfrank M (1993) An introduction to fuzzy control. Springer, Berlin
28. Dubois D, Prade H (1980) Fuzzy sets and systems: Theory and applications. Academic, New York
29. Dubois D, Prade H (1985) Théorie des possibilités. Applications a la représentation des connaissances en informatique. Masson, Paris
30. Dubois D, Prade H (1988) Possibility theory: An approach to computerized processing of uncertainty. Plenum, New York
31. Dubois D, Prade H (1996) Fuzzy sets and systems (Reedition on CR-ROM of [28]. Academic, New York
32. Fullér R (2000) Introduction to neuro-fuzzy systems. Springer, Heidelberg
33. Gaines BR (1977) Foundations of fuzzy reasoning. Int J Man-Mach Stud 8:623–668
34. Gil Aluja J (2004) Fuzzy sets in the management of uncertainty. Springer, Heidelberg
35. Gil-Lafuente AM (2005) Fuzzy logic in financial analysis. Springer, Heidelberg

36. Glöckner I (2006) Fuzzy quantifiers. A computational theory. Springer, Heidelberg
37. Goguen JA (1967) *L*-fuzzy sets. J Math Anal Appl 18:145–174
38. Goguen JA (1969) The logic of inexact concepts. Synthese 19:325–373
39. Goodman IR, Nguyen HT (1985) Uncertainty models for knowledge-based systems. North-Holland, Amsterdam
40. Hájek P (1998) Metamathematics of fuzzy logic. Kluwer, Dordrecht
41. Hanss M (2005) Applied fuzzy arithmetic. An introduction with engineering applications. Springer, Heidelberg
42. Kacprzyk J (1983) Multistage decision making under fuzziness, Verlag TÜV Rheinland, Cologne
43. Kacprzyk J (1992) Fuzzy sets and fuzzy logic. In: Shapiro SC (ed) Encyclopedia of artificial intelligence, vol 1. Wiley, New York, pp 537–542
44. Kacprzyk J (1996) Multistage fuzzy control. Wiley, Chichester
45. Kacprzyk J, Fedrizzi M (eds) (1988) Combining fuzzy imprecision with probabilistic uncertainty in decision making. Springer, Berlin
46. Kacprzyk J, Orlovski SA (eds) (1987) Optimization models using fuzzy sets and possibility theory. Reidel, Dordrecht
47. Kandel A (1986) Fuzzy mathematical techniques with applications. Addison Wesley, Reading
48. Kaufmann A, Gupta MM (1985) Introduction to fuzzy mathematics – theory and applications. Van Nostrand Reinhold, New York
49. Klement EP, Mesiar R, Pap E (2000) Triangular norms. Springer, Heidelberg
50. Klir GJ (1987) Where do we stand on measures of uncertainty, ambiguity, fuzziness, and the like? Fuzzy Sets Syst 24:141–160
51. Klir GJ, Folger TA (1988) Fuzzy sets, uncertainty and information. Prentice-Hall, Englewood Cliffs
52. Klir GJ, Wierman M (1999) Uncertainty-based information. Elements of generalized information theory, 2nd end. Springer, Heidelberg
53. Klir GJ, Yuan B (1995) Fuzzy sets and fuzzy logic: Theory and application. Prentice-Hall, Englewood Cliffs
54. Kosko B (1992) Neural networks and fuzzy systems. Prentice-Hall, Englewood Cliffs
55. Kruse R, Meyer KD (1987) Statistics with vague data. Reidel, Dordrecht
56. Kruse R, Gebhard J, Klawonn F (1994) Foundations of fuzzy systems. Wiley, Chichester
57. Kuncheva LI (2000) Fuzzy classifier design. Springer, Heidelberg
58. Li Z (2006) Fuzzy chaotic systems modeling, control, and applications. Springer, Heidelberg
59. Liu B (2007) Uncertainty theory. Springer, Heidelberg
60. Ma Z (2006) Fuzzy database modeling of imprecise and uncertain engineering information. Springer, Heidelberg
61. Mamdani EH (1974) Application of fuzzy algorithms for the control of a simple dynamic plant. Proc IEE 121:1585–1588
62. Mareš M (1994) Computation over fuzzy quantities. CRC, Boca Raton
63. Mendel J (2000) Uncertain rule-based fuzzy logic systems: Introduction and new directions. Prentice Hall, New York
64. Mendel JM, John RIB (2002) Type-2 fuzzy sets made simple. IEEE Trans Fuzzy Syst 10(2):117–127
65. Mordeson JN, Nair PS (2001) Fuzzy mathematics. An introduction for engineers and scientists, 2nd edn. Springer, Heidelberg
66. Mukaidono M (2001) Fuzzy logic for beginners. World Scientific, Singapore
67. Negoiţa CV, Ralescu DA (1975) Application of fuzzy sets to system analysis. Birkhäuser/Halstead, Basel/New York
68. Nguyen HT, Waler EA (205) A first course in fuzzy logic, 3rd end. CRC, Boca Raton
69. Nguyen HT, Wu B (2006) Fundamentals of statistics with fuzzy data. Springer, Heidelberg
70. Novák V (1989) Fuzzy sets and their applications. Hilger, Bristol, Boston
71. Novák V, Perfilieva I, Močkoř J (1999) Mathematical principles of fuzzy logic. Kluwer, Boston
72. Peeva K, Kyosev Y (2005) Fuzzy relational calculus. World Scientific, Singapore
73. Pedrycz W (1993) Fuzzy control and fuzzy systems, 2nd edn. Research Studies/Wiley, Taunton/New York
74. Pedrycz W (1995) Fuzzy sets engineering. CRC, Boca Raton
75. Pedrycz W (ed) (1996) Fuzzy modelling: Paradigms and practice. Kluwer, Boston
76. Pedrycz W, Gomide F (1998) An introduction to fuzzy sets: Analysis and design. MIT Press, Cambridge
77. Petry FE (1996) Fuzzy databases. Principles and applications. Kluwer, Boston
78. Piegat A (2001) Fuzzy modeling and control. Springer, Heidelberg
79. Ruspini EH (1991) On the semantics of fuzzy logic. Int J Approx Reasining 5:45–88
80. Rutkowska D (2002) Neuro-fuzzy architectures and hybrid learning. Springer, Heidelberg
81. Rutkowski L (2004) Flexible neuro-fuzzy systems. Structures, learning and performance evaluation. Kluwer, Dordrecht
82. Seising R (2007) The fuzzification of systems. The genesis of fuzzy set theory and its initial applications – developments up to the 1970s. Springer, Heidelberg
83. Smithson M (1989) Ignorance and uncertainty. Springer, Berlin
84. Sousa JMC, Kaymak U (2002) Fuzzy decision making in modelling and control. World Scientific, Singapore
85. Thole U, Zimmermann H-J, Zysno P (1979) On the suitability of minimum and product operator for the intersection of fuzzy sets. Fuzzy Sets Syst 2:167–180
86. Turkşen IB (1991) Measurement of membership functions and their acquisition. Fuzzy Sets Syst 40:5–38
87. Türksen IB (2006) An ontological and epistemological perspective of fuzzy set theory. Elsevier, New York
88. Wang Z, Klir GJ (1992) Fuzzy measure theory. Kluwer, Boston
89. Wygralak M (1996) Vaguely defined objects. Representations, fuzzy sets and nonclassical cardinality theory. Kluwer, Dordrecht
90. Wygralak M (2003) Cardinalities of fuzzy sets. Springer, Heidelberg
91. Yager RR (1983) Quantifiers in the formulation of multiple objective decision functions. Inf Sci 31:107–139
92. Yager RR, Filev DP (1994) Essentials of fuzzy modeling and control. Wiley, New York
93. Yager RR, Kacprzyk J (eds) (1996) The ordered weighted averaging operators: Theory, methodology and applications. Kluwer, Boston

94. Yazici A, George R (1999) Fuzzy database modeling. Springer, Heidelberg
95. Zadeh LA (1965) Fuzzy sets. Inf Control 8:338–353
96. Zadeh LA (1968) Probability measures of fuzzy events. J Math Anal Appl 23:421–427
97. Zadeh LA (1973) Outline of a new approach to the analysis of complex systems and decision processes. IEEE Trans Syst, Man Cybern SMC-2:28–44
98. Zadeh LA (1975) Fuzzy logic and approximate reasoning. Synthese 30:407–428
99. Zadeh LA (1975) The concept of a linguistic variable and its application to approximate reasoning. Inf Sci (Part I) 8:199–249, (Part II) 8:301–357, (Part III) 9:43–80
100. Zadeh LA (1978) Fuzzy sets as a basis for a theory of possibility. Fuzzy Sets Syst 1:3–28
101. Zadeh LA (1983) A computational approach to fuzzy quantifiers in natural languages. Comput Math Appl 9:149–184
102. Zadeh LA (1985) Syllogistic reasoning in fuzzy logic and its application to usuality and reasoning with dispositions. IEEE Trans Syst Man Cybern SMC-15:754–763
103. Zadeh LA (1986) Fuzzy probabilities. Inf Process Manag 20:363–372
104. Zadeh LA, Kacprzyk J (eds) (1992) Fuzzy logic for the management of uncertainty. Wiley, New York
105. Zadeh LA, Kacprzyk J (eds) (1999) Computing with words in information/intelligent systems. 1 Foundations. Springer, Heidelberg
106. Zadeh LA, Kacprzyk J (eds) (1999) Computing with words in information/intelligent systems. 2 Applications. Springer, Heidelberg
107. Zang H, Liu D (2006) Fuzzy modeling and fuzzy control. Birkhäuser, New York
108. Zimmermann H-J (1976) Description and optimization of fuzzy systems. Int J Gen Syst 2:209–215
109. Zimmermann H-J (1987) Fuzzy sets, decision making, and expert systems. Kluwer, Dordrecht
110. Zimmermann H-J (1996) Fuzzy set theory and its applications, 3rd edn. Kluwer, Boston
111. Zimmermann H-J, Zysno P (1980) Latent connectives in human decision making. Fuzzy Sets Syst 4:37–51

Fuzzy System Models Evolution from Fuzzy Rulebases to Fuzzy Functions

I. Burhan Türkşen
Head Department of Industrial Engineering, TOBB-ETÜ, (Economics and Technology University of the Union of Turkish Chambers and Commodity Exchanges), Ankara, Republic of Turkey

Article Outline

Glossary

Z-FRB Zadeh's linguistic fuzzy rule base.

TS-FR Takagi–Sugeno fuzzy rule base.

c^* the number of rules in the rule base.

nv the number of input variables in the system.

$X = (x_1, x_2, \ldots, x_{nv})$ input vector.

x_j the input (explanatory variable), for $j = 1, \ldots, nv$.

A_{ji} the linguistic label associated with jth input variable of the antecedent in the ith rule.

B_i the consequent linguistic label of the ith rule.

R_i ith rule with membership function $\mu_i(x_j)\colon x_j \to [0, 1]$.

A_i multidimensional type 1 fuzzy set to represent the ith antecedent part of the rules defined by the membership function $\mu_i(x)\colon x \to [0, 1]$.

$a_i = (a_{i,1}, \ldots, a_{i,nv})$ the regression coefficient vector associated with the ith rule.

b_i the scalar associated with the ith rule in the regression equation.

SFF-LSE "Special Fuzzy Functions" that are generated by the Least Squares Estimation.

SFF-SVM "Special Fuzzy Functions" estimated by Support Vector Machines.

The estimate of y_i would be obtained as $Y_i^* = \beta_{i0}^* + \beta_{i1}^* \Gamma_i + \beta_{i2}^* X$ with SFF-LSE

y the dependent variable, assumed to be a linear function.

β_j, $j = 0, 1, \ldots, nv$, indicate how a change in one of the independent variables affects the dependent variable.

$X = (x_{j,k} \mid j = 1, \ldots, nv, \ k = 1, \ldots, nd)$ the set of observations in a training data set.

m the level of fuzziness, $m = 1.1, \ldots, 2.5$.

c the number of clusters, $c = 1, \ldots, 10$.

J the objective function to be minimized.

$\|\cdot\|_A$ a norm that specifies a distance-based similarity between the data vector x_k and a fuzzy.

$A = I$ the Euclidean norm.

$A = COV^{-1}$ the Mahalonobis norm.

COV the covariance matrix.

m^*, c^* the optimal pair.

$v_{X|Y,i} = (x_{1,i}, x_{2,i}, \ldots, x_{nv,i}, y_i)$ the cluster centers for $m = m^*$ and each cluster $i = 1, \ldots, c^*$.

$v_{X,i} = (x_{1,i}, x_{2,i}, \ldots, x_{nv,i})$ the cluster centers of the m^* "input space" for $m = m^*$ and $c = 1, \ldots, c^*$.

$\gamma_{ik}(x_k)$ the normalized membership values of x data sample in the ith cluster, $i = 1, \ldots, c^*$.

$\Gamma_i = (\gamma_{ik} \mid i = 1, \ldots, c^*; k = 1, \ldots, nd)$ the membership values of x in the ith cluster.

X_i', X_i'', X_i''' potential augmented input matrices in SFF-LSE.

$f(\vec{x}_k) = \hat{y}_k = \langle \vec{w}, \vec{x}_k \rangle + b$ linear Support Vector Regression (SVR) equation.

$l_\varepsilon = |y_k - f(\vec{x}_k)|_\varepsilon = \max\{0, |y - f(x)| - \varepsilon\}$ ε-insensitive loss function.

$\vec{w}, b$ the weight vector and bias term.

$c > 0$ the tradeoff between the empirical error and the complexity term.

$\xi_k \geq 0$ and $\xi_k^* \geq 0$ the slack variables.

α_k and α_k^* Lagrange multipliers.

$K\langle \vec{x}_{k'}, \vec{x}_k \rangle$ the kernel mapping of the input vectors.

$\hat{y}_{ik'}^* = \hat{f}(\vec{x}_{ik'}, \alpha_i, \alpha_i^*) = \sum_{k=1}^{nd} (\alpha_{ik} - \alpha_{ik}^*) K\langle \vec{x}_{ik'}, \vec{x}_{ik} \rangle + b_i$ output value of kth data sample in ith cluster with SFF-SVM.

$\tilde{A} = \{(x, (u, f_x(u))) \mid x \in X, u \in J_x \subseteq [0, 1]\}$ type 2 fuzzy set, $\tilde{A}$.

$f_x(u): J_x \to [0, 1], \forall u \in J_x \subseteq [0, 1], \forall x \in X$ *secondary membership function.*

$f_x(u) = 1, \forall x \in X, \forall u \in J_x, J_x \subseteq [0, 1]$ *interval value type 2 membership function.*

[27] the domain of the primary membership is discrete and the secondary membership values are fixed to 1. Thus, the proposed method is utilizing discrete interval valued type 2 fuzzy sets in order to represent the linguistic values assigned to each fuzzy variable in each rule. These fuzzy sets can be mathematically defined as follows:

$\tilde{A} = \int_{x \in X} \left[\sum_{u \in J_x} 1/u \right] / x$ Discrete Interval Valued Type 2 Fuzzy Sets (DIVT2FS)

$x \in X \subseteq \Re, u \in J_x = \{J_{xr}\}, r = 1, \ldots, NM$ a system variable in continuous domain.

u the primary membership value with discrete domain.

J_{xr} rth primary membership value associated with x.

Definition of the Subject

Fuzzy System Modeling (FSM) is one of the most prominent tools that can be used to identify the behavior of highly nonlinear systems with uncertainty. In the past, FSM techniques utilized type 1 fuzzy sets in order to capture the uncertainty in the system. However, since type 1 fuzzy sets express the belongingness of a crisp value x' of an input variable x in a fuzzy set A by a crisp membership value $\mu_A(x')$, they cannot fully capture the uncertainties associated with higher order imprecision in identifying membership functions. In the future, we are likely to observe higher types of fuzzy sets, such as type 2 fuzzy sets. The use of type 2 fuzzy sets and linguistic logical connectives drew a considerable amount of attention in the realm of fuzzy system modeling in the last two decades. In this paper, we first review type 1 fuzzy system models known as Zadeh, Takagi–Sugeno and Türkşen models; then we review potentially future realizations of type 2 fuzzy systems again under the headings of Zadeh and Takagi–Sugeno and Türkşen fuzzy system models, in contrast to type 1 fuzzy system models. Type 2 fuzzy system models have a higher predictive power. One of the essential problems of type 2 fuzzy system models is computational complexity. In data-driven fuzzy system modeling methods discussed here, Fuzzy C-Means (FCM) clustering algorithm is used in order to identify the system structure.

Introduction

Fuzzy system models are the most successful models to handle uncertainties in decision-making. The major advantages of fuzzy system models are their robustness and transparency. Fuzzy system modeling achieves robustness by using fuzzy sets which incorporates imprecision in system models. In addition, unlike some other system models, such as neural networks, the fuzzy system models are highly descriptive, i. e., transparent.

In the last two decades, researchers proposed several data driven type 1 fuzzy system modeling approaches that can extract the hidden rules of a system behavior automatically by using historical data. The system modeling methods, proposed by Nakanishi et al. [31], Takagi–Sugeno [39], Sugeno and Yasukawa [38], Emami et al. [17], are among the most notable ones. Since these methods utilize only the historical data, i. e., since they do not require expert knowledge, they are strictly data-driven modeling techniques. Thus, in addition to being robust and transparent, these system-modeling techniques can identify system model structure objectively for a given performance measure.

In these traditional fuzzy system models of the past, the structure is characterized by type 1 fuzzy sets. Type 1 fuzzy sets, defined on a universe of discourse, maps an element onto a precise number in the unit interval [0, 1]. This conflicts with the basic philosophy of fuzzy set and logic theory. In the future, it is expected, fuzzy system models will be formed with higher order fuzzy sets, such as type 2 fuzzy sets, which was first proposed by Zadeh [50]. They will be

used more and more in order to capture the uncertainty associated with membership functions. A type 2 fuzzy set can be informally defined as a fuzzy set that is characterized by a fuzzy membership function, i. e., membership values also are in the unit interval [0, 1] in the computational level.

In this paper, we propose to review first the type 1 fuzzy system models as the historically significant but successful modeling activity of the past in the domain of fuzzy control system problems. Next we review the type 2 fuzzy system models as the potential future modeling activity in the domain of fuzzy decision support system problems.

In an historical sense, Zadeh [50], and Takagi–Sugeno [39] versions of type 1 fuzzy system models are typically basic fuzzy rule base models. In contrast, type 1 "special fuzzy function" models, recently proposed by Türkşen [46], are alternate models to fuzzy rule base models. They give better predictions in comparison to type 1 fuzzy rule base models [3,46]. Furthermore, fuzzy rule base models, in general, can not capture the interactive nature of a problem space due to projection deficiency. Whereas fuzzy function models are able to capture the interactions of all variables since they are not subject to projection deficiencies.

For future developments, currently there are mainly two essentially different schools of thought in type 2 fuzzy system modeling research. The first one is based on the properties of the linguistic connectives which causes the generation of interval valued type 2 fuzzy sets. Türkşen [42,43,44,45] mathematically showed that Fuzzy Conjunctive Canonical Forms (FCCF) and Fuzzy Disjunctive Canonical Forms (FDCF) are no longer equivalent to each other for the 16 basic linguistic expressions formed with linguistic connectives "AND", "OR", etc. Furthermore, it is shown that FCCF contains FDCF for certain families of t-norms and t-co-norms. Zimmerman and Zysno [51] empirically showed that linguistic connectives were characterized differently by different people and Türkşen [44] provided the groundwork for this with the Interval Valued type 2 fuzzy sets. These type 2 system models represent uncertainties generated by different linguistic connectives that combine type 1 membership functions. Such combinations generate their FDCF and FCCF expressions and hence identify a particular sort of Interval Valued type 2 representations of systems.

In the second school of thought in type 2 fuzzy system modeling, traditional connectives, i. e., t-norms and co-norms, are utilized directly with the assumption that a t-norm directly corresponds to a linguistic "AND" in its FDCF and a t-co-norm directly corresponds to an "OR" in its FCCF while ignoring FCCF of "AND" and FDCF of

"OR". But, each variable is represented by using a type 2 fuzzy set at the beginning of computations. In a series of papers Mendel, Karnik and Lian [23,24,25] extended traditional type 1 system models to type 2 fuzzy system models and proposed inference methods in order to process type 2 fuzzy sets. These studies were explained thoroughly by Mendel in [28]. Several other researchers such as, Starczewski and Rutkowski [37], John and C. Czarnecki [20,21], and Chen and Kawase [10] worked on type 2 fuzzy system models and inference methods.

In general the main inhibitor of the use of the type 2 fuzzy system models is the computational complexity of the inference mechanism that is used to infer a model output by using type 2 fuzzy system models for a given input data vector. Liang and Mendel [26] proposed Interval-Valued (IV) type 2 fuzzy system models in place of full type 2 fuzzy system models together with inference methods to remedy this problem. Thus they use a simplified version of type 2 fuzzy sets, namely, Interval-Valued type 2 fuzzy sets (IVT2FS) rather than full type 2 fuzzy sets (FT2FS).

Recently, Uncu and Türkşen [48] proposed discrete "interval valued type 2 fuzzy sets (DIVT2FS)" which are generated by a variation of the level of fuzziness around fixed cluster centers in applications of FCM. The proposed model representation enables us to identify a computationally efficient inference mechanism.

The rest of this paper is organized as follows: the basic notation, terminology and the three well-known type 1 fuzzy system model structures will be briefly reviewed in Sect. "Introduction" with an emphasis on the more recent "special fuzzy functions". Then the extensions of these well-known fuzzy system modeling structures in type 2 formation will be discussed for potential future studies in Sect. "Future of Fuzzy System Models". Finally, the conclusions will be drawn and the future research directions will be provided in Sect. "Conclusions and Future Directions".

Type 1 Fuzzy System Models of the Past

In general fuzzy system models identify an underlying relationship between input and output variables of a system. In this paper, we deal with Multi-Input Single Output (MISO) version fuzzy system models. Generally fuzzy system models represent relationships between the input and output variables as a collection of: (a) either if-then rules that utilize linguistic labels (i. e., fuzzy sets) or (b) "special fuzzy functions" that takes on membership values and/or their transformations as well as selected input variables as their arguments.

Type 1 Fuzzy Rulebases

In general, the fuzzy rule base structure can be written as follows:

$$R: \underset{i=1}{\overset{c^*}{\text{ALSO}}} \left(\text{IF antecedent}_i \text{ THEN consequent}_i \right), \quad (1)$$

where c^* is the number of rules in the rule base. There are several well known fuzzy rule base structures which mainly differ in the representation of their consequents. If the consequent is represented with fuzzy sets then the rule base can be categorized as the Zadeh Fuzzy Rule base, Z-FRB [50] Whereas, if the consequent is represented with a linear equation of input variables, then the rule base structure is known as Takagi–Sugeno Fuzzy Rulebase (TS-FRB) structure [39]. The Z-FRB and TS-FRB structures can be formalized as follows:

Let nv be the number of input variables in the system. Then, the multidimensional antecedent, X, can be defined as $X = (x_1, x_2, \ldots, x_{nv})$, where x_j is the jth input variable of the antecedent in the domain of X. X, can be defined as $X = X_1 \times X_2 \times \ldots \times X_{nv}$, where $X_j \subseteq \Re$ is the domain of variable x_j. Similarly, the domain of the output variable, will be denoted as $Y \subseteq \Re$. Then, the ith rule, r_i, and rulebase, R, in Z-FR structure can be defined as:

$$R_i: \text{ IF } \underset{j=1}{\overset{NV}{\text{AND}}} (x_j \in X_j \text{ isr } A_{ji}) \text{ THEN } y \in Y \text{ isr } B_i,$$
$$\forall i = 1, \ldots, c^* \quad (2)$$

$$R: \underset{i=1}{\overset{c^*}{\text{ALSO}}} \left(\text{ IF } \underset{j=1}{\overset{NV}{\text{AND}}} (x_j \in X_j \text{ isr } A_{ji}) \right.$$
$$\left. \text{THEN } y \in Y \text{ isr } B_i \right), \quad (3)$$

where A_{ji} is the linguistic label associated with jth input variable of the antecedent in the ith rule, R_i, with membership function $\mu_i(x_j): x_j \to [0, 1]$ and similarly B_i is the consequent linguistic label of the ith rule with membership function $\mu_i(y): y \to [0, 1]$, and c^* is the number of rules in the model. The above structure assumes non-interactivity between input variables because the membership functions of every A_{ji} is obtained by the projection of *nv+1 dimensional internal system representation*. In order to eliminate the non-interactivity assumption, Delgado et al. [12], Babuska et al. [1], and Uncu and Türkşen [48] used multidimensional type 1 fuzzy sets to represent the antecedent part of the rules. Hence, the Z-FRB structure can be expressed as follows:

$$R: \underset{i=1}{\overset{c^*}{\text{ALSO}}} \left(\text{IF } x \in X \text{ isr } A_i \text{ THEN } y \in Y \text{ isr } B_i \right), \quad (4)$$

where the multidimensional antecedent fuzzy set of ith rule is defined as $\mu_i(x): x \to [0, 1]$. The other well-known fuzzy rulebase structures, namely Takagi–Sugeno (TS-FRB) fuzzy rulebase structure, can be expressed, respectively, as follows:

$$\underset{i=1}{\overset{c^*}{\text{ALSO}}} \left(\text{IF antecedent}_i \text{ THEN } y_i = a_i x^{\text{T}} + b_i \right), \quad (5)$$

where, $a_i = (a_{i,1}, \ldots, a_{i,nv})$ is the regression coefficient vector associated with the ith rule, b_i is the scalar associated with the ith rule.

Type 1 "Special Fuzzy Functions" of the Recent Past

There are at least two ways to form special fuzzy functions: (i) "Special Fuzzy Functions" that are generated by the Least Squares Estimation, SFF-LSE, of Türkşen [46] and (ii) "Special Fuzzy Functions" estimated by Support Vector Machines, SFF-SVM's, of Çelikyilmaz and Türkşen [3]. These "Special Fuzzy Functions" are structurally different from (1) Zadeh's [50] linguistic fuzzy rule bases (Z-FRB), (2) Takagi–Sugeno fuzzy rule base TS-FR [39] (3) "Fuzzy Regression" models of Tanaka et al. [40,41] and its variations, and (5) Hathaway and Bezdek [18] model. Because the proposed "Special Fuzzy Functions" introduces membership values and their transformations as new input variables in addition to the original scalar input variables in function estimations. In particular, these "Special Fuzzy Functions" are structurally different from (1) "Fuzzy Regression" models of Tanaka, et al. [40,41], and its variations, and (2) Hathaway and Bezdek [18] models. This is why we call them "Special Fuzzy Functions".

It ought to be noted that the introduction of membership values and their transformations as new input variables are acceptable since the membership values are obtained from FCM which is a highly nonlinear transformation of the original scalar input variable. Hence there is no co-linearity and thus they can be included in the proposed LSE and SVM structure. For this purpose, first one executes a fuzzy clustering algorithm such as FCM with original selected input variables after an execution of a feature selection algorithm; and then determines (local) optimum number of fuzzy clusters and hence the associated membership values. Then a special fuzzy function to represent each fuzzy cluster, i. e., fuzzy rule, separately can be identified. Thus there are as many fuzzy functions as there are fuzzy clusters similar to Hathaway and Bezdek's [18] Fuzzy C-Regression model (FCRM). But Hathaway and Bezdek use membership values as the weights to be used in the estimation of the functions using weighted least squares algorithm. FCRM updates the membership values as the similarity measure by using estimation error from these functions.

"Special Fuzzy Functions", SFF, are estimated after one generates membership values of each cluster from FCM algorithm. Therefore it is structurally a new and unique approach for the determination of fuzzy system models in place of fuzzy rule bases. This the reason we call them "Special Fuzzy Functions", SFF. These "Special Fuzzy Functions" represent fuzzy rule bases in functional form and structure. When the relationship between input variables and the output variable of the system can be linearly explained in the original dimension space of the data, it is quite reasonable and faster to estimate such "Special Fuzzy Functions" using least squares estimation. When this relationship is more complex and there needs to be a non-linear transformation of the original input variables, it is better to map the input dataset into a higher dimensional space, e. g., a *hyper-space*, where the input dimension is large (maximum n). One of the powerful methods to find these "Special Fuzzy Functions" which define a linear relation between input and output variables in the higher dimension, but a non-linear relationship in the original dimension is the support vector machines which was first proposed by Vapnik [49]. For the regression case, support vector machines for regression algorithm can be applied to find these "Special Fuzzy Functions". Hence, in the next sections, we are going to specify the details of these "Special Fuzzy Functions" estimated using the Least Squares, LSE, and Support Vector machine for Regression, SVR, algorithms.

It is to be noted for the sake of emphasis that the estimated parameters of the inputs are not fuzzy sets in our proposed approach. It should be recalled that membership values and their transformations are augmented into the original selected input set as new and additional variables. In our experience, it is found that this approach is most suitable for those analysts who are familiar with a function estimation technology, e. g., the least squares technology, support vector machines, ridge regression, etc. They only need to develop an understanding of some fuzzy clustering algorithm without studying many aspects of fuzzy theory. All they have to understand is the notion of membership values and how they can be obtained from a fuzzy clustering algorithm such as FCM and/or its variations in addition to their usual background knowledge of a function estimation technique, e. g., LSE, or SVR, etc.

Thus this is a novel approach in order to provide an easy entry into fuzzy system modeling for mathematicians and statisticians who are working in industry and for other novices. For this purpose, we present next our generalization of the LSE algorithm, which includes membership values and their transformations in addition to the original scalar input variables.

Special Fuzzy Functions with LSE (SFF-LSE) Method

In Ordinary LSE (OLSE) method, the dependent variable, y, is assumed to be a linear function of one or more independent, input, variables, x, plus an error component as follows:

$$y = \beta_0 + \beta_1 x_1 + \ldots + \beta_{nv} x_{nv} + \varepsilon \,,$$

where y is the dependent output, x_j's are the inputs (explanatory variables), for $j = 1, \ldots, nv$, nv is the number of selected inputs and ε is the independent error term which is typically assumed to be normally distributed. The goal of the least squares method is to obtain estimates of the unknown parameters, β_j's, $j = 0, 1, \ldots, nv$, which indicate how a change in one of the independent variables affects the dependent variable as follows:

$$\beta = (X'X)^{-1} X'Y \,, \tag{6}$$

where $\beta = (\beta_0, \beta_1, \ldots, \beta_{nv})$.

The proposed generalization of OLSE as **SFF-LSE**, requires that a fuzzy clustering algorithm, such as FCM [2], be available to determine the interactive (joint) membership values of input-output variables in each of the fuzzy clusters that can be identified for a given training data set.

Let (x_k, y_k), $k = 1, \ldots, nd$, be the set of observations in a training data set, such that

$$X = (x_{j,k} \mid j = 1, \ldots, nv, \;\; k = 1, \ldots, nd) \,. \tag{7}$$

First, one determines the optimal (m^*, c^*) pair for a particular performance measure, i. e., a cluster validity index, with an iterative search and an application of FCM algorithm, where m is the level of fuzziness (in our experiments, we usually take $m = 1.1, \ldots, 2.5$), and c is the number of clusters (in our experiments, we usually take $c = 2, \ldots, 10$). The well known FCM algorithm can be stated as follows:

$$\min J(U, V) = \sum_{k=1}^{nd} \sum_{i=1}^{c} (u_{ik})^m (\|x_k - v_i\|)_A$$

$$\text{s.t.} \quad 0 \le u_{ik} \le 1 \,, \quad \forall i, k$$

$$\sum_{i=1}^{c} u_{ik} = 1 \,, \quad \forall k$$

$$0 \le \sum_{k=1}^{nd} u_{ik} \le nd \,, \quad \forall i \,,$$

where J is objective function to be minimized, $\|.\|_A$ is a norm that specifies a distance-based similarity between the data vector x_k and a fuzzy cluster center v_i. In particular, $A = I$ is the Euclidean norm and $A = COV^{-1}$ is

the Mahalonobis norm, etc., where COV is the covariance matrix.

The optimal pair, (m^*, c^*), can be determined with a user defined cluster validity index, partition entropy or partition coefficient [2]. Another alternative of selecting the optimum pair would be running the overall SFF-LSE model for every (m, c) pair specified by the user and determining the optimal pair from the training RMSE values of each model. The following definitions adopt the idea of using a user defined cluster validity index for the determination of the optimal pair. The experiments in this paper follow the second alternative.

Once the optimal pair (m^*, c^*) is determined with the application of FCM algorithm, one next identifies the cluster centers for $m = m^*$ and each cluster $i = 1, \ldots, c^*$ as:

$$v_{X|Y,i} = (x_{1,i}, x_{2,i}, \ldots, x_{nv,i}, y_i) . \qquad (8)$$
$$_{m^*}$$

From this, we identify the cluster centers of the "input space" for $m = m^*$ and $c = 1, \ldots, c^*$ as:

$$v_{X,i} = (x_{1,i}, x_{2,i}, \ldots, x_{nv,i}) . \qquad (9)$$
$$_{m^*}$$

Next, one computes the normalized membership values of each data sample in the training data set with the use of the cluster center values determined in the previous step. There are generally two steps in these calculations:

(a) first we determine the (local) optimum membership values u_{ik}'s and then determine μ_{ik}'s that are above an α-cut in order to eliminate harmonics generated by FCM as:

$$u_{ik} = \left(\sum_{j=1}^{c} \left(\frac{\|x_k - v_{X,i}\|}{\|x_k - v_{X,j}\|} \right)^{\frac{2}{m-1}} \right)^{-1} ,$$

$$\mu_{ik} = \{u_{ik} \geq \alpha\} , \quad (10)$$

where μ_{ik} denotes the membership value of the kth vector, $k = 1, \ldots, nd$, in the ith rule, $i = 1, \ldots, c^*$ and x_k denotes the kth vector.

(b) next, we normalize them as:

$$\gamma_{ik}(x_k) = \frac{\mu_{ik}(x_k)}{\sum\limits_{i'=1}^{c} \mu_{i'k}(x_k)} , \qquad (11)$$

where $\gamma_{ik}(x_k)$'s are the normalized membership values of x data sample in the ith cluster, $i = 1, \ldots, c^*$, which in turn indicate the membership values that will constitute as a new input variable in our proposed scheme of function identification for the representation of ith cluster. Let

$\Gamma_i = (\gamma_{ik} | i = 1, \ldots, c^*; \ k = 1, \ldots, nd)$ be the membership values of x, a data sample, in the ith cluster, i. e., ith rule.

Next we determine a new augmented input matrix of x for each of the clusters, which could take on several forms depending on which *transformation* of membership values we want to or need to include in our system structure identification for our intended system analyses. Examples of possible augmented input matrices are:

$$X_i' = [1, \Gamma_i, X] , \text{ or}$$
$$X_i'' = \left[1, \Gamma_i^2, X\right] , \text{ or} \qquad (12)$$
$$X_i''' = \left[1, \Gamma_i^2, \Gamma_i^m, \exp(\Gamma_i), X\right] , \text{ etc.},$$

where X_i', X_i'', X_i''' are potential augmented input matrices to be used in least squares estimation of a new system structure identification and $\Gamma_i = (\gamma_{ik} | i = 1, \ldots, c^*; \ k = 1, \ldots, nd)$. The choice amongst X_i', X_i'', X_i''' depends on whether we want to or need to include just the membership values or some of their transformations as new input variables in order to obtain the best representation of a system behavior. A new augmented input matrix having a single input variable in the original input space when only membership values themselves are augmented to the dataset, i. e., X_i' may look like this:

$$X_i' = [1, \Gamma_i, X] = \begin{bmatrix} 1 & \gamma_{i,1} & x_{i,1} \\ \vdots & \vdots & \vdots \\ 1 & \gamma_{i,nd} & x_{i,nd} \end{bmatrix}$$

Up to this point, in the proposed system modeling approach, we have defined how the augmented input matrix for each cluster could be formed with the output of an FCM algorithm. Both the proposed SFF-LSE and SFF-SVM approaches implement these steps. From this point forward, the estimation of "Special Fuzzy Functions" takes place for each cluster, where one can implement any function estimation methodology, e. g., LSE or SMV. Different approaches are followed in the estimation of "Special Fuzzy Functions" using a augmented matrix. Here we continue to specify the SFF-LSE models. In the next section the SFF-SVM models will be introduced.

Thus the function of a single input single output model, which includes only the membership values as the additional input variable, $Y_i = \beta_{i_0} + \beta_{i_1} \Gamma_i + \beta_{i2} X$, that represents the ith rule corresponding to the ith interactive (joint) cluster in (Y_i, Γ_i, X) space, would be estimated with SFF-LSE approach as follows:

$$\beta_i^* = \left(X_i'^T X_i' \right)^{-1} \left(X_i'^T Y_i \right) , \qquad (13)$$
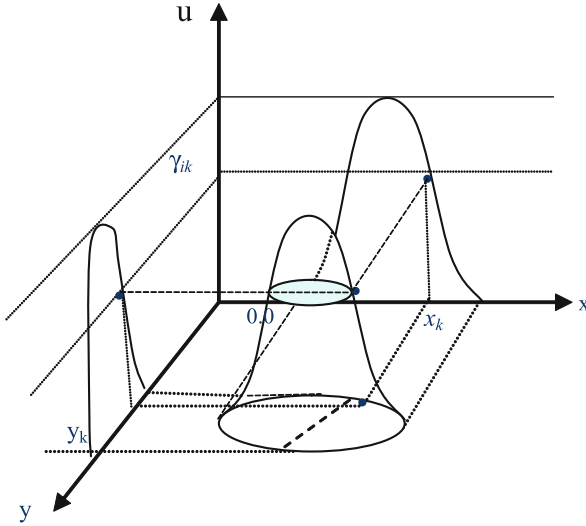
Fuzzy System Models Evolution from Fuzzy Rulebases to Fuzzy Functions, Figure 1
A fuzzy cluster in [$U \times X \times Y$] space

where $\beta_i^* = (\beta_{i0}^*, \beta_{i1}^*, \beta_{i2}^*)$ and $X_i' = [1, \Gamma_i, X]$, provided the inverse of covariance, $(X_i'^T X_i')^{-1}$, exists. The estimate of y_i would be obtained as:

$$Y_i^* = \beta_{i0}^* + \beta_{i1}^* \Gamma_i + \beta_{i2}^* X . \quad (14)$$

The single output value is calculated using each output value, one from each cluster, and weighting them with their corresponding membership values as follows:

$$Y_i^* = \frac{\sum_{i=1}^{c^*} \gamma_i Y_i^*}{\sum_{i=1}^{c^*} \gamma_i} . \quad (15)$$

Within the proposed framework, the general form of the shape of a cluster for the case of a single input variable X_j and for the ith cluster can be conceptually captured, in a stylistic, imaginary manner, say, by a second order (cone) function when one introduces the square of membership values into the augmented input matrix in the space of [$U \times X \times Y$] which can be illustrated with a prototype shown in Fig. 1.

In a number of real life case studies, we have in fact found out that generally some second order or exponential function provide a good approximation from amongst some 20 alternatives we have experimented with in the past.

Before we specify the details of the proposed SFF-SVM method, we first review briefly the background of the support vector machines for regression algorithm.

Support Vector Machines for Regression

Support Vector Machine, SVM, is a data-mining tool to build a model of a given system. The foundations of SVM have been developed by Vapnik [49]. SVM is a type of optimization technique in which prediction error and model complexity are simultaneously minimized. Let the training samples be denoted as:

$$X = \{(\vec{x}_k, \vec{y}_k) | k = 1, \ldots, nd\} , \quad (16)$$

where X denotes the space of input-output patterns $\vec{x}_k = (x_{jk} | j = 1, \ldots, j_{nv}, \ k = 1, \ldots, k_{nd})$ represents each input data vector, and y_k is the output value of the kth, data vector in the dataset. Support vector machines are used to solve classification problems as well as regression models where the output variable is scalar. In linear Support Vector Regression (SVR), the aim is to find a pair($\vec{w}, b$), where $\vec{w}$ is the weight vector, and b is the bias term in this regression equation, such that the value of the point, y_k, can be predicted according to a real-valued function:

$$f(\vec{x}_k) = \hat{y}_k = \langle \vec{w}, \vec{x}_k \rangle + b , \quad (17)$$

where $\langle . \rangle$ is the dot product representation. The goal is to find a function, that has at most ε deviation from the actually obtained targets, y_k, for all the training data. This concept of ε-insensitive loss function, l_ε, was first introduced by Vapnik [28] as follows:

$$l_\varepsilon = |y_k - f(\vec{x}_k)|_\varepsilon = \max\{0, |y - f(x)| - \varepsilon\} . \quad (18)$$

The loss function does not penalize errors below some error, $\varepsilon \geq 0$. Thus the goal of learning is to find a function with a small risk on test samples. This would mean good generalization. This type of SVR is called the ε-insensitive which embodies the Structural Risk Minimization (**SRM**) as displayed as follows:

$$R_{\exp}[f] \leq R_{\emp}[f] + R_{\text{complexity}}[f] . \quad (19)$$

In **SRM** of **SVM**, the aim is not only minimize the empirical risk from training samples, $R_{\emp}$, but also find a simple function to minimize the complexity of the model, $R_{\text{complexity}}$. The more flat the functions are, the less complex they would be, in other words they would get simpler, and therefore they would be closer to the linear functions. The more flat function, the smaller the complexity term of the expected risk in SRM, $R_{\exp}$, gets and the smaller would be the weight vector. In support vector regression the complexity term is expressed as weights assigned to all points in the training sample. In order to ensure that the

weight vector is small, Euclidean Norm, i. e., $\|w\|^2$ is used. In mathematical terms the objective function of SVM for regression with two conditions can be stated as follows.

$$\min_{\vec{w},b,\xi_k,\xi_k^*} \quad \varsigma(\vec{w},\xi,\xi^*) = \tfrac{1}{2}\|\vec{w}\|^2 + \tfrac{C}{nd}\sum_{k=1}^{nd}\left(\xi_k + \xi_k^*\right)$$
$$\text{subject to} \quad y_k - \langle \vec{w}, \vec{x}_k \rangle - b \le \varepsilon + \xi_k$$
$$\langle \vec{w}, \vec{x}_k \rangle + b - y_k \le \varepsilon + \xi_k^*$$
$$\xi_k \ge 0$$
$$\xi_k^* \ge 0,$$
$$(20)$$

where ε represents the ε insensitive value which does not penalize the points whose estimated deviations are lesser/greater than ε, and $\vec{w}$, b unknown values that represent the weight vector and bias term, respectively. The $c > 0$ determines the tradeoff between the empirical error and the complexity term. The slack variables $\xi_k \ge 0$ and $\xi_k^* \ge 0$ are introduced to the model to soften the optimization problem in order to prevent infeasible solutions. The assumption in (20) is that, it is possible to find such a function that approximates all pairs $(\vec{x}_k, y_k)$ with ε precision. The optimization problem is a convex quadratic program, which can be solved by using the well-known Lagrange multiplier method. Therefore by introducing Lagrange multipliers α_i and β_i, one can construct Lagrange function, and the solution to the optimization theorem is given by the saddle point of the Lagrange function using the Karush–Kuhn–Tucker theorem [49] where the primal model is translated into dual quadratic programming problem as follows:

$$\max_{\alpha,\alpha^*} = \frac{1}{2}\sum_{k,k'=1}^{nd}\left(\alpha_k - \alpha_k^*\right)\left(\alpha_{k'} - \alpha_{k'}^*\right)\langle \vec{x}_k, \vec{x}_{k'} \rangle$$
$$- \varepsilon \sum_{k=1}^{nd}\left(\alpha_k + \alpha_k^*\right) + \sum_{k=1}^{nd}\left(\alpha_k - \alpha_k^*\right)y_k \quad (21)$$
$$\text{subject to} \quad \sum_{k=1}^{nd}\left(\alpha_k - \alpha_k^*\right) = 0$$
$$\alpha_k, \alpha_k^* \in [0, C] .$$

In model (21), we search for the parameters α_k and α_k^*, which are Lagrange multipliers. The weight vector can now be explained using the Lagrange multipliers as:

$$\vec{w} = \sum_{k=1}^{nd}\left(\alpha_k - \alpha_k^*\right)\vec{x}_k , \quad (22)$$

where, $\vec{x}_k$ is the kth observation and α_k and α_k^* are the Lagrange multipliers for the kth observation, and the estima-

tion function of a vector is given as follows:

$$f(\vec{x}_{k'}) = \sum_{k=1}^{nd}\left(\alpha_k - \alpha_k^*\right)\langle \vec{x}_{k'}, \vec{x}_k \rangle + b$$
$$= \sum_{k=1}^{nd}\left(\alpha_k - \alpha_k^*\right)\vec{x}_k^{\mathrm{T}}\vec{x}_{k'} + b , \quad (23)$$

where $\vec{x}_{k'}$ is the new vector whose output is to be predicted, T represents the transpose operation on vectors and α_k and α_k^* are the Lagrange multipliers for the kth observation.

In most cases, there is a non-linear relationship between input and output variables and a non-linear support vector regression algorithm is needed. In order to make the support vector model non-linear, the input vectors are mapped into a higher dimensional feature space using a mapping function, $\phi(x)$. However, in most of the cases, explicit mapping results in infeasible solutions that are computationally hard to obtain. The feasible way to convert a linear SMV's into non-linear SMV is to use kernel mapping which maps the input vectors into a higher dimensional feature space, i. e., $k(x, x') = \langle \phi(x), \phi(x') \rangle$ which changes the SMV algorithm as:

$$\max_{\alpha,\alpha^*} = \frac{1}{2}\sum_{k,k'=1}^{nd}\left(\alpha_k - \alpha_k^*\right)\left(\alpha_{k'} - \alpha_{k'}^*\right)k\langle \vec{x}_k, \vec{x}_{k'} \rangle$$
$$- \varepsilon \sum_{k=1}^{nd}\left(\alpha_k + \alpha_k^*\right) + \sum_{k=1}^{nd}\left(\alpha_k - \alpha_k^*\right)y_k \quad (24)$$
$$\text{subject to} \quad \sum_{k=1}^{nd}\left(\alpha_k - \alpha_k^*\right) = 0$$
$$\alpha_k, \alpha_k^* \in [0, C] .$$

Note that one may choose various different kernel functions, e. g., Gaussian radial base kernel, polynomial kernel, satisfying the Mercer's condition [49] and the output value of the k'th input vector is calculated using the following function:

$$\hat{y}_{k'} = \hat{f}(\vec{x}_{k'}, \alpha, \alpha^*) = \sum_{k=1}^{nd}\left(\alpha_k - \alpha_k^*\right)K\langle \vec{x}_{k'}, \vec{x}_k \rangle + b, \quad (25)$$

where the $K\langle \vec{x}_{k'}, \vec{x}_k \rangle$ represents the kernel mapping of the input vectors. From Eq. (25), the dependent variable is estimated using the kernel mapping of the input vectors and their Lagrange multipliers of each vector that are calculated from the optimization algorithm. Note that in Eq. (25), it is not required to calculate the weight vector

explicitly, the input vectors who's Lagrange multipliers are not zero are used in estimating the output and they are called the "**support vectors**". In a sense, the complexity of the functions, i. e., the $\frac{1}{2}\|\vec{w}\|^2$ term, represented by support vectors is independent of the dimensionality of the input space x, and only depends on the number of support vectors.

An additional result of the application of Karush–Kuhn–Tucker (KKT) theorem in SVR is:

$$(C - \alpha_k)\xi_k = 0$$
$$(C - \alpha_k^*)\xi_k^* = 0 \,. \tag{26}$$

One of the several conclusions one might make from (26) is that $\alpha_k \alpha_k^* = 0$, i. e., there can never be a set of dual variables for an observation k which are both simultaneously non-zero as this would require non-zero slacks in both directions. Since $c > 0$, then $\xi_k \xi_k^* = 0$ must also be true. In the same sense, there can never be two slack variables $\xi_k, \xi_k^* > 0$ which are both non-zero and equal.

Special Fuzzy Functions with SVM (FF-SVM) Method

As one can build ordinary least squares for the estimation of the Special Fuzzy Functions when the relationship between input variables and the output variable can be linearly defined in the original input space, one may also build support vector regression models to estimate the parameters of the non-linear Special Fuzzy Functions using support vector regression methods. The augmented input matrix is determined from FCM algorithm such that there is one SVR in SFF-SVM for each cluster same as SFF-LSE model. One may choose any membership transformation depending on the input dataset. Then one can apply support vector regression, SVR, algorithm instead of LSE to each augmented matrix, which are comprised of the original selected input variables and the membership values and/or their transformations. Support vector machines' optimization algorithm is applied to each augmented matrix of each cluster (rule) i, $i = 1, \ldots, c^*$, to optimize their Lagrange multipliers, α_{ik} and α_{ik}^*, and find the candidate support vectors, $k = 1, \ldots, nd$. Hence, using SFF-SVM, one finds Lagrange multipliers of each kth train data sample one for each cluster, i. Then the output value of kth data sample in ith cluster is estimated using the Equation (27) as follows:

$$\hat{y}_{ik'}^* = \hat{f}(\vec{x}_{ik'}, \alpha_i, \alpha_i^*)$$
$$= \sum_{k=1}^{nd} (\alpha_{ik} - \alpha_{ik}^*) K\langle \vec{x}_{ik'}, \vec{x}_{ik}\rangle + b_i \,. \tag{27}$$

Where the $\hat{y}_{ik'}^*$ is the estimated output of the k'th vector in ith cluster which is estimated using the support vector regression function with the Lagrange multipliers of the ith cluster. The augmented kernel matrix denotes the kernel mapping of the augmented input matrix (as described in SFF-LSE approach) where the membership values and their transformations are used as additional input variables. After the optimization algorithm finds the optimum Lagrange multipliers, one can estimate the output value of each data point in each cluster using Eq. (27).

The inference structure of SFF-SVM is adapted from the Special Fuzzy Functions with least squares where one can estimate a single output of a data point (see Eq. (15) by taking the membership value weighted averages of its output values calculated for each cluster using Eq. (27).

Future of Fuzzy System Models

In the future, fuzzy system models are expected to be structured by type 2 fuzzy sets. For this purpose, we next present basic definitions.

Basic Definitions

Definition 1 A type 2 fuzzy set $\tilde{A}$ on universe of discourse, x, is a fuzzy set which is characterized by a fuzzy membership function, $\tilde{\mu}_A(x)$, where $\tilde{\mu}_A(x)$ is a mapping as shown below:

$$\tilde{\mu}_A(x)\colon X \to [0, 1]^{[0,1]} \,. \tag{28}$$

Then type 2 fuzzy set, $\tilde{A}$, can be characterized as follows:

$$\tilde{A} = \{(x, (u, f_x(u)))|x \in X, \, u \in J_x \subseteq [0, 1]\} \,. \tag{29}$$

Where u is defined as the primary membership value, $J_x \subseteq [0, 1]$ is the domain of u and $f_z(u)$ is the secondary membership function. An alternative definition of type 2 fuzzy set $\tilde{A}$, used by Mendel [28] and inspired by Mizumoto and Tanaka [30], can be given as follows:

Given that x is a continuous universe of discourse, the same type 2 fuzzy set can be defined as:

$$\tilde{A} = \int_{x \in X}\left[\int_{u \in J_x} f_x(u)/u\right]/x \,. \tag{30}$$

And for discrete case, the same type 2 fuzzy set can be defined as:

$$\tilde{A} = \sum_{x \in X}\left[\sum_{u \in J_x} f_x(u)/u\right]/x \,. \tag{31}$$

The secondary membership function, $f_x(u)$, can be defined as follows:

Definition 2 Secondary membership function, $f_x(u)$, is a function that maps membership values of universe of discourse x onto unit interval $[0, 1]$. Thus, $f_x(u)$ can be characterized as follows:

$$f_x(u): J_x \to [0, 1], \forall u \in J_x \subseteq [0, 1], \quad \forall x \in X. \quad (32)$$

With the secondary membership function defined as above, the membership function of type 2 fuzzy set $\tilde{A}$, $\tilde{\mu}_A(x)$, can then be written for continuous and discrete cases, respectively, as follows:

$$\tilde{\mu}_A(x) = \int_{u \in J_x} f_x(u)/u, \quad \forall x \in X, \quad (33)$$

$$\tilde{\mu}_A(x) = \sum_{u \in J_x} f_x(u)/u, \quad \forall x \in X. \quad (34)$$

An Interval Valued Type 2 Fuzzy Set (IVT2FS), which is a special case of type 2 fuzzy set, can be defined as follows:

Definition 3 Let $\tilde{A}$ be a linguistic label with type-2 membership function on the universe of discourse of base variable x, $\tilde{\mu}_A(x): X \to f_x(u)/u, u \in J_x, J_x \subseteq [0, 1]$. The following condition needs to be satisfied in order to consider $\tilde{\mu}_A(x)$ as an interval value type 2 membership functions:

$$f_x(u) = 1, \forall x \in X, \forall u \in J_x, J_x \subseteq [0, 1]. \quad (35)$$

Thus, the interval valued type 2 membership function is a mapping as shown below:

$$\tilde{\mu}_A(x) : X \to 1/u, u \in J_x, J_x \subseteq [0, 1]. \quad (36)$$

General Structure of Type 2 Fuzzy System Models

In a series of papers, Mendel, Karnik and Liang [22,23,24,25] extended traditional type 1 inference methods such that these methods can process type 2 fuzzy sets. These studies were explained thoroughly by Mendel in [28]. The classical Zadeh and Takagi–Sugeno type 1 models are modified as type 2 fuzzy rule bases (T2Z-FR and T2TS-FR,), respectively, as follows:

$$\underset{i=1}{\overset{c^*}{\text{ALSO}}} \left[\text{IF} \underset{j=1}{\overset{NV}{\text{AND}}} \left(x_j \in X_j \text{ isr } \tilde{A}_{ji} \right) \right.$$
$$\left. \text{THEN } y \in Y \text{ isr } \tilde{B}_i \right], \quad (37)$$

$$\underset{i=1}{\overset{c^*}{\text{ALSO}}} \left[\text{IF} \underset{j=1}{\overset{NV}{\text{AND}}} \left(x_j \in X_j \text{ isr } \tilde{A}_{ji} \right) \right.$$
$$\left. \text{THEN } y_i = a_i x^{\text{T}} + b_i \right]. \quad (38)$$

Mendel, Karnik and Liang [22,23,24,25] assumed that the antecedent variables are separable (i. e., non-interactive). After formulating the inference for a full type 2 fuzzy system model, Karnik et al. [22,23,24,25] simplified their proposed methods for the interval values case. In order to identify the structure of the IVT2FS, it was assumed that the membership functions are Gaussian. A clustering method was utilized to identify the mean parameters for the Gaussian functions. However, the clustering method has not been specified. It was assumed that the standard error parameters of the Gaussian membership functions are exactly known. The number of rules was assigned as eight due to the nature of their application. However, the problem of finding the suitable number of rules was not discussed in the paper.

Liang and Mendel [26] proposed another method to identify the structure of IVT2FS. It has been suggested to initialize the inference parameters and to use steepest-descent (or other optimization) method in order to tune these parameters of an IVT2FS. Two different approaches for the initialization phase have been suggested in [26]. Partially dependent approach utilizes a type 1 fuzzy system model to provide a baseline for the type 2 fuzzy system model design. Totally independent approach starts with assigning random values to initialize the inference parameters. Liang and Mendel [26] indicated that the main challenge in their proposed tuning method is to determine the active branches.

Mendel [28] indicated that several structure identification methods, such as one-pass, least-squares, back-propagation (steepest descent), singular value-QR decomposition, and iterative design methods, can be utilized in order to find the most suitable inference parameters of the type 2 fuzzy system models. Mendel [28] provided an excellent summary of the pros and cons of each method.

Several other researchers such as Starczewski and Rutkowski [37], John and Czarnecki [20,21] and Chen and Kawase [10] worked on T2-FSM. Starczewski and Rutkowski [37] proposed a connectionist structure to implement interval valued type 2 fuzzy structure and inference. It was indicated that methods such as, back propagation, recursive least squares or Kalman algorithm-based methods, can be used in order to determine the inference parameters of the structure. John and Czarnecki [20,21] extended the ANFIS structure such that it can process the type 2 fuzzy sets.

All of the above methods assume non-interactivity between antecedent variables. Thus, the general steps of the inference can be listed as: fuzzification, aggregation of the antecedents, implication, and aggregation of the consequents, type reduction and defuzzification.

With the general structure of type 2 fuzzy system models and inference techniques, we next propose discrete interval valued type 2 rule base structures.

Discrete Interval Valued Type 2 Fuzzy Sets (DIVT2FS)

In Discrete Interval Valued Type 2 Fuzzy Sets (DIVT2FS) [48] the domain of the primary membership is discrete and the secondary membership values are fixed to 1. Thus, the proposed method is utilizing discrete interval valued type 2 fuzzy sets in order to represent the linguistic values assigned to each fuzzy variable in each rule. These fuzzy sets can be mathematically defined as follows:

$$\tilde{A} = \int_{x \in X} \left[\sum_{u \in J_x} 1/u \right] /x \, , \tag{39}$$

where $x \in X \subseteq \Re, u \in J_x = \{J_{xr}\}, \ r = 1, \ldots, NM$, x is a system variable with continuous domain, u is the primary membership value with discrete domain and J_{xr} is the rth primary membership value associated with x. The discrete interval valued type 2 fuzzy set $\tilde{A}$, can be considered as the union of type 1 fuzzy sets $a^r, r = 1, \ldots, nm$, where nm is the number of embedded type 1 fuzzy sets that form the "discrete interval valued type 2 fuzzy set $\tilde{A}$". (for the purposes of this paper, we assume that nm is known. Currently we are conducting further research to determine nm at the Knowledge-Intelligence Laboratory, University of Toronto). Consequently, the membership function of discrete interval type 2 fuzzy set $\tilde{A}$, $\mu_{\tilde{A}}(x)$, can be represented as a collection of type 1 membership functions as follows:

$$\mu_{\tilde{A}}(x) = \{\mu_A^r(x)\} \, , r = 1, \ldots, NM \, , \tag{40}$$

where nm is the number of discrete membership values assigned to each value of system variable x, $\mu_A^r(x)$ is the membership function associated with rth embedded type 1 fuzzy set, which can be defined as follows:

$$\mu_A^r(x) \colon X \to J_{xr}, x \in X \subseteq \Re \, . \tag{41}$$

One of the goals of this study is to eliminate the non-interactivity assumption used in the existing type 2 fuzzy system models. Hence, we have extended the fuzzy rule base structure proposed by Delgado et al. [12], Babuska et al. [1] and Uncu and Türkşen [48]. Thus, the proposed Discrete Interval Type 2 Zadeh Fuzzy Rule base (DIT2-Z-FR) structure can be written as follows:

$$R \colon \left\{ \underset{i=1}{\overset{c^*}{\mathrm{ALSO}}} \left(\text{IF } x \in X \text{ isr } A_i^r \text{ THEN } y \in Y \text{ isr } B_i^r \right) \right\},$$
$$r = 1, \ldots, NM \, , \tag{42}$$

where, A_i^r is the rth embedded type 1 multidimensional fuzzy set associated with the antecedent of the ith rule. A_i^r will be represented with the membership function $\mu_i^r(x)$ in the membership value domain. Similarly, B_i^r is the rth embedded type 1 fuzzy set associated with the consequent of the ith rule. B_i^r will be represented with the membership function $\mu_i^r(y)$ in the membership value domain.

The rule base structure given in (42) can be thought as a collection of embedded type 1 fuzzy rule bases. Hence, the proposed fuzzy rule base structure is named as Discrete Interval Type 2 Fuzzy Rule base (DIT2-Z-FR) structure.

The other well-known fuzzy rule base structures are also extended to type 2 by using the above idea. Hence, Discrete Interval Valued Type 2 Takagi–Sugeno Fuzzy Rule base (DIT2-TS-FR) can be written respectively as follows:

$$R \colon \left\{ \underset{i=1}{\overset{c^*}{\mathrm{ALSO}}} \left(\text{IF } x \in X \text{ isr } A_i^r \text{ THEN } y = a_i^r x^{\mathrm{T}} + b_i^r \right) \right\},$$
$$r = 1, \ldots, NM \, , \tag{43}$$

where, $a_i^r = \left(a_{i,1}^r, \ldots, a_{i,NV}^r \right)$ is the regression coefficient vector associated with the ith rule of the rth embedded type 1 fuzzy system model and b_i^r is the scalar associated with the ith rule of the rth embedded type 1 fuzzy system model. As one can observe the Takagi–Sugeno structure is not only extended by representing the antecedent fuzzy sets with discrete type 2 fuzzy sets but also by letting uncertainty in crisp inference parameters in consequents. Thus, the problem of building type 2 fuzzy system models is reduced to finding embedded type 1 fuzzy system models.

Discrete Interval Valued Type 2
"Special Fuzzy Functions", (SFF)

In a similar manner to Discrete Interval Valued Type 2 Fuzzy Rule bases, we could construct Discrete Interval Valued Type 2 "Special Fuzzy Functions" for the cases of SFF-LSE and SFF-SVR. For example, the function of a single input single output model, which includes only the membership values as the additional input variable, $Y_i^r = \beta_{i0}^r + \beta_{i1}^r \Gamma_i^r + \beta_{i2}^r X, r = 1, \ldots, NM$ that represents the rth "Special Fuzzy Function" to the ith interactive (joint) type 2 cluster in $(Y_i^r, \Gamma_i^r, X^r), i = 1, \ldots, c^*, r = 1, \ldots, NM$ space, would be estimated with SFF-LSE approach as follows:

$$\beta_i^{r*} = \left(X_i^{r'T} X_i^{r'} \right)^{-1} \left(X_i^{rT} Y_i^r \right), \quad r = 1, \ldots, NM \, , \tag{44}$$

where $\beta_i^{r*} = (\beta_{i0}^r*, \beta_{i1}^r*, \beta_{i2}^r*)$ and $X_i^{r'} = [1, \Gamma_i^r, X]$, provided the inverse of covariance, $(X_i^{r'T} X_i^{r'})^{-1}$, exists. The

estimate of y_i would be obtained as:

$$Y_i^{r^*} = \beta_{i0}^{r^*} + \beta_{i1}^{r^*} \Gamma_i^r + \beta_{i2}^{r^*} X . \tag{45}$$

The single output value is calculated using each output value, one from each cluster, and weighting them with their corresponding membership values as follows:

$$Y_i^{r^*} = \frac{\sum_{i=1}^{c^*} \gamma_i^r Y_i^{r^*}}{\sum_{i=1}^{c^*} \gamma_i^r}, \quad i = 1, \dots, c^*, \quad r = 1, \dots, NM . \tag{46}$$

The development of such Discrete Interval Valued Type 2 "Special Fuzzy Functions" with LSE will be studied in our future investigations. In a similar manner, we propose to develop Discrete Interval Valued Type 2 "Special Fuzzy Functions" with SVM methodology.

Case Study Applications

In order to test the proposed model performances as opposed to fuzzy rule base systems three input-output datasets are considered in this investigation. These are:

(i)   Daily price of a stock in stock market.
(ii) Customer Income Prediction model for a major bank.
(iii) The amount of chemicals for a desulphurization process for a steel processing company.

The specifications of each datasets are displayed in the following parts:

Daily Stock Price Dataset

Daily stock price dataset comprises of the daily trend data of stock prices. This dataset was introduced by Sugeno and Yasukawa [38]. The same dataset has been used in various other studies one of which compares six different fuzzy reasoning methods using this dataset.

Out of 100 observations, 50 of them are used for the training purposes and the other 50 was hold-out for testing purposes. There were originally 11 input variables and single output variable in the dataset [48]. Preliminary input selection was applied using Random Forests (RF) method, [48] which estimates variable importance. Based on the results of RF method, only 4 of input variables i. e., x_2, x_4, x_8, x_{10}, were found to have importance on the output variable. The rest of the variables had insignificant effect on the output.

Income Prediction Dataset

The purpose of the Income Prediction Model was to predict the income of the future customers based on the current customer information and 1996 year census data. There were more than 200 variables and hundred of thousands of customers. According to business needs, the dataset was partitioned into 9 different parts based on age, number of different types of investment accounts hold by the customer and different regions of residency. In this study, only one partition was investigated.

We have only selected 10% of the i.i.d. data samples to do our research on using only single partition explained above. The data was cleaned form the outliers using the expert's knowledge and 900 training and 900 testing observations are selected randomly. The dataset was comprised of 11 input variables [48]. Based on the correlation analysis, 3 input variables were discarded from the dataset resulting in 8 input variables. There were no census variables among the selected input variables.

Desulphurization Dataset

A torpedo car desulphurization facility removes sulfur from hot metal leaving the blast furnaces before it is sent to the next process. Generally, desulphurization is carried out by injecting two different powered reagents directly into the hot metal via a lance. The reagents react with the sulfur in the hot metal and residue, which is rich in sulfur, is separated from the iron.

The aim of the data-mining project was to determine the right amounts of the reagents to be added into the hot metal. These regents are expensive materials and precise estimation is required. There are vast quantities of data available on the desulphurization process, which has various characteristics. The original input and output variables are shown in [27]. There are 750 training and 900 verification samples used in the experiments. Based in the variable selection using random forest regression method, only 5 variables are found to be important.

Experimental Design

Using the three different input-output datasets, we build four different fuzzy system model structures, SFF-LSE, SFF-SVM, SY-FRB and TS-FRB. In order to keep the consistency between each model structure, the same training and testing datasets are used for the four fuzzy system models with the same input variables. The categorical variables are transformed into probabilities using logistic regression and are used as additional inputs only in Income Prediction and Desulphurization Datasets in all of the four models.

Sugeno–Yasukawa Models and Takagi–Sugeno Models

The proposed FSM models are compared to two well known Fuzzy Rule Base Models: (i) Sugeno and Yasukawa's fuzzy logic based approach using Partition Type Fuzzy Model, **SY-FRB,** [38] (ii) Takagi and Sugeno's fuzzy system modeling approach, **TS-FRB** [39]. In Sugeno and Yasukawa's FSM approach, they use linguistic variables for both the consequent and the antecedent part of the fuzzy rules and the system learns all inference parameters from the data without the expert intervention. The variable selection method defined in their paper is not applied to these 3 datasets in order to compare the models on the same basis.

In Takagi and Sugeno's FRB (TS-FRB) structure [39], they assume that the antecedent membership functions are to be characterized with triangular membership functions. In their approach, each input variable space is assumed to be partitioned into two clusters and logical connective AND is taken as MIN. Then, the structure identification problem is just to identify the regression equation coefficients for each rule and the antecedent parameters for each input variable in each rule. Researchers proposed several structure identification methods to identify the membership functions from the data, e. g., Delgado et al. [12], Babuska and Verbruggen, etc. In this paper we have used Babuska et al.'s [1] modified Takagi–Sugeno study where the membership functions of the antecedents are identified using fuzzy c-means clustering and projected onto each input vector. The degree of fulfillment of each rule is then calculated using a t-norm operator, i. e. product. Only one aggregate input membership function is identified for each rule. The inference parameters are same as the traditional Takagi–Sugeno inference method [39].

Special Fuzzy Functions Models (SFFM)

In this paper, the modeling performance from the Special Fuzzy Functions with LSE and SVM models, SFF-LSE and SFF-SVM, are compare to the Fuzzy Rule Base structures. Instead of using cluster validity indices to select the optimum model parameters, we measured the optimum model based on the best model performance using RMSE by applying a grid search for each parameter. Note that fuzzy functions with LSE system models have 2 parameters, which are the FCM parameters, i. e., degree of fuzziness and cluster size. Special Fuzzy Functions with SVR models have 4 parameters: FCM parameters and the SVR parameters which are *C-regularization* and *ε-insensitive value (epsilon).*

In both of these special fuzzy function models, membership values and their exponential transformations are used as additional input variables. We applied the LIB-SVM program [6] within our special fuzzy function codes for support vector optimization in estimating the "Special Fuzzy Functions". We chose Gaussian RBF as the kernel function, $K(x, x') = exp(-\gamma \|x - x'\|^2)$ in all the experiments and the default values of the kernel parameters in LIBSVM [10] are used. Recall that the SVR regression has two parameters that is set by the user: ε-insensitive zone (*epsilon or* ε) and the regularization parameter, C). The parameters of SVM regression, C and ε, are generated by the grid search, i. e., $C = \{2^{-3}, 2^{-1}, \ldots, 2^3, 2^5\}$, $epsilon(\varepsilon) = \{0.1, \ldots, 0.5\}$, as well as the FCM parameters, i. e., $c = 3, 5, \ldots, 10$, $m = 1.1, \ldots, 2.5$ in all the experiments. Hence, for the fuzzy functions with LSE models 2 parameters are specified for each model, and for the fuzzy functions using SVM models, 4 parameters, *m, c, C,* and γ are determined using a grid search where the model performance of each model is determined using Root Mean Square Error of the models as follows:

$$\text{RMSE} = \sqrt{\frac{1}{n} \sum_{i=1}^{N} (y_i - \hat{y}_i)^2} \qquad (47)$$

y_i, and $\hat{y}_i$ are the actual and estimated output values of a single observation, N is the total number of observations in the dataset.

Experimental Model Results

The 4 fuzzy system models are applied on the daily stock price of a stock market, income prediction, and reagents estimation in desulphurization process datasets. The results are displayed in Tables 1,2 and 3.

Special Fuzzy Function (SFF) models, when estimated with either *SVM or LSE* algorithms, show better generalization than the fuzzy rule base models. The fuzzy function models can increase the model performance by up to 35% depending on the dataset.

Table 4 displays the optimum parameters of the models from each experiment whose results are displayed in Tables 1–3. In Table 4, *C-reg* indicates the regularization parameter, ε is the ε-insensitive region (epsilon), *m* refers

Fuzzy System Models Evolution from Fuzzy Rulebases to Fuzzy Functions, Table 1
Daily Stock Price of a Stock in Stock Market

	SFF-SVM	SFF-LSE	TS-FRB	SY-FRB
RMSE(train)	2.43	3.82	2.76	7.16
RMSE(test)	3.64	5.61	5.68	9.93

Fuzzy System Models Evolution from Fuzzy Rulebases to Fuzzy Functions, Table 2
Income Prediction Dataset*

	SFF-SVM	SFF-LSE	TS-FRB	SY-FRB
RMSE(train)	0.40	0.52	0.49	0.58
RMSE(test)	0.64	0.64	0.80	0.70

* The RMSE values are calculated from standardized output values.

Fuzzy System Models Evolution from Fuzzy Rulebases to Fuzzy Functions, Table 3
Reagent Estimation for Desulphurization Process

	SFF-SVM	SFF-LSE	TS-FRB	SY-FRB
Reagent1				
RMSE(train)	30	40	35	69
RMSE(test)	42	45	45	72
Reagent2				
RMSE(train)	4.80	6.49	5.62	10.01
RMSE(test)	6.59	7.19	7.19	10.80

Fuzzy System Models Evolution from Fuzzy Rulebases to Fuzzy Functions, Table 4
Optimum Model Parameters of three datasets

Dataset	Model Type	Optimum model Parameters
Daily Stock Price	SFF-SVM	$C - reg = 32$, $\varepsilon = 0.2$, $c = 8$, $m = 1.9$, $\#sv = 28(\sigma_{sv} = 1.5)$
	SFF-LSE	$c = 8$, $m = 1.6$
Income Prediction	SFF-SVM	$C - reg = 64$, $\varepsilon = 0.2$, $c = 7$, $m = 1.2$, $\#sv = 264(\sigma_{sv} = 5.7)$
	SFF-LSE	$c = 8$, $m = 1.6$
Desulphurization	SFF-SVM	Reagent 1: $C - reg = 64$, $\varepsilon = 0.2$, $c = 7$, $m = 1.4$, $\#sv = 121(\sigma_{sv} = 4.8)$ Reagent 2: $C - reg = 64$, $\varepsilon = 0.2$, $c = 7$, $m = 1.5$, $\#sv = 101(\sigma_{sv} = 7.7)$
	SFF-LSE	Reagent 1: $c = 5$, $m = 1.4$ Reagent 2: $c = 6$, $m = 1.5$

to the degree of fuzziness (weighting exponent) of the fuzzy c-mean clustering algorithm, c indicates the number of cluster, and $\#sv$ refers to the average number of support vectors from each support vector regression model build for each cluster of the *SFF-SVM* models. In order to show the dispersion of the number of support vectors among each cluster we also included the standard deviation (σ_{sv}) of the support vectors of the optimum models.

The grid search algorithms applied in this paper try to find the best RMSE value from training data in each experiment and assign these parameters as the optimum model parameters. The algorithm searches for the minimum regression error. Then, verification dataset output is inferred using the optimum parameters. The issue with these grid search algorithms is that, sometimes, the models get stuck in the local minimum which is smaller than the global minimum and this might cause generalization problems. An example to this concept is shown in income prediction dataset (Table 2). The model parameters best fit to the training data when FF-SVM is used but this causes generalization problems. It should also be reminded that, when there is a linear relationship between the inputs and the output, then LSE model performances will be as good as the other model performances. On the other hand, FF-LSE models, in three of the datasets, show more reliable results than the SVM models. One should run both models and determine the optimum model parameters after observing the results from both models.

Conclusions and Future Directions

In this paper, we have outlined basic well known three type 1 system models of the recent past and suggested that type 2 fuzzy system models are likely to be studied more extensively in the future. In particular, we have reviewed: (1) Type 1 Fuzzy Rule bases and (2) Type 1 "Special Fuzzy Functions". Furthermore we discussed the Type 2 Fuzzy Rule bases. But we left the Type 2 "Special Fuzzy Functions" for a future study after providing a structure for their development . As well, we have demonstrated that "Type 1 Special Fuzzy Functions" provide better results then "Type 1 Fuzzy Rule Base" models in three specific case studies.

Bibliography

Primary Literature

1. Babuska R, Verbruggen HB (1997) Constructing fuzzy models by product space clustering. In: Hellendoorn H, Driankov D (eds) Fuzzy model identification: selected approaches. Springer, Berlin, pp 53–90
2. Bezdek JC (1973) Fuzzy mathematics in pattern classification. Ph.D Thesis, Applied Mathematics Center. Cornell University, Ithaca
3. Celikyilmaz A, Türkşen IB (2007) Fuzzy functions with support vector machines. Inf Sci 177:5163–5177
4. Celmins A (1987) Least squares model fitting to fuzzy vector data. Fuzzy Sets Syst 22:245–269
5. Celmins A (1987) Multidimensional least squares model fitting of fuzzy models. Math Model 9:669–690
6. Chang C, Lin C (2001) LIBSVM: a library for support vector machines. Software available at http://www.csie.ntu.edu.tw/~cjlin/libsvm

7. Chang PT, Lee ES (1994) Fuzzy linear regression with spreads unrestricted in sign. Compt Math Appl 28:61–71

8. Chang YHO, Ayyub BM (1993) Reliability analysis in fuzzy regression. In: Proc. Annual Conf. of NAFIPS'93. Allentown, IEEE, New York, pp 93–97

9. Chen MS, Wang SW (1999) Fuzzy clustering analysis for optimizing fuzzy membership functions. Fuzzy Sets Syst 103(2):239–254

10. Chen Q, Kawase S (2000) On fuzzy-valued fuzzy reasoning. Fuzzy Sets Syst 113:237–251

11. Chiu S (1994) Fuzzy model identification based on cluster estimation. J Intell Fuzzy Syst 2(3)267–278

12. Delgado M, Gomez–Skermata AF, Martin F (1997) Rapid prototyping of fuzzy models". In: Hellendoorn H, Driankov D (eds) Fuzzy model identification: selected approaches. Springer, Berlin, Germany, pp 53–90

13. Demirci M (1999) Fuzzy functions and their fundamental properties. Fuzzy Sets Syst 106:239–246

14. Demirci M (2003) Foundations of fuzzy functions and vague algebra based on many-valued equivalence relations, Part I: fuzzy functions and their applications. IJ Gen Syst 32:123–155

15. Demirci M, Recasens J (2004) Fuzzy groups, fuzzy functions and fuzzy equivalence relations. Fuzzy Sets Syst 144:441–458

16. Diamond P (1998) Fuzzy least squares. Inf Sci 46:141–157

17. Emami MR, Türkşen IB, Goldenberg AA (1998) Development of a systematic methodology of fuzzy logic modeling. IEEE Tran Fuzzy Syst 63(3):346–361

18. Hathaway RJ, Bezdek JC (1993) Switching regression models and fuzzy clustering. IEEE Trans Fuzzy Syst 1(3):195–203

19. Jang JSR (1993) Anfis: adaptive-network-based fuzzy inference systems. IEEE Trans Syst Man Cybern 23(3):665–685

20. John RI, Czarnecki C (1998) A type 2 adaptive fuzzy inference system. In: Proc. IEEE Conf. Systems, Man and Cybernetics, vol 2. IEEE, New York, pp 2068–2073

21. John RI, Czarnecki C (1999) An adaptive type-2 fuzzy system for learning linguistic membership grades. In: Proc. IEEE International Fuzzy Systems Conference, vol 3. IEEE, New York, pp 1552–1556

22. Karnik NN, Mendel JM (1998) Introduction to type-2 fuzzy logic systems. In: Proc. IEEE Conf. On computational intelligence, vol 2. IEEE, New York, pp 915–920

23. Karnik NN, Mendel JM (1998) Type-2 fuzzy logic systems: type reduction. In: Proc. IEEE Conf. On Systems, Man and Cybernetics, vol 2. IEEE, New York, pp 2046–2051

24. Karnik NN, Mendel JM, Liang Q (1999) Type-2 fuzzy logic systems. IEEE Trans On Fuzzy Syst 7(6):643–658

25. Karnik NN, Mendel JM (2000) Applications of type-2 fuzzy logic systems: handling the uncertainty associated with surveys. In: Proc. IEEE Conf. On Fuzzy Systems, vol 3. pp 1546–1551

26. Liang Q, Mendel JM (2000) Interval type-2 fuzzy logic systems: theory and design. IEEE Trans On Fuzzy Syst 8(5):535–550

27. Mamdani EH, Assilian S (1981) An experiment in linguistic synthesis with a fuzzy logic controller. In: Mamdani EH, Gains BR (eds) Fuzzy Reasoning and Its Applications. Academic Press, New York

28. Mendel JM (2001) Uncertain rule-based fuzzy logic systems: introduction and new directions. Prentice, Upper Saddle River

29. Mizumoto M (1989) Method of fuzzy inference suitable for fuzzy control. J Soc Instrum Control Eng 58:959–963

30. Mizumoto M, Tanaka K (1976) Some properties of fuzzy sets of type 2. Inf Control 31:312–340

31. Nakanishi H, Türkşen IB, Sugeno M (1993) A review and comparison of six reasoning methods. Fuzzy Sets Syst 57:257–295

32. NR Pal, Bezdek JC (1995) On cluster validity for the fuzzy c-means model. IEEE Trans Fuzzy Syst 3(3)370–379

33. Rutkowska D (2002) Type 2 fuzzy neural networks: an interpretation based on fuzzy inference neural networks with fuzzy parameters. In: Proc. IEEE Conf. On Fuzzy Systems, vol 2. IEEE, New York, pp 1180–1185

34. Savic D, Pedryzc W (1991) Evolution of fuzzy linear regression models. Fuzzy Sets Syst 39:51–63

35. Smola AJ, Scholkopf B (1998) A tutorial on support vector regression. NeuroColT2 Technical Report Series, NC2-Tr-1998–030

36. Sproule BA, Bazoon M, Shulman KI, Turkşen IB, Naranjo CA (2000) Fuzzy logic pharmacokinetic modeling: an application to lithium concentration prediction. Clinical Pharmacology Therapy 62:29–40

37. Starczewski J, Rutkowski L (2002) Connectionist structures of type 2 fuzzy inference systems. In: Wyrzykowski R et al (eds) PPAM 2001, LCNS 2328. Springer, Heidelberg, pp 634–642

38. Sugeno M, Yasukawa T (1993) A fuzzy logic based approach to qualitative modeling. IEEE Trans Fuzzy Syst 1:7–31

39. Takagi T, Sugeno M (1985) Fuzzy identification of systems and its applications to modeling and control. IEEE Trans Syst Man Cybern SMC-15(1):116–132

40. Tanaka H, Ishibuchi H, Yoshikawa S (1995) Exponential possibility regression analysis. Fuzzy Sets Syst 69:305–318

41. Tanaka H, Vegima S, Asai K (1982) Linear regression analysis with fuzzy model. IEEE Trans Syst Man Cybern SMC-2:903–907

42. Türkşen IB (1986) Interval valued fuzzy sets based on normal forms. Fuzzy Sets Syst 20:191–210

43. Türkşen IB (1992) Interval-valued fuzzy sets and 'compensatory AND'. Fuzzy Sets Syst 51:295–307

44. Türkşen IB (1995) Fuzzy normal forms. Fuzzy Sets Syst 69:319–346

45. Türkşen IB (2002) Type 2 representation and reasoning for cww. Fuzzy Sets Syst 127:17–36

46. Türkşen IB (2008) Fuzzy Functions with LSE. Applied Soft Computing 8(3):1178–1182

47. Uncu Ö, Türkşen IB (2007) A novel feature selection approach: combining feature wrappers and filters. Inf Sci 177:449–466

48. Uncu Ö, Türkşen IB (2007) Discrete Interval Type 2 Fuzzy System Models Using Uncertainty in Learning Parameters. IEEE Fuzzy Syst 15(1):90–106

49. Vapnik NV (1998) Statistical learning theory. Wiley, New York

50. Zadeh LA (1975) The concept of a linguistic variable and its application to approximate reasoning. Inf Sci 8:199–249

51. Zimmermann HJ, Zysno P (1980) Latent connectives in human decision-making. Fuzzy Sets Syst 4:37–51

Books and Reviews

Kilic K (2002) A proposed fuzzy system modeling algorithm with an application in pharmacokinetic modeling. Ph.D Thesis, Department of Mechanical and Industrial Engineering. University of Toronto, Toronto