Acta Physica Academiae Scientiarum Hungaricae, 53 (1—2), pp. 263—277 (1982)

COMPUTER MODELS FOR SOME ASPECTS OF EVOLUTION*

F. DEÁK and G. MARX

DEPARTMENT OF ATOMIC PHYSICS, ROLAND EÖTVÖS UNIVERSITY
1088 BUDAPEST, HUNGARY

> « L' esprit va dans son travail, de son désordre à son ordre. Il importe qu'il se conserve jusqu'à la fin des ressources de désordre, et que l'ordre qu'il a commencé de se donner ne se lie pas si complètement, ne lui soit pas un si rigide maître, qu'il ne puisse le changer et user de sa liberté initiale. »
>
> (Paul Valéry)

In this model a certain sequence of characters is supposed to have a "survival advantage" compared to any random sequence. The multiplication happens with errors of a fixed probability. By making use of this model, the evolutionary role of sequence length, life expectancy and mutation rate has been studied.

Evolution

Lifeless objects wear out, they decay, their manifold heads towards complete disorder. A population of living beings does not wear out, it is even able to improve itself. If one looks for the origin of this qualitative difference, it turns out that the essential threshold of life is the ability of self reproduction. In a noisy environment not all offsprings of a living being are identical, so there is always a spectrum of genetic information present in the population. In the actual environment the allele (the variant) with the highest survival value and the highest fecundity multiplies the fastest. Other alleles are decimated or eliminated by natural selection, but they are always reproduced by the mutations of the fittest allele. If the environment changes, another allele of the spectrum will fit the new environment in the best way, and this allele then takes over. In this way the population will adjust itelf to new conditions. In this information-theoretical sense "life is the undertaking of an information carrier, to produce as much copies as possible". This is an updated version of the saying that "the hen is an undertaking of an egg to make more eggs". From this point of view three layers of natural phenomena can be distinguished:

* Dedicated to Prof. I. Tarján, pioneer of exact approach to biology, on his 70th birthday.

I. The behaviour of material objects with few degrees of freedom can be described by deterministic equations of motion (equations of Newton, Maxwell, Schrödinger, Dirac...). These are symmetric with respect to time reversal, the motions described by their solutions are reversible. (E.g a mass point, a mathematical pendulum, a rigid body, a lone electron.)

II. The behaviour of an aggregate of matter with very many degrees of freedom is irreversible due to statistics. It is characterized by increasing disorder, by wearing out, by dissipating any concentrated energy from a single degree of freedom to many degrees of freedom of the aggregate. (Second Law of thermodynamics.) Examples: the universe or any piece of real matter.

III. A population of self reproducing structures is able to improve itself in a noisy environment, to increase the survival chances and multiplication rate of the member structures spontaneously. Natural selection eliminates the deficient copies, so the population of self reproducing structures does not wear out. In fact just the opposite is true: by making use of random mutations, the population is able to adjust itself to a changing environment. Its fate is irreversible as well, but now in a positive sense. (Darwinian evolution.) In accordance with the Second Law, the necessary condition for spreading and developing is a steady flow of free enthalpy to the self reproducing structures, so that the increase of entropy in the environment overcompensates the increase of organisation and information within the population.

In this information-theoretical sense the main task of life science is to understand the spontaneous origin, spreading and increase of the genetic information. No wonder this challenge has attracted the attention not only of biologists, but of chemists, physicists, mathematicians and computer scientists as well [1].

The first computer model

Since the discovery of the genetic code, the mechanism for the evolution of genetic information has been sought by a number of authors [2]. To understand the evolution of genetic information, Manfred Eigen and Peter Schuster have introduced a very simple, but very illustrative model [3]. Let us start with a random sequence of characters. (E.g. BAK GEVLNT GUPIF LESTKKM.) The given environment has been supposed to have the highest preference for a "sensible sequence" of these many characters (e.g. TAKE ADVANTAGE OF MISTAKE). If the characters of an actual sequence agree with the ideal sequence in several places, this actual sequence will have a higher multiplication rate than any other random sequence with a lesser degree of coincidence. In this way the "sensible sequence" will be selected from the manifold of random sequences within a few generations. Evidently, if the copying were 100 percent faithful, a population not containing the ideal sentence from the beginning would not have a possibility of evolution towards the "most sensible sentence". But the interplay of a slight error rate per character with natural selection makes such a spontaneous increase of sensible information possible. A very high mutation ,rate, however, is harmful: it may wipe out any achievements of the evolution, before its fixation in the population with the help of natural selection. Eigen has worked with sentences of definite length, with a fixed preference for a unique "ideal sentence".He used the computer model to find the maximum size of inheritable information at a given mutation rate and to explore the advantages of a cyclic information structure (like *TAKE* ADVANTAGE OF MIS*TAKE*).

By generalizing Eigen's model, the present paper explores the influence of different internal parameters on the speed of evolution [4]. Such parameters are the length of words, their mutation rate, their life expectancy, the capability of changing the information length, etc. In this paper also we shall restrict ourselves to a fixed "target word". The case of "changing environment", that of "divergent evolution" with different target words, the possibility of a "symbiosis of information" will be explored in a subsequent paper.

Let us use an alphabet with 16 characters (e.g. the Hawaiian alphabet). A word may contain L characters. It will be compared to a previously fixed target word (e.g. LIFE). If only R letters are right ones at right places (e.g. in the case of LOTE one has $L=4$, but only $R=2$), the word is less "fit", so it will produce 2^{R+1} offsprings per generation. The "sensible word" ($R=L$) has the most offsprings: 2^{L+1}. The parent perishes after one generation, after having produced the offspring.

In general the offspring are copies of the parent word, but there is a chance for a "typing error" (mutation) at each copy. With a mutation probability μ any character of the parent word may be replaced by another, randomly chosen character of the alphabet. (E.g. an offspring of LIFE may be LOTE, another may be LIFE. The value of R may remain, may decrease or may increase.)

The number of words in the population will grow very fast from generation to generation, at the end like a fast diverging geometrical sequence. In order to control the number of digits, we express the composition of the population in percentages.

Influence of the mutation rate

Let us consider a population of words with the fixed length L. The number of the words, containing R right characters at right places is given by $y(R)$. These numbers make the $L+1$ dimensional population vector $\mathbf{y}$. The population vector will change from generation to generation due to selective multiplication and due to mutations. Its value is $\mathbf{y}_n$ in the n-th generation. For the next generation, any R-word will be replaced by its 2^{R+1} offspring (this is described by the spreading matrix S). Some of the offspring are mutants, because any character of the parent word has a chance μ of being miscopied (this is described by the mutation matrix $\mathbf{M}$). So the population vector of the next generation can be obtained by a linear transformation from the former generation:

$$\mathbf{y}_{n+1}=\mathbf{MSy}_n\,. \tag{1}$$

The multiplication is described by the $(L+1)\times(L+1)$ sized diagonal spreading matrix

$$\mathbf{S}=\begin{bmatrix} 1 & 0 & 0 & \dots & 0 \\ 0 & 2 & 0 & \dots & 0 \\ 0 & 0 & 2^2 & \dots & 0 \\ \dots & \dots & \dots & \dots & \dots \\ 0 & 0 & 0 & \dots & 2^{L+1} \end{bmatrix}. \tag{2}$$

A single mutation may replace either a right character with a wrong one ($R\rightarrow R-1$), with the probability

$$p_{R-1,R}=\mu\cdot\frac{R}{L}\cdot\frac{16-1}{16-1}.$$

(Mutation rate times the chance of affecting a right character times the chance of choosing a wrong character from the alphabet.) Or the mutation may transform a wrong character into another wrong one ($R\rightarrow R$), with a probability

$$p_{R,R}=\mu\cdot\frac{L-R}{L}\cdot\frac{16-2}{16-1}.$$

Or the mutation may transform any wrong character into the corresponding right one ($R\rightarrow R+1$), with a probability

$$p_{R+1,R}=\mu\cdot\frac{L-R}{L}\cdot\frac{1}{16-1}.$$

Substituting a right character with the right one (at a given place it is the same!) is not mutation, it is faithful copying. So the total mutation probability of one character is

$$p_{R-1,R}+p_{R,R}+p_{R+1,R}=\mu.$$

A single-character mutation can be described by the following matrix:

$$\mathbf{P}=\mu\begin{bmatrix} \frac{14}{15}\frac{L}{L}, & \frac{15}{15}\frac{1}{L}, & 0, & 0, & \dots \\ \frac{1}{15}\frac{L}{L}, & \frac{14}{15}\frac{L-1}{L}, & \frac{15}{15}\frac{2}{L}, & 0, & \dots \\ 0, & \frac{1}{15}\frac{L-1}{L}, & \frac{14}{15}\frac{L-2}{L}, & \frac{15}{15}\frac{3}{L}, & \dots \\ 0, & 0, & \frac{1}{15}\frac{L-2}{L}, & \frac{14}{15}\frac{L-3}{L}, & \dots \\ \dots, & \dots, & \dots, & \dots, & \dots \end{bmatrix} \tag{3}$$

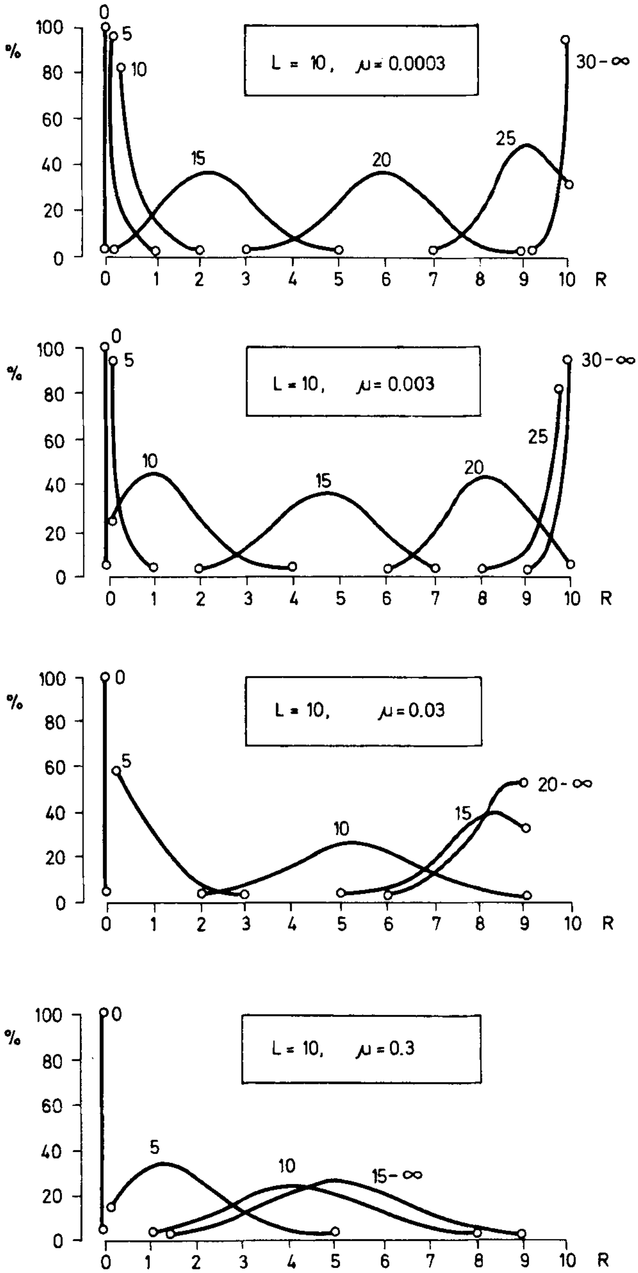
L = 10, μ = 0.0003
L = 10, μ = 0.003
L = 10, μ = 0.03
L = 10, μ = 0.3
%
100
80
60
40
20
0
0 1 2 3 4 5 6 7 8 9 10 R
0
5
10
15
20
25
30-∞
20-∞
15-∞

Fig. 1

The expected number of mutations is $\langle N \rangle = \mu L$ in a word of L characters, they may be described by a Poisson distribution. So the overall mutation matrix is constructed in the following way:

$$\mathbf{M} = \sum_{N=0}^{\infty} e^{-\mu L} \frac{(\mu L)^N}{N!} \mathbf{P}^N . \tag{4}$$

Let us now start with a 10 character word, all characters wrong ($R=0$). The evolution of the population towards the "target word" is shown in Fig. 1 for different mutation probabilities (μ). The curve of the population distribution is drawn for each fifth generation. (The serial number of the generation is indicated on the corresponding curve.) Table I shows the final population distribution. This limit is essentially the eigenvector of the product matrix $\mathbf{M} \cdot \mathbf{S}$, belonging to the highest eigenvalue.

For a fixed "target word" (in a fixed "environment") the "best fitting population" would consist of only perfect words ($R=10$), in this population the multiplication rate would be the highest: $2^{11}=2048$. This could be realized only with an exact reproduction, $\mu=0$, but this would make any evolution, any adjustment to a new environment impossible. A similar multiplication rate ($\simeq 2000$) can be realized by any value $\mu < 1\%$, so there is no practical advantage to suppress the mutations below that in this specific model.

If, however, the environment cannot be expected to last longer than 20 generations, if the target word will be changed at the 20th generation, there is no time enough to build up the limiting population by natural selection. Fig. 2 shows that within 20 generations the mutation probability $\mu=1\%$ builds up the highest multiplication rate. If the available time is so limited, a more faithful reproduction were

Table I

μ					R					
	1	2	3	4	5	6	7	8	9	10
0.	–	–	–	–	–	–	–	–	–	100
0.000 3	–	–	–	–	–	–	–	–	1	99
0.001	–	–	–	–	–	–	–	–	2	98
0.003	–	–	–	–	–	–	–	–	6	94
0.01	–	–	–	–	–	–	–	2	7	92
0.03	–	–	–	–	–	–	2	10	34	54
0.1	–	–	–	–	2	8	19	30	29	12
0.15	–	–	1	2	8	17	26	26	15	4
0.2	–	–	2	416	24	25	17	7	1	
0.3	1	4	11	20	25	21	12	5	1	–

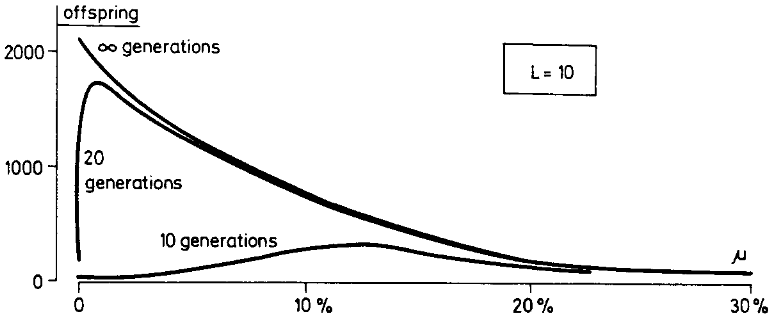

Fig. 2

even disadvantageous. In a fast changing world, in the case when the target word will change after 10 generations, the highest multiplication rate can be achieved by a larger copying error: $\mu = 10\%$. So the moral of this chapter is the following: under strictly controlled environmental conditions a small mutation rate is preferred, because this results in a narrow population spectrum. (This is the idea behind the hybrid chicken.) In a fast changing environment, however, a higher multiplication rate has a definite selective advantage. (This is a lesson learned well by the flu virus.)

Influence of the multiplication advantage

In this calculation the multiplication advantage per correct characters was taken to be 2. If one uses a higher multiplication advantage (Eigen used 2.718 and 10 per bit), it will make the final spectrum narrower. So a higher multiplication rate may compensate a larger mutation probability μ.

Influence of the longevity

In our model it has been assumed up to now that the parent word disappears immediately after having produced the first generation of her offspring. Let us change this condition: the life time of an individual word be 3 generations, i.e. any individuum bears three times. This certainly increases the overall population number, but the increase becomes negligible if the multiplication rate $\gg 1$. During the fast change of the population (R growing from 0 towards L) the surviving parents act as conservative elements, slowing down the increase of the average multiplication rate. (Table II gives the increase of the average multiplication rate in the first generations, if $L = 10$, $\mu = 0.000\,3$, the life time of a word is 1 or 3 generations. Moral: the longevity of words

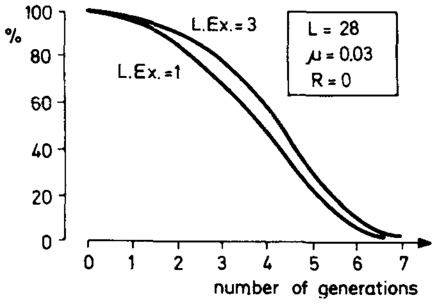

Fig. 3

Table II

Number of generations	Lifetime	
	1	3
0	2	2
1	2	2
2	2	2
3	2	2
4	2	2
4	2.02	2
6	2.02	2.02
7	2.04	2.04
8	2.10	2.08
9	2.18	2.14
10	2.40	2.32
11	2.74	2.56
12	3.75	3.24
13	4.76	4.34
	identical	

has a special value mainly for a population of small multiplication rate. (Fig. 3 shows the delayed decrease of the number of completely wrong words, $R=0$ for a special case.)

In the following chapters we shall go back to the "one word — one generation" assumption.

Influence of the length of the genetic information

Eigen has claimed [3] that in the case of $\mu L < 1$ the correct words will dominate the final population. This does not seem to be valid in our computer model. The reason for this is simple. Let y_l describe an ideal population, consisting exclusively of the

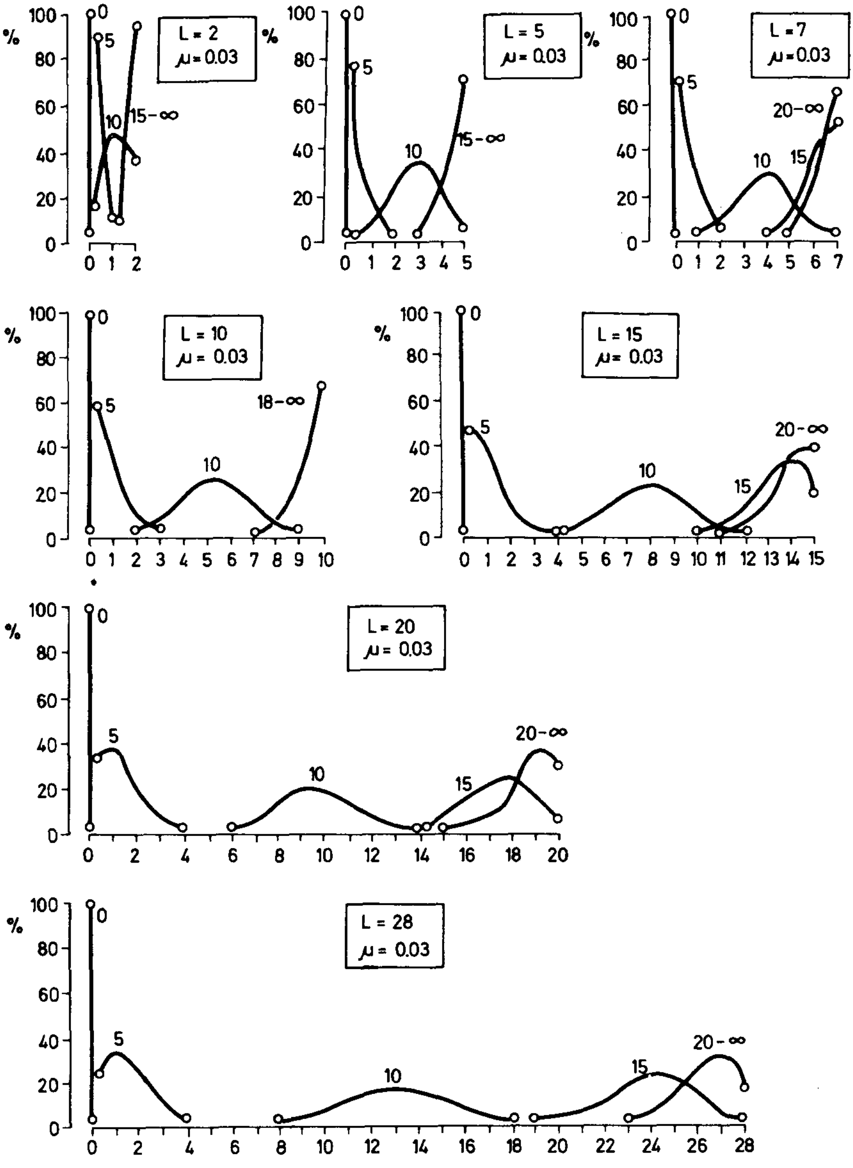

Fig. 4

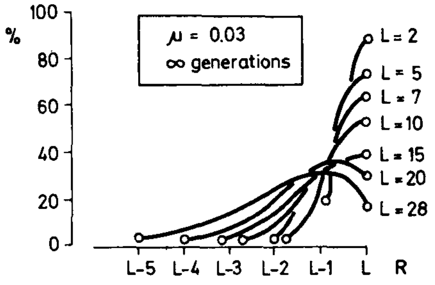

Fig. 5

correct words with $R=L$. Now a new generation cannot be better than $\mathbf{My}_{\mathrm{I}}$. If the error per character is μ, then the probability for a copy to be incorrect is $\simeq\mu L$. E.g. for $\mu=1/15$ and $L=10$ one gets $\mu L=2/3$, i.e. a good half of any generation will contain incorrect characters! It is true on the other hand that the new generation will be produced mainly by the perfect fraction, because the perfect words have at least twice as high multiplication rate as the imperfect ones.

Fig. 4 depicts the evolution of words with different lengths in time. The mutation rate per character is the same for all: $\mu=3\%$. The speed of evolution increases with growing L. But for longer words, even if they are correct, there is a small chance to produce correct copies. Fig. 5 shows that the width of the final population increases by growing L. This certainly limits the size of the inheritable genetic information. It is only the "selection pressure", not inheritance, which prevents a complete randomization of the text. As Eigen has emphasized: for a fixed value of μ natural selection will optimalize the size of the genetic information. This phenomenon will be studied in the next Chapter.

Growth of the information content

Peter Schuster [5] presented a simple model to show how the mutation probability per character limits the size of the genetic information. Here he assumes for simplicity that only correct words multiply, any incorrect character is lethal. This assumption is useful to get a relation between μ and $L_{\max}$. But to understand the increase of the genetic information to its optimum length one has to relax this strict condition. That is what we do in our second computer model.

Let us consider a population of "words" with different lengths. (E.g. we shall put $L=2, 3, 4, 5$. Lone characters are not considered to be sensible words, but only building blocks available in unlimited quantity. The maximum value of L is limited by the long build-up time and by computer capacity. Let us see, if the evolution dynamics makes any further restriction.) For each length there is a "target word", the longer target words are obtained from the shorter ones by adding a new character to the end. (E.g. a series of target words may be AB, ABC, ABCD, ABCDE, or in a more English way: TO, TOO, TOOL, TOOLS. An arbitrary word made of L characters may contain R characters identical with those of the target word with the same length. (There are R identical characters on the correct place). The number of such words will be denoted by $y(L, R)$ in the population. Evidently $0 \leqq R \leqq L$, as before. The number $y(L, N)$ will change in time.

A longer chain of characters needs more time to be copied, so the time gap between two generations is supposed to be proportional to the number L of characters in the word. — A word carrying a more sophisticated information may have a longer life time. ("More enzymes offer a more elaborate defence system against wearing out.") In our second computer model the average life time is assumed to be

$$T(L, R) = 3^R \tau .$$

If one has a population of "correct" words of different lengths ($R = L$) and no error copies are allowed, the time dependence of $y(L, L)$ will be described by the differential equation

$$\dot{y}(L, L) = \left(\frac{1}{L} - \frac{1}{3^L \tau}\right) y(L, L) .$$

The optimum value of L, offering the fastest multiplication, can be obtained from the algebraic equation

$$\frac{\mathrm{d}}{\mathrm{d}L}\left(\frac{1}{L} - \frac{1}{3^L \tau}\right) = 0 ,$$

i.e. from

$$L^2 3^{-L} \lg 3 = \tau . \qquad (5)$$

So for any τ one can estimate the optimum length of words (see Fig. 6).

If the word contains also "wrong" characters, its multiplication rate is suppressed by a factor of 1/2 per wrong character. So an (L, R) word has the multiplication value

$$S(L, R) = \frac{1}{L} \frac{2^R}{\sum\limits_{R'=0}^{L} 2^{R'}} . \qquad (6)$$

These are the elements of the diagonal spreading matrix S.

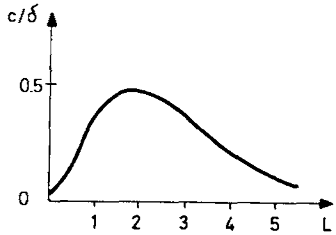

Fig. 6

Not all offspring are true copies. The mutation matrix (4) can be approximated for a small value of μ by the expression

$$M = 1 + \mu LP\,. \tag{7}$$

Now the population contains words of different lengths, so both **S** and **M** are composite matrices containing diagonal submatrices indicated by (6) and (7).

Let us allow that a random character may stick to the end of any word with a probability γ per unit time. (γ for "growth"). There is a 1/16 chance that the new added character is a "right" one (increasing L to $L+1$ and increasing R to $R+1$). There is a 15/16 chance that the new added character is a "wrong" one (increasing L to $L+1$ but leaving R unchanged). This effect can be described by a nondiagonal matrix G.

$$\mathbf{G} = \left[\begin{array}{ccccccccc}
-1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & \ldots \\
0 & -1 & 0 & 0 & 0 & 0 & 0 & 0 & \ldots \\
0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & \ldots \\
\frac{15}{16} & 0 & 0 & -1 & 0 & 0 & 0 & 0 & \ldots \\
\frac{1}{16} & \frac{15}{16} & 0 & 0 & -1 & 0 & 0 & 0 & \ldots \\
0 & \frac{1}{16} & \frac{15}{16} & 0 & 0 & -1 & 0 & 0 & \ldots \\
0 & 0 & \frac{1}{16} & 0 & 0 & 0 & -1 & \ldots & \\
0 & 0 & \frac{1}{16} & 0 & 0 & 0 & -1 & \ldots & \\
\ldots & \ldots & \ldots & \ldots & \ldots & \ldots & \ldots & &
\end{array}\right]. \tag{8}$$

It will be assumed in the new model that a word at the site of a wrong character is more vulnerable than elsewhere. There is a probability β for a break of the word at the wrong character. The wrong character can stick to the first or to the second fragment with equal probability. If both fragments consist of at least two characters, we have got two shorter words instead of the original long one. The $L-R-1$ other wrong characters of the original word distribute in average uniformly between the two fragments, so the first fragment (containing the first L_1 characters of the original word) will consist of

$$R_1 = L_1 - \frac{1}{2}\left[L_1 \frac{L-R-1}{L-1}\right] + \frac{1}{2}\left[(L_1-1)\frac{L-R-1}{L-1} + 1\right]$$

right characters. The right characters of the original word will become wrong certainly in the second fragment, but some of the wrong characters of the original word may turn out to be right, with a probability 1/16 per character. All these changes can be taken into account algebraically with the help of a nondiagonal matrix **B**.

The decay of the words can be described with a mean life time

$$T(L,R)=3^R\tau\,,$$

i.e. with the expression $\tau^{-1}\mathbf{D}$, where

$$\mathbf{D}=\begin{bmatrix} 1 & 0 & 0 & 0 & =0 & 0 & 0 & \dots \\ 0 & 3^{-1} & 0 & 0 & 0 & 0 & 0 & \dots \\ 0 & 0 & 3^{-2} & 0 & 0 & 0 & 0 & \dots \\ 0 & 0 & 0 & 1 & 0 & 0 & 0 & \dots \\ 0 & 0 & 0 & 0 & 3^{-1} & 0 & 0 & \dots \\ 0 & 0 & 0 & 0 & 0 & 3^{-2} & 0 & \dots \\ 0 & 0 & 0 & 0 & 0 & 0 & 3^{-3} & \dots \\ \dots & \dots & \dots & \dots & \dots & \dots & \dots & \dots \end{bmatrix}.$$

Finally, the differential equation describing the fate of the population reads

$$\mathbf{y}=[(1+\mu L\mathbf{P})\mathbf{S}+\gamma\mathbf{G}+\beta\mathbf{B}-\tau^{-1}\mathbf{D}]\mathbf{y}\,.$$

This takes into account several aspects influencing the information content of the words. To see a specific example, the numerical parameters were chosen in the following specific way:

$$\mu=0.10\,,\qquad \gamma=0.02\,,\qquad \beta=0.02\,,\qquad \tau=0.4\,.$$

The computation started with a single nonsense word of two characters ($L=2$, $R=0$, like FZ). The emergence of longer and sensible words is shown in Fig. 7. This Figure gives a motion picture, how the two character word improved (e.g. from FZ to TO), then one experiences a very slow transition to three characters and improving of the meaning of the three characters (e.g. making TOO). In the final steady population about 3/4 of the population is made of a sensible four-character word (e.g. TOOL), coexisting with a few sensible three-character expressions (like TOO) and with a few sensible five-character expressions (like TOOLS). Copies with one character error are also present. The evolution stops at this stage. There would be no advantage to try to build up lengthy words (like TOOLSMITH), because they need too much time for formation and too high a chance to pick up misprints.

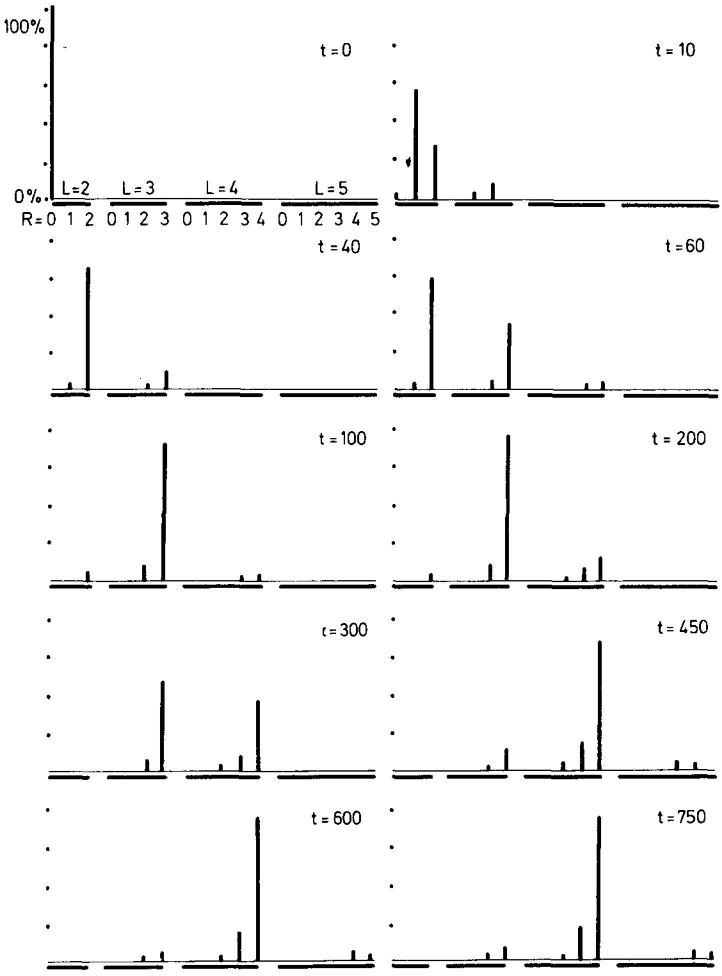

Fig. 7

Evidently the above set of characters has been used only as illustration. The computer program works with other parameter values as well, it can be used as a tool to model more complex situations concerning the evolution of genetic information.

The authors are very much indebted to Prof. Vilmos Csányi and to Prof. Peter Schuster for inspiring and critical discussions.

References

1. See e.g. K. G. Dengibh: An Inventive Universe. Hutchinson, London, 1975. Manfred Eigen-Ruthild Winkler: Das Spiel. Piper, München, 1975.
2. See e.g. Hubert P. Yockey, Journal of Theoretical Biology, **91,** 13., 1981. M. Coates and S. Stone, Journal of Molecular Evolution, **17,** 311, 1981. U. Niesert, D. Harnasch and C. Bresh, Journal of Molecular Evolution, **17,** 348, 1981.
3. Manfred Eigen, Berichte der Bunsen-Gesellschaft für physikalische Chemie 1059, 1976. Manfred Eigen, Peter Schuster, Die Naturwissenschaften, **64,** 541, 1977.
4. G. Marx, Biológia (Budapest) **28,** 177, 1980. (in Hungarian)
5. Peter Schuster: Prebiotic Evolution, in Biochemical Evolution, Cambridge University Press (1982), in print. See Figure 17.