10 Galaxies in the Cosmological Context

Gabriella De Lucia
INAF – Astronomical Observatory of Trieste, Trieste, Italy

T.D. Oswalt, W.C Keel (eds.), *Planets, Stars and Stellar Systems. Volume 6: Extragalactic Astronomy and Cosmology*, DOI 10.1007/978-94-007-5609-0_10, © Springer Science+Business Media Dordrecht 2013

Abstract: In the last decades, a number of observational experiments have converged to establish the cold dark matter model as the "de facto" standard model for structure formation. While the cosmological paradigm appears to be firmly established, a theory of galaxy formation remains elusive, and our understanding of the physical processes that determine the observed variety of galaxy properties and their evolution as a function of cosmic time and environment is far from complete. Although much progress has been made, both on the theoretical and observational side, understanding how galaxies form and evolve remains one of the most outstanding questions of modern astrophysics. This chapter provides an introduction to ideas and concepts that underpin modern models of galaxy formation and evolution, in the currently favoured cosmological context.

1 Introduction

It was not until the seventeenth century that Galileo discovered that the swathe of light visible on a dark night from horizon to horizon was not made up of some sort of "celestial fluid" but was instead composed of myriads of unresolved stars. More and more "patches of light" started to be observed – *nebulae* or *Island Universes*, using the definition given by Immanuel Kant. A comet hunter – Charles Messier – and a musician who become a skilled maker of the most powerful telescopes of his time – Wilhelm Herschel – independently produced the first catalogs of nebulae. The designations introduced by Messier, for basically all the nebulae that can be seen with small telescopes, are still in use today (e.g., the nearest spiral galaxy to the Milky Way – Andromeda – is also known as M31). Despite a *Great Debate*[1] held in 1920 to establish the nature of these objects, the controversy remained unresolved until 1925 when Edwin Hubble, using distances estimated from Cepheid variables in M31, provided the definitive demonstration of their extragalactic nature. Since then, astronomers have made huge progress in the observation of extragalactic systems and have collected a vast amount of detailed information, in different portions of the electromagnetic spectrum, for millions of galaxies. Despite almost one century having passed since the birth of extra-galactic astronomy, and despite much progress from both the observational and theoretical side having been made, many questions about the formation and the evolution of galaxies remain unanswered.

How do the nebulae form? And how do they evolve as a function of cosmic time and environment? The first detailed models for the formation of galaxies were proposed only about 40 years after the confirmation of their extragalactic nature. In their classical paper, Eggen et al. (1962) analyzed the properties and motion of 221 dwarf stars and showed that those of lower metallicity tended to move on more highly eccentric orbits. The observed trends were interpreted as a signature that the stars that are observed as a spheroidal halo in our galaxy formed during a rapid radial collapse that later continued to form the stellar disk. This scenario was later worked out in more detail in early numerical simulations carried out by Larson (1975, 1976). His work showed that, with appropriate choices of the parameters, these dissipative

[1]The National Academy of Sciences in Washington invited two astronomers, Harlow Shapley and Heber Curtis, to "debate" about the scale of the universe and the nature of the *nebulae*. The debate had no winner or looser. Although Curtis turned out to be correct as he believed that the nebulae were galaxies external to our own, Shapley was correct in arguing that our galaxy was larger than previously thought and for showing that our Sun was not at the center of its galaxy.

collapse models can reproduce the observed basic properties of both elliptical and spiral galaxies, provided that the star formation is much slower in proto-spirals than in proto-ellipticals. The numerical work by Larson, however, also pointed out that if some means of redistributing angular momentum is not included (e.g., viscosity), these models are unable to obtain the high-surface brightnesses that are observed in real galaxies. We now know that one of the main problems with these early studies was that they neglected the presence of dark matter.

The first observational evidence of a *missing mass* problem dates back to the 1930s, when Zwicky (1937) estimated that the speeds of galaxies in the Coma cluster are too large to keep the system gravitationally bound, unless the dynamical mass is at least 100 times larger than the mass contained in galaxies. The reality of the problem, however, gained a hold upon the astronomical community only in the mid-1970s, when different studies showed that the rotation curves of spiral galaxies are either flat or rising at the optical edge of the galaxies, contrary to the Keplerian fall off that is expected if the visible stars and gas were the only mass in the system (Rubin and Ford 1970; Einasto et al. 1974; Ostriker et al. 1974). These observations led to the conclusion that dark matter must play an important role in galaxy formation, and motivated the two-stage theory proposed by White and Rees (1978). In this scenario, dark matter haloes form first, and the physical properties of galaxies are then determined by cooling and condensation of gas within the potential well of the haloes. This model contains many of the ideas that are at the basis of the tools that are nowadays used to study the formation and evolution of galaxies, and that will be discussed in more detail in this chapter.

In the 1980s, much work focused on the nature of the unseen dark matter component. Initially, many studies focused on neutrinos as the most likely candidates for the dark matter. It was soon realized, however, that in a neutrino-dominated universe, structure would form by fragmentation (top-down), with the largest superclusters forming first in a sort of flat "pancake"-like sheets (Zeldovich et al. 1982). These must then fragment to form smaller structures like galaxy groups and galaxies – a picture that conflicts with observation, as shown by detailed simulations of structure formation (White et al. 1983). During the same years, a number of different dark matter candidates were provided by particle physics models based on supersymmetry. These weakly interacting massive particles (WIMPS) are today considered the most likely candidates for dark matter. Because their masses are much larger (and therefore their velocities[2] are much smaller) than those of neutrinos, these particles are said to be "cold." Cold dark matter (CDM) decouples from the radiation field long before recombination so that its density fluctuations can grow significantly before the baryons decouple from the radiation. When this happens, baryons are free to fall in the dark matter potential wells that have formed and that allow structure formation to occur at a rate sufficient to be consistent with the large-scale structure observed at present (Davis et al. 1985). The CDM theory has now become the preferred scenario for galaxy formation and is the framework that will be adopted in this chapter. In a CDM universe, structure grows hierarchically (bottom-up), with small objects collapsing first and later merging in a continuous hierarchy to form more and more massive systems.

The aim of this chapter is to provide an introduction to the ideas and concepts that underpin modern models of galaxy formation and evolution, in a universe in which cosmic structures originate from small initial perturbations and build up hierarchically through gravitational instabilities. The layout of this chapter is as follows: ❯ Section 2 provides a brief description of the cosmological model that is currently accepted as the standard model for structure formation, while ❯ Sect. 3 deals with the physical processes that govern the formation and the

[2]Their velocities are nonrelativistic at the epoch of radiation-matter equality.

evolution of galaxies. Section 4 provides a brief review of the numerical techniques that are currently used to study galaxy formation in a cosmological context and highlights their most recent successes and open problems. Finally, Sect. 5 gives some concluding remarks. Due to space limits, this chapter does not contain a detailed overview of the observational properties of local and/or distant galaxies. The interested reader is referred to other chapters of this volume, as well as to the textbooks by, e.g., Binney and Merrifield (1998) and Mo et al. (2010), where also a more detailed exposition of some of the topics discussed in the following can be found.

2 The Framework: The Dissipationless Universe

This section provides a brief review of the cosmological framework in which galaxy formation and evolution take place, focusing on those ingredients that can be considered as the initial and boundary conditions for any galaxy formation model. For a more rigorous and detailed exposition of the subject, the reader is referred to classical textbooks by, e.g., Padmanabhan (1993) and Peacock (1999).

2.1 The Cosmological Model

During the last decade, a variety of observational tests have ushered in a new era of "precision cosmology" and have converged to establish the CDM model (Peebles 1982; Blumenthal et al. 1984) as the *de facto* standard cosmological model for structure formation. In the currently favored cosmogony (the ΛCDM universe), about 75% of the energy density is due to a yet unknown form of *dark energy* that tends to increase the rate of expansion of the universe, about 21% to a nonbaryonic *cold dark matter* that has yet to be detected in the laboratory, and only about 4% is made of baryonic matter out of which stars and galaxies are made. In the past years, it has been shown that this cosmological model is able to match simultaneously a variety of observational measurements, among which are the power spectrum of low-redshift galaxies, the structure that is seen in the Lyman α forest at $z \sim 3$, the present acceleration of the cosmic expansion as inferred from supernovae observations, and the temperature fluctuations in the cosmic microwave background. By combining these experiments, the parameters of this cosmological model are currently known with uncertainties of only a few percent (e.g., Komatsu et al. 2011), thus effectively removing a large part of the parameter space in galaxy formation studies.

The initial fluctuations are assumed to follow a Gaussian random distribution and to have expanded to cosmological scales by inflation[3] – a brief period of time during which the scale factor of the early universe increased exponentially. The dark matter component that has no pressure undergoes gravitational collapse, which makes the perturbations grow. The early evolution of these perturbations can be accurately described using the linear approximation which breaks down, however, when the density contrast becomes nearly unity. In the nonlinear regime, the evolution can be studied analytically if some simplifying assumptions are made (e.g., the

[3]The inflationary hypothesis was introduced by Guth (1981). While inflation is understood principally by its detailed predictions of the initial conditions for the hot early universe, the detailed particle physics mechanism responsible for it is not known.

spherical top-hat model, see, e.g., Chap. 8 of Padmanabhan) or, more directly and accounting for the full geometrical complexity of the problem, using cosmological N-body simulations.

For the purposes of modeling galaxy formation, the following information should be available: (i) the distribution of the dark matter halo masses at any given redshift, (ii) the structural properties of the dark matter haloes, and (iii) a statistical representation of their assembly history (that is what in the jargon is called a "merger tree").

2.2 The Halo and Subhalo Mass Functions

The first calculation of the abundance of gravitationally bound structures was carried out by Press and Schechter (1974), long before the CDM model was introduced. By assuming a Gaussian density field smoothed using a spherical top-hat window, and by varying the radius of the smoothing window, one can consider structures of different mass $M = 4/3\pi\rho R^3$. The abundance of haloes above a given mass depends on the fraction of spheres for which the density contrast (this is usually expressed as $\delta = \rho(x)/\bar{\rho} - 1$) exceeds some critical value δ_c. A natural choice for the critical value of the density contrast is provided by the spherical top-hat model and corresponds to the linearly extrapolated density contrast at which haloes are expected to virialize ($\delta_c \sim 1.69$).

Assuming that the probability that $\delta > \delta_c$ is the same as the fraction of mass elements that are contained in haloes with mass larger than M, one obtains

$$\frac{dn}{dM}(M,t) = \left(\frac{2}{\pi}\right)^{1/2} \frac{\rho_0}{M^2} \frac{\delta_c}{\sigma(M)} \left|\frac{d\ln\sigma}{d\ln M}\right| \exp\left[-\frac{\delta_c^2}{2\sigma^2(M)}\right] \tag{10.1}$$

where ρ_0 is the mean density of the universe, $\sigma(M)$ is the fractional root variance in the density field smoothed using a top-hat filter that contains, on average, a mass M, and $\delta_c(t)$ is the critical overdensity for spherical top-hat collapse at time t. The Press and Schechter derivation neglects underdense regions that can be enclosed within larger overdense regions and that would have a finite probability of being included in a larger collapsed object. To correct this, Press and Schechter introduced a "fudge factor" equal to 2 in front of the derived expression (this is included in the equation above) but did not give a proper demonstration of the correction adopted. An alternative derivation of the halo mass function was given by Bond et al. (1991), using what is usually referred to as the "excursion set formalism." A detailed expositionl of this formalism can be found in White (1994, see also Sect. 7.2 of Mo et al. 2010).

The halo mass function predicted by this simple calculation agrees surprisingly well with the results obtained from N-body simulations. This is shown in the left panel of *Fig. 10-1*. The colored symbols are results from the Millennium Simulation, which follows the evolution of $N = 2,160^3$ particles of mass $8.6 \times 10^8\ h^{-1}\mathrm{M}_\odot$ within a comoving box of size $500\ h^{-1}$Mpc on a side. Dashed lines are the Press-Schechter predictions at $z = 0$ and $z = 10$ and show that this formula underpredicts the high-mass end of the mass function by up to an order of magnitude, with the disagreement becoming worse at earlier cosmic epochs. Solid lines are predictions from the fitting formula proposed by Jenkins et al. (2001) that appears to describe results from the N-body simulation remarkably well, over the redshift and mass range well sampled.

Until the late 1990s, dissipationless simulations suffered from the so-called "overmerging" problem, i.e., substructures disrupted very quickly within dense environments so that haloes were smooth and featureless. The problem was initially explained by the lack of dissipation in N-body simulations (Katz et al. 1992; Summers et al. 1995): it was thought that baryons would

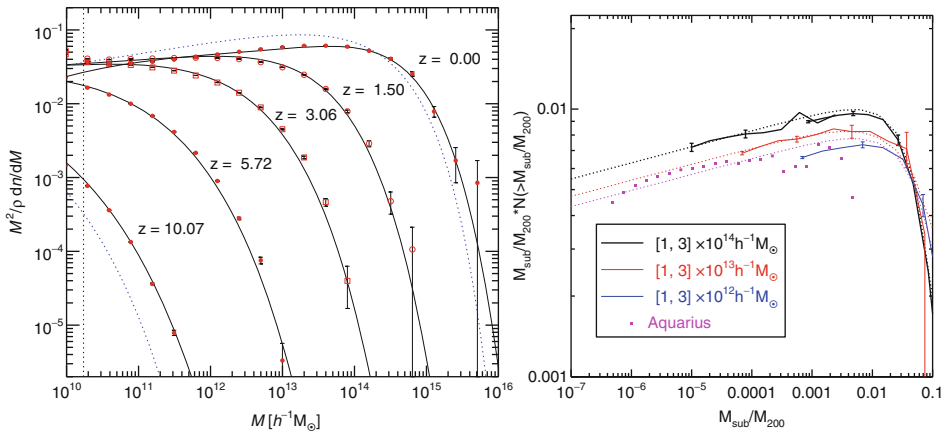

Fig. 10-1

***Left panel*: from Springel et al. (2005b). The differential halo mass function at different epochs. The mass function has been multiplied by M^2 to take out the dominant mass dependence. *Solid lines* are predictions from an analytic fitting function proposed in Jenkins et al. (2001) while *dashed lines* show the Press-Schechter mass function at $z = 0$ and $z \sim 10$. *Colored* symbols are obtained from the Millennium Simulation (Springel et al. 2005b). *Right panel*: from Gao et al. (2011). *Solid lines* show the averaged cumulative subhalo mass functions for three intervals of host halo mass as indicated in the legend. The error bars on selected points show the error on the mean for the three mass ranges indicated. The *filled squares* show the mean of the cumulative subhalo mass functions of the six Aquarius haloes with typical mass of $\sim 10^{12}\ h^{-1}\mathrm{M}_\odot$**

sink into the center of dark matter haloes making them more resilient to disruption by the tidal field of the parent halo. Both analytic work and high-resolution simulations, however, demonstrated later that the cores of dark matter haloes that fall into a larger system can actually survive as self-gravitating objects orbiting in the smooth dark matter background of the halo, provided high enough force and mass resolution are used. A wealth of dark matter substructures are now routinely identified using different techniques (see below). If any, we are now facing the opposite problem of having "too much" substructure, at least on the galactic scale, where simulations predict more substructures than visible galaxies by almost two orders of magnitude (see Sect. 4 in Tasitsiomi 2003). As an example of the performance achieved by numerical N-body simulationsh in the last years, *Fig. 10-2* reproduces the density map of a 10-year-old high-resolution simulation of a galaxy cluster in the right panel, and what could be considered the state-of-the-art numerical simulation on the same scale only about three decades ago in the left panel.

The identification of substructures in dark matter haloes is a difficult technical problem, and many different algorithms have been developed to accomplish this task in the last years. Each of these has its own advantages and weaknesses. For example, in the hierarchical friends-of-friends algorithm (HFOF, Klypin et al. 1999), the linking length[4] is reduced in discrete

[4]The friend-of-friend (FOF, Davis et al. 1985) algorithm is a percolation algorithm that links together all the particles with a separation less than b times the mean interparticle separation. It has been shown that, with an appropriate choice of the linking length, it is possible to select groups close to the virial overdensity predicted by the spherical collapse model.

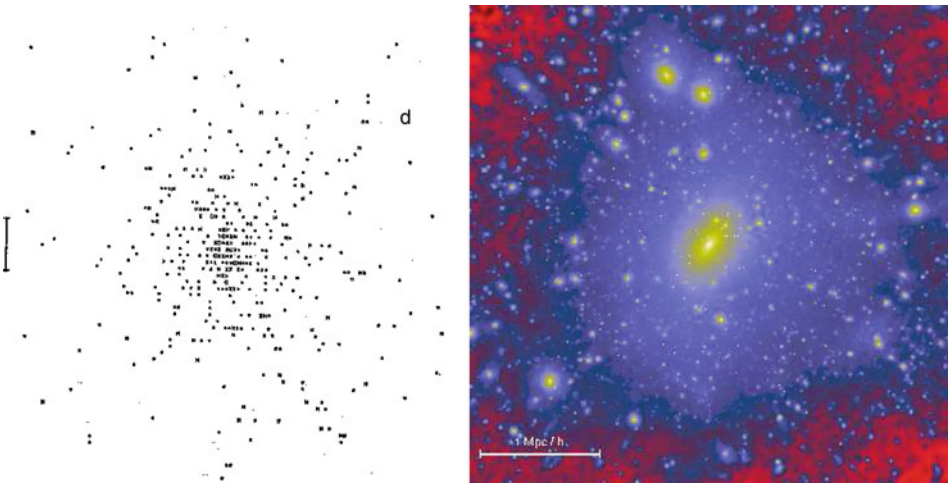

Fig. 10-2
***Left panel*: from White (1976). Projected distribution of a 700-body system with mass comparable to the virial mass of the Coma cluster. *Right panel*: from Springel et al. (2001). Density map of a high-resolution cluster resimulation. The cluster has a virial mass of 8.4×10^{14} $M_\odot$ and the high-resolution region of the simulation contains about 66 million particles**

steps, thus selecting groups of higher and higher overdensity. The choice of the levels of linking lengths is somewhat arbitrary, and the algorithm requires an iterative procedure. The bound density maximum algorithm (BDM, Klypin et al. 1999) iteratively determines a bound subset of particles in a sphere around a local density maximum. Since this method separates background particles from particles that are bound to the halo, the BDM algorithm estimates the physical properties of substructures more accurately. It implicitly assumes, however, that the halo is spherically symmetric, while the HFOF algorithm can deal with haloes of arbitrary shapes. Another approach is given by the SKID algorithm (see http ref: http://www-hpcc.astro.washington.edu/tools) in which the density around each particle is evaluated using a smoothing kernel. The particles are then moved along the density gradients toward a local density maximum. Particles that end up in the same local maximum are linked together using an FOF algorithm and then checked for self-boundness. Only self-bound groups with more than a user-specified minimum number of particles are kept as genuine substructures. An algorithm that has been frequently used in recent years is, finally, SUBFIND (Springel et al. 2001) which combines ideas used in other group-finding techniques with a topological approach for finding substructure candidates.

Typically, only about 10% of the total mass of a dark matter halo is found in substructures. The abundance of relatively massive substructures increases systematically (albeit weakly) with host halo mass, as shown in the right panel of *Fig. 10-1*. This trend reflects the fact that more massive haloes are both less centrally concentrated and younger (i.e., they assembled later) than their less massive counterparts. Therefore, they exert weaker tidal forces and have had less time to disrupt their substructures. As discussed above, different algorithms can be used to identify dark matter substructures, and different criteria for defining the boundaries and membership of these substructures are bound to lead to systematic differences. Several recent studies, however,

find very similar slopes for the subhalo mass function, which suggests that the expected differences can be probably corrected by simple scale factors (a recent detailed comparison between different algorithms is given in Knebe et al. 2011).

2.3 Halo Structure

The internal structure of dark matter haloes has been studied extensively using N-body simulations. These show that the density profiles of dark matter haloes are shallower than r^{-2} at small radii and steeper at large radii. The density profile extracted from N-body simulations is well described by the following equation:

$$\rho(r) = \rho_{\rm crit} \frac{\delta_{\rm char}}{(r/r_s)(1 + r/r_s)^2} \tag{10.2}$$

where r_s is a scale radius and $\delta_{\rm char}$ is a characteristic overdensity. The above profile has been shown to provide a good representation of the equilibrium density profiles of dark matter haloes of all masses in all CDM-like cosmogonies (Navarro et al. 1997, NFW). In (◉ 10.2), the local logarithmic slope gradually changes from a value of −3 in the outer parts to an asymptotic slope of −1 in the inner parts. The spatial scale r_s of this transition is treated as a fitting parameter and is often parameterized in terms of the concentration $c = r_h/r_s$ of the halo, which in fact is a reparametrization of $\delta_{\rm char}$ relative to the critical density:

$$\delta_{\rm char} = \frac{\Delta_h}{3} \frac{c^3}{\ln(1 + c) - c/(1 + c)}$$

where the limiting radius of a dark matter halo (r_h) is defined as the radius within which the mean matter density is

$$\rho_h = \Delta_h \bar{\rho} = \Delta_h \rho_{\rm crit} \Omega_m$$

and $\bar{\rho}$ is the mean matter density of the universe at the time considered and $\rho_{\rm crit}$ is the corresponding critical density for closure. Different definitions of the radius of a halo can be found in the literature. The most commonly adopted definition corresponds to R_{200}, that is, the radius that contains a mean overdensity equal to 200 times the critical density at the redshift considered. The corresponding enclosed mass is usually referred to as M_{200}, and in this case, $\Delta_h = 200/\Omega_m$.

For a given cosmology, the NFW profile is then completely characterized by the halo mass and by the concentration parameter $c = r_h/r_s$. At any given epoch, less massive haloes are more concentrated than their more massive counterparts (Neto et al. 2007 and references therein), a finding that can be interpreted as reflecting the density of the universe at the time of halo formation. More recent N-body studies (e.g., Navarro et al. 2004) show that the density profiles of highly resolved simulated haloes deviate from the NFW profile, particularly in the inner regions, and demonstrate that they are better described by an Einasto (1965) profile:

$$\rho(r) = \rho_{-2} \exp\left[\frac{-2}{\alpha}\left[\left(\frac{r}{r_{-2}}\right)^{\alpha} - 1\right]\right]$$

with r_{-2} equal to the radius at which the logarithmic slope of the density distribution is equal to −2 and $\rho_{-2} = \rho(r_{-2})$. The shape parameter of the Einasto profile (α) appears to vary systematically with halo mass (e.g., Hayashi and White 2008), a result that indicates a (small) deviation of the mean density profiles from a "universal" shape.

N-body simulations also show that dark matter haloes have strongly triaxial shapes, with a slight preference for nearly prolate systems (Jing and Suto 2002; Hayashi et al. 2007), and that they are supported by nearly isotropic velocity dispersions (Wojtak et al. 2005). Another important property of a dark matter halo is its angular momentum, traditionally parameterized as

$$\lambda = \frac{J\,E^{1/2}}{G\,M^{5/2}}$$

where J, E, and M are the total angular momentum, energy, and mass of the halo, respectively. Numerical simulations have shown that the distribution of spin parameters for dark matter haloes is well fit by a log-normal distribution:

$$p(\lambda)\lambda = \frac{1}{\sqrt{2\pi}\sigma_{\ln\lambda}} \exp\left[-\frac{\ln^2(\lambda/\bar{\lambda})}{2\sigma^2_{\ln\lambda}}\right]\frac{d\lambda}{\lambda}$$

with $\bar{\lambda} \sim 0.035$ and $\sigma_{\ln\lambda} \sim 0.5$. The median and width of this distribution appear to depend weakly on halo mass, redshift, and cosmology (Bett et al. 2007; Macciò et al. 2007).

2.4 Halo Merger Trees

A statistical description of the assembly history of haloes, i.e., a description of the merging events and of the masses of the haloes involved, can be obtained using a Monte Carlo approach by sampling the distribution of progenitor masses predicted from the extended Press-Schechter theory (Lacey and Cole 1993) or by using outputs from N-body simulations. In the jargon, this is called "merger tree." Its first schematic representation was presented by Lacey and Cole and is reproduced in ❯ *Fig. 10-3*. In the figure, cosmic time increases from top to bottom, and the widths of the branches reflect the masses of the individual merging haloes.

The excursion set approach of the extended Press-Schechter formalism provides a neat way to calculate the distribution of halo progenitor masses M_1 at redshift z_1, for a halo of mass M_2 at later redshift z_2. This can be written as

$$\frac{dN}{dM_1} = \left(\frac{2}{\pi}\right)^{1/2} \frac{d\ln\sigma}{d\ln M_1} M_2 \frac{\sigma_1^2}{M_1^2} \frac{\delta_{c1}-\delta_{c2}}{(\sigma_1^2-\sigma_2^2)^{3/2}} \exp\left[-\frac{(\delta_{c1}-\delta_{c2})^2}{(\sigma_1^2-\sigma_2^2)}\right] \tag{10.3}$$

where $\sigma_1 = \sigma(M_1)$, $\sigma_2 = \sigma(M_2)$, $\delta_{c1} = \delta_c(z_1)$, $\delta_{c2} = \delta_c(z_2)$. Repeating the procedure at different redshifts, and imposing that the mass is conserved so that in each individual realization, the sum of the progenitor masses is equal to the mass of the parent halo, one can construct merger trees of haloes of different mass, with arbitrary high resolution. In practice, finding a suitable algorithm is not trivial, and different methods have been proposed (see Sect. 7.3 of Mo et al. 2010). In general, the Press-Schechter formalism and its Monte Carlo extension capture the qualitative behavior of all statistics that can be extracted from N-body simulations. However, recent studies have shown that some discrepancies are found between analytic merger trees and the corresponding statistics extracted from N-body simulations. The level of this disagreement, which becomes more important with increasing redshift, can be reduced by empirically tuning the progenitor distributions, but no theoretical justification exists for the form of the proposed corrections (Parkinson et al. 2008).

A fundamental assumption that underlies the Monte Carlo approach is that the formation history of a halo of a given mass does not depend on the "environment." This assumption

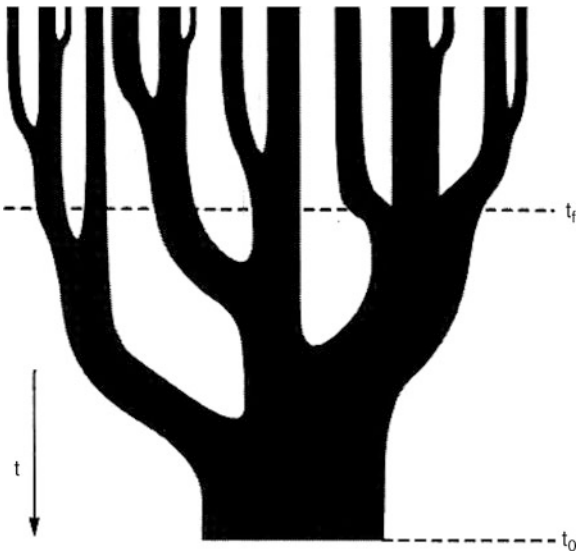

Fig. 10-3
From Lacey and Cole (1993). Illustration of a merger tree. Time increases from *top* to *bottom*, and the widths of the tree branches encode the masses of the merging haloes

was supported by early numerical work who found no dependence of halo clustering on concentration or formation time[5] (see, e.g., Lemson and Kauffmann 1999; Percival et al. 2003). A reanalysis of the same data, however, showed that close pairs of haloes form at slightly higher redshifts than more widely separated halo pairs, suggesting that haloes in dense regions form at slightly earlier times than haloes of the same mass in less dense regions (Sheth and Tormen 2004). These results were later confirmed by more recent numerical work that analyzed the properties of dark matter haloes in large volumes with high resolution, and found a clear dependency of the clustering amplitude on the halo formation time (Gao et al. 2005). This is illustrated in *Fig. 10-4* which shows the two-point correlation function[6] for haloes in four different mass ranges (each panel corresponds to a different mass bin, as indicated in the legend) and for the 20% oldest (red lines) and youngest (blue lines) haloes in each mass range. The figure shows that older haloes are more clustered than their younger counterparts with similar mass and that the dependence on the formation time is strongest for galactic mass haloes. It should be noted that these haloes were not well resolved by earlier numerical work that addressed the same issue.

Strictly speaking, this result invalidates the Monte Carlo approach in terms of using a one-parameter model (i.e., the halo mass) to construct the merger tree. In addition, as discussed above, this effect is strongest for haloes similar in mass to that of our Milky Way, which represent a large fraction of the galaxies in typical observational surveys. Since it is plausible that galaxy

[5]The formation time of a halo is typically defined as the time when the most massive progenitor of the halo first contains half the final mass.
[6]The two-point correlation function describes the probability, in excess of Poisson probability, to find two galaxies at a given relative distance.

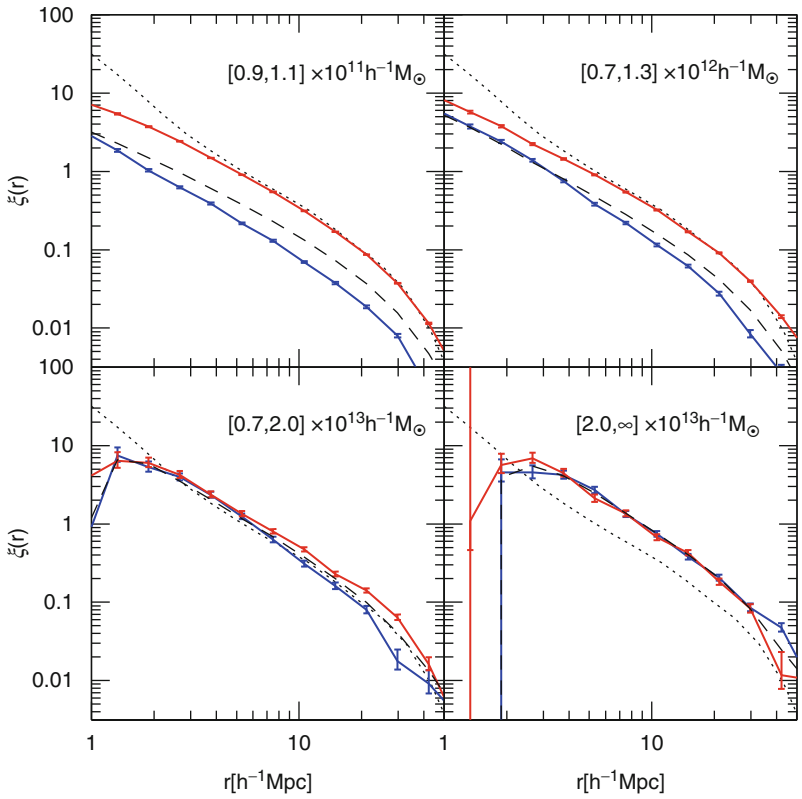

Fig. 10-4

From Gao et al. (2005). Two-point correlation functions for haloes in four mass ranges. Each panel gives results for haloes in the mass range indicated in the label. The *dotted black line*, repeated in all panels, is the correlation function of the underlying mass distribution. *Dashed black lines* give the correlation functions for the full sample of haloes in each mass range. The *red* and *blue* curves give correlation functions for the 20% oldest and 20% youngest of these haloes, respectively. Error bars are based on Poisson uncertainties in the pair counts

properties depend on the assembly history of their haloes, these results suggest that models that ignore the dependence on the large-scale structure will be in error, although the extent of the problem does likely depend on the specific galaxy formation model considered. Recent tests suggest that the effect discussed above influences the galaxy-galaxy and galaxy-mass correlation function by 5–10%, which is within their current statistical uncertainty (Croton et al. 2007). The trends discussed, however, are likely to play a more important role for studies of extreme objects that may be thought, for example, to form particularly early or late.

Alternatively, merger trees can be constructed using outputs from N-body simulations. This is not a trivial task: a discrete number of simulation outputs is available; one may want to include substructures which complicate significantly the merger tree structure; the mass of a halo can decrease with time; haloes may spatially overlap at a given time output and therefore be blended together by the specific group-finding algorithm employed, then separate at

the next time output, and eventually come back together again later on; etc. (for a discussion of problems commonly encountered when building merger trees from N-body simulations, see Tweed et al. 2009). The main advantage of using merger trees extracted from simulations to graft on galaxy formation models is that they can give predictions for the positions of galaxies within haloes. This allows realistic "mock catalogues" to be constructed which contain not only the physical properties of all model galaxies (e.g., luminosities, masses, star formation rates) but also dynamically consistent redshift and spatial information, like in real galaxy redshift surveys. In addition, numerical merger trees are immune to the problem discussed above because they automatically take into account the dependence of halo clustering on age. On the other hand, N-body merger trees suffer of a finite mass resolution and of the "technical" problems mentioned above. Both approaches, extracting the trees directly from an N-body simulation and growing Monte Carlo trees, have therefore their advantages and weaknesses, and both are still widely used as input for some classes of galaxy formation models that will be discussed in more detail in Sect. 4.

Figure 10-5 shows the merger tree of a cluster-size halo, extracted from an N-body simulation. The branch highlighted in green is obtained by connecting the halo at each time step to the progenitor with the largest mass (the "main" progenitor). The rightmost branches are merger trees of secondary substructures (only those with more than 500 particles are shown) present in the FOF group at $z = 0$. Circles mark objects that belong to the same FOF group as the main progenitor, while triangles mark objects that have not yet joined the FOF group. Typically, when a halo is accreted onto a bigger system (i.e., becomes a "subhalo"), it loses mass efficiently due to tidal stripping (De Lucia et al. 2004a; Gao et al. 2004). A nice example of this process is

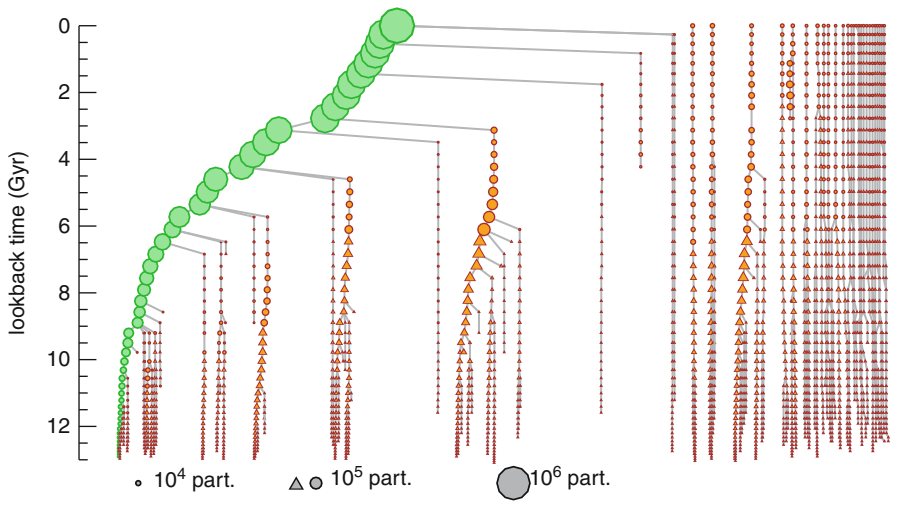

Fig. 10-5
From De Lucia and Blaizot (2007). Merger tree of a FOF group. Only the trees of subhalos with more than 500 particles at $z = 0$ are shown. Their progenitors are shown down to a 100-particle limit. The leftmost tree is that of the main subhalo of the FOF, while the trees on the right correspond to other substructures identified in the FOF group at $z = 0$

shown by the halo branch located roughly at the center of *Fig. 10-5*. The simulation work mentioned above also shows that a significant fraction of the substructures residing in cluster-size haloes at the present day were accreted at redshifts $z \lesssim 1$ and that the infall time and the retained mass of a subhalo are both strongly increasing functions of clustercentric radius. This implies that subhaloes in the inner regions of cluster haloes today were generally more massive in the past than similar mass but more recently accreted subhaloes in the outer regions. This is an important result to consider when linking the properties of luminous galaxies to those of dark matter (sub)haloes.

3 The Physics of Galaxy Formation

So far, this chapter has focused on the formation and evolution of structure under the influence of gravity alone. In order to make a close link between theoretical models of structure formation and observational data, it is necessary to consider the gas-dynamical and radiative processes that drive the evolution of the baryonic component of dark matter haloes. These processes are far more difficult to deal with than gravitational instability, as they cover several orders of magnitude in physical size and timescales, and are intertwined in an entangled network of actions, back reactions, and self-regulations. This section provides an overview of the main physical processes and ingredients that have to be considered when modeling the formation and evolution of galaxies in the cosmological set discussed in the previous section, highlighting the current status of observational and numerical studies.

3.1 Gas Accretion

During the linear regime, the density perturbation fields of the baryons and dark matter are expected to be equal on scales above the Jeans length. After halo formation, hydrodynamical forces come into play, and further collapse of the gaseous component associated with dark matter haloes is regulated by a combination of gravity, cooling, and hydrodynamical processes.

If the halo virial temperature exceeds the temperature of the accreting gas, then the gas will accrete supersonically, which will give rise to an accretion shock. Both analytic work and numerical simulations have early shown that when the cooling times are longer than the dynamical times, the shock occurs at a radius that is comparable (or slightly larger than) the virial radius (Bertschinger 1985; Evrard 1990). In reality, the accreting gas is not smooth but lumpy so that there is no well-defined accretion shock but rather a complex network of shocks. These heat the gas by thermalizing its kinetic energy up to the virial temperature of the halo. For an isothermal sphere, this can be written as

$$T_{\rm vir} = \frac{\mu m_{\rm p}}{2k_{\rm B}} V_{\rm c}^2 \simeq 3.6 \times 10^5 {\rm K} \left(\frac{V_{\rm c}}{100\,{\rm km\,s^{-1}}} \right)^2$$

where $m_{\rm p}$ is the proton mass and μ is the mean molecular weight of the gas. This gas will form a hydrostatically supported atmosphere which will obey the hydrostatic equilibrium equation:

$$\frac{{\rm d}P}{{\rm d}r} = \frac{{\rm d}(k_{\rm B}T\rho/\mu m_{\rm p})}{{\rm d}r} = -\rho(r)\frac{{\rm d}\Phi}{{\rm d}r} = -\rho(r)\frac{GM(r)}{r^2}$$

where P is the gas pressure, $\rho(r)$ the gas density, and $M(r)$ the total (i.e., dark matter plus baryonic) mass within the radius r. This gas will then cool radiatively and eventually lose energy and, consequently, pressure support. At this point, the gas will fall toward the center of the gravitational potential provided by the dark matter halo, conserving its angular momentum and settling in a denser gas disk.

In the regime where the cooling times are much shorter than the dynamical times, the shock forms at much smaller radii, close to the forming galaxy. The gas is still heated to very high temperatures (actually larger than in the slow cooling regime because the preshock velocity of the infalling gas will be larger than in this case of a virial shock) but will cool so rapidly that it cannot maintain the pressure needed to support a quasistatic hot atmosphere. The distinction between these "rapid" and "slow" cooling regimes was clearly understood when the first hierarchical galaxy formation models were presented (Rees and Ostriker 1977; Binney 1977; White and Frenk 1991). This picture has been validated by 1D hydrodynamical simulations (Birnboim and Dekel 2003, see also unpublished work by Forcada-Miro and White 1997) and by more recent 3D hydrodynamical simulations (Kereš et al. 2005; Ocvirk et al. 2008) that show that most of the accretion on haloes with mass $\lesssim 10^{12} M_\odot$ tend to be directed along filaments, and is often referred to as "cold accretion." As Croton et al. (2006) have stressed and as noted above, the term cold accretion is a misnomer. In fact, what differentiates mainly the two modes of accretion is not the temperature to which infalling gas is shocked but rather the time spent by the gas at the postshock temperature before its energy is radiated away. It is worth noting that the transition mass between the rapid and slow cooling regimes found in the most recent simulations is very close to that identified in early analytical work (see discussion in Benson and Bower 2011). Finally, it should be noted that the rates computed in simulations often correspond to accretion rate onto the haloes and that these are different from the accretion rates onto the galaxies. The latter can be strongly affected by metal line cooling and by feedback from supernovae and/or active galactic nuclei (Benson and Bower 2011; van de Voort et al. 2011).

3.2 Gas Cooling

The primary cooling processes relevant for structure formation are two-body radiative processes. A gas with primordial composition (only hydrogen and helium) is almost entirely ionized at temperatures above 10^6 K, while a gas of nonzero metallicity is fully ionized at temperatures above a few 10^7 K. At these high temperatures, the cooling is dominated by the bremsstrahlung continuum due to the deceleration of electrons as they encounter atomic nuclei. At lower temperatures (i.e., $10^4 \mathrm{K} < T < 10^6 \mathrm{K}$), collisional ionization, recombination, and collisional excitation become important. At even lower temperatures ($T < 10^4$ K), most of the electrons have recombined so that atomic cooling is very inefficient. Cooling can still take place (albeit at very low rates) if the gas is enriched, but the dominant cooling in this regime is given by the excitation (through collisions) of rotational or vibrational energy levels of molecular hydrogen (or of other molecules if present) and subsequent decay. Since the dominant cooling processes are two-body processes, one can write the cooling rate per unit volume as

$$\mathcal{L} = n_{\mathrm{H}}^2 \Lambda(T, Z)$$

where n_{H} is the number density of hydrogen (both neutral and ionized) and $\Lambda(T, Z)$ is the cooling function that, as explained above, will depend (strongly) both on the temperature and on the chemical composition of the gas.

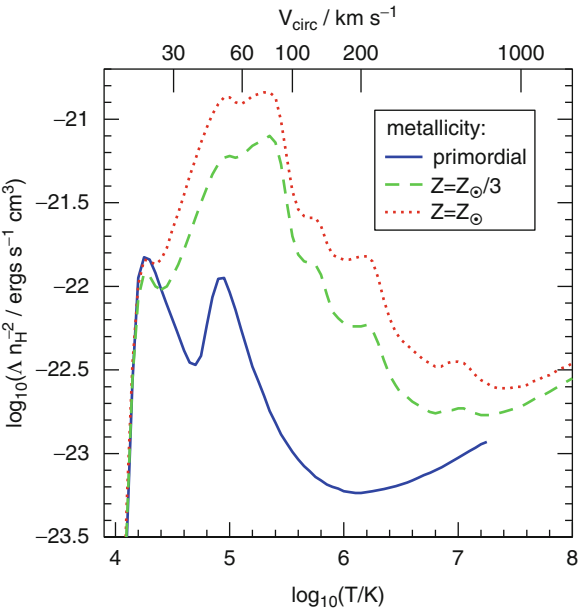

Fig. 10-6

From Baugh (2006), based on model results from Sutherland and Dopita (1993). The cooling rate is plotted as a function of the virial temperature of the hot halo gas. The equivalent circular velocity of the halo is indicated on the *top* axis. The different *curves* show how the cooling rate depends upon the metallicity of the gas, as indicated by the legend

Figure 10-6 shows how the cooling rate varies as a function of the temperature of the hot halo gas and how it depends upon the chemical composition of the gas. Cooling is dominated by bremsstrahlung at the high temperature end, where $\Lambda \propto T^{1/2}$. The peaks in the primordial cooling function at ~15,000 K and ~10^5 K are due to the collisionally excited electronic levels of hydrogen and singly ionized helium, respectively. For an enriched gas, cooling is significantly enhanced at temperatures $\gtrsim 10^5$ K due to the collisionally excited levels of ions of oxygen, carbon, nitrogen, etc. Above ~10^6 K, other metal lines contribute significantly, in particular neon, iron, and silicon. The cooling functions shown in *Fig. 10-6* are based on model results from Sutherland and Dopita (1993) and assume ionization equilibrium, i.e., that the densities of all ions are equal to their equilibrium values. This approximation is correct if the time scales of the radiative processes are much shorter than the hydrodynamical time scales of the gas, which might not be the case in shocks or in the case of a very dilute gas (where reaction rates are low). In these cases, a more appropriate treatment requires cooling rates to be recomputed using nonequilibrium densities.

At high redshifts, an additional cooling channel has to be taken into account: inverse Compton scattering of cosmic microwave background photons by electrons in the ionized plasma inside dark matter halos. This channel is effective if the temperature of the plasma exceeds that of the microwave background $T_\gamma \approx 2.73(1+z)$ K. It can be shown that

$$\frac{t_{\rm Compton}}{t_{\rm age}} \approx 350\Omega_m^{1/2} h (1+z)^{-5/2}$$

where $t_{\rm age}$ is the age of the universe at redshift z. For $\Omega_m = 0.3$ and $h = 0.7$, one obtains $t_{\rm Compton}/t_{\rm age} = 1$ at $z \sim 6$. So Compton cooling against the cosmic microwave background becomes important only at $z \gtrsim 6$. The cooling rate per unit mass associated with Compton cooling is proportional to the electron temperature and independent on the gas density (see Sect. 8.1.2 of Mo et al. 2010). So, assuming an isothermal distribution and a constant electron fraction, gas that is able to cool via this process will do so at all radii.

In a spherically symmetric gaseous system, a local cooling time can be defined dividing the thermal energy density of the gas by the cooling rate per unit volume:

$$t_{\rm cool}(r) = \frac{3}{2}\frac{kT\rho_{\rm g}(r)}{\bar{\mu}m_{\rm p}n_{\rm e}^2(r)\Lambda(T,Z)}$$

where $n_{\rm e}(r)$ is the electron density and $\rho_{\rm g}(r)$ is the gas density at a radius r. A simple estimate of the instantaneous cooling rate onto the central object can be obtained by following the method proposed by White and Frenk (1991): a cooling radius, $r_{\rm cool}$, can be defined as the radius at which the local cooling time is equal to a suitably defined age for the halo, e.g., the Hubble time. At early times and for low-mass haloes, the cooling radius can be larger than the virial radius. In this case, the hot gas is never expected to be in hydrostatic equilibrium, and the cooling rate is essentially limited by the accretion rate (the halo is in the rapid cooling regime discussed above). At late times and in massive systems, the cooling radius lies within the virial radius, and the gas can be assumed to cool quasistatically with a cooling rate that can be modeled by a simple inflow equation (this is the slow cooling regime):

$$\frac{{\rm d}M_{\rm cool}}{{\rm d}t} = 4\pi\rho_{\rm g}(r_{\rm cool})r_{\rm cool}^2\frac{{\rm d}r_{\rm cool}}{{\rm d}t}$$

The cooling model just described is extremely simplified and does not account, for example, for the fact that the gas distribution can readjust itself once gas starts to cool out (Viola et al. 2008). Nevertheless, it has been shown to provide results that are statistically in relatively good agreement with more detailed hydrodynamical simulations that adopt the same physics (e.g., Benson et al. 2001; Yoshida et al. 2002, but see also Saro et al. 2010). A number of assumptions need to be made, however, to implement the above simple prescriptions in analytic models of galaxy formation (e.g., about the gas profile and for the calculation of the cooling radius). Recent work has shown that the different assumptions adopted can give rise to significant differences, in particular at scales larger than those typical of our own galaxy (De Lucia et al. 2010).

3.3 Star Formation

It is generally accepted that the rate at which galaxies can form stars is determined by its ability to form dense molecular clouds. This is supported by direct observations of associations of young stars in the Milky Way and other nearby galaxies, as well as by observations of CO emission from starburst galaxies. From the theoretical point of view, that of star formation remains a poorly understood mechanism where processes like turbulence, magnetic fields, dust, molecular cooling, etc., all play an important role (a recent review can be found in McKee and Ostriker 2007). In terms of building a galaxy formation model, it is important to understand: (i) where and when the first generation of stars formed and their properties, (ii) the rate at which stars form in disks and starbursts, and (iii) the distribution of stellar masses produced in episodes of star formation.

3.3.1 The First Generation of Stars

If we believe that the structures in the universe grew hierarchically, the first objects that became nonlinear are expected to have masses much smaller than those of typical galaxies. The first generation of stars is expected to be extremely metal poor, because heavy elements can only be produced in the interior of stars. These stars are referred to as Population III stars. In a CDM model, the virial temperature of a halo is related to its mass through the following equation:

$$T_{\rm vir} \sim 442\Omega_m^{1/3}\left(\frac{M}{10^4 h^{-1}{\rm M}_\odot}\right)^{2/3}\left(\frac{1+z_{\rm vir}}{100}\right)K$$

where it has been assumed that the average density of dark matter haloes is 200 times the critical density, $H(z) = H_0\Omega_m^{1/2}(1+z)^3/2$ for $z >> 1$, and $z_{\rm vir}$ is the assembly redshift of the halo. At $z \gtrsim 200$, Compton scattering plays an important role, and the temperature of the intergalactic medium is

$$T_{\rm gas} = T_\gamma = 2.73\,(1+z)\,K$$

where T_γ is the temperature of the cosmic microwave background. At these redshifts, only haloes with $M \gtrsim 10^4{\rm M}_\odot$ can trap significant amount of baryonic gas. At lower redshift, the temperature of the intergalactic medium decreases faster than that of the microwave background, and lower mass haloes start trapping baryonic gas. The gas that is trapped will eventually be heated to the virial temperature of the parent haloes by shocks. If this gas can cool, it will condense and eventually fragment allowing the formation of stars in these early haloes.

As discussed in the previous section, at temperatures lower than ~10^4 K, the only significant cooling is due to molecular hydrogen, so the chemistry of this molecule governs the formation of the first objects. ➤ *Figure 10-7* shows the minimum mass of haloes within which H_2 cooling is sufficiently effective to lead to gas collapse (for a more rigorous exposition, see Ciardi and Ferrara 2005 and references therein). This minimum halo mass turns out to be between $10^4\,{\rm M}_\odot$ at $z \gtrsim 100$ and $10^8{\rm M}_\odot$ at lower redshift. In a CDM universe, haloes in this mass range start forming in large numbers only at $z \lesssim 30$. Significant uncertainties are involved in the discussion outlined above. For example, the presence of UV photons can dissociate hydrogen molecules and therefore suppress significantly the cooling efficiency. On the other hand, induced formation of hydrogen molecules behind shocks driven by the first stars can outweigh its photodissociation. In addition, once star formation begins, mechanical and radiative feedback from the first stars can remove a large fraction of the remaining halo gas.

Because of the very low metallicities, the cooling time of the gas may be significantly longer than the time scale of the gravitational collapse so that the cloud may not be able to fragment. First simulations of this initial collapse phase have confirmed this scenario and concluded that the first stars formed in isolation and were very massive (of the order of 60–100 ${\rm M}_\odot$ but with large uncertainties, see, e.g., Abel et al. 2002; Yoshida et al. 2006). More recent numerical studies have shown that metal-free gas clouds can fragment strongly, with the details of the process depending on the degree of turbulence in the halo. As a consequence, the mass spectrum of Pop III stars might be relatively flat ranging from ~0.1 to ~10 ${\rm M}_\odot$ (Clark et al. 2011; Greif et al. 2011).

3.3.2 A Star Formation Law

The problems related to the formation of the first stars propagate into galaxy formation theory if we wish to understand the rate at which stars form in these systems and any consequences that

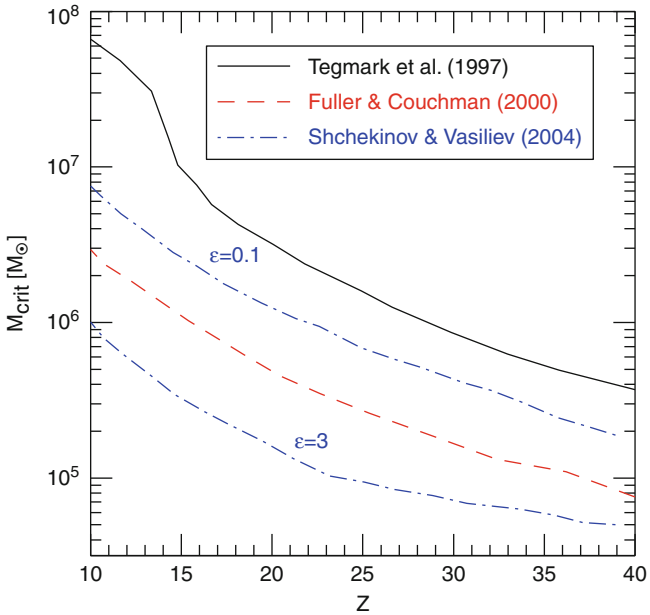

Fig. 10-7

From Ciardi and Ferrara (2005). Minimum mass able to cool and collapse as a function of redshift as calculated in the studies indicated in the legend. The *blue dot-dashed lines* are derived for two different values of the rate of ionizing photon production by ultra-high energy cosmic rays

star formation might have for further evolution. Our limited understanding of the physical processes involved prevents us from constructing a "star formation law" from first principles that describes how the star formation rate $\Sigma_{\rm star}$ depends on the physical conditions of the interstellar medium. In order to make progress, one can appeal to empirical laws. A power-law relation of the form

$$\Sigma_{\rm star} \propto \Sigma_{\rm gas}^{N} \tag{10.4}$$

has been known since a long time (Schmidt 1959) and has been shown to provide a very good parametrization of the global star formation rate over a large range of surface densities, from the gas-poor spiral disks to the cores of the most luminous starburst galaxies (Kennicutt 1998). The best-fit observational data is

$$\Sigma_{\rm star} = (2.5 \pm 0.7) \times 10^{-4} \left(\frac{\Sigma_{\rm gas}}{\mathrm{M}_{\odot \mathrm{pc}^{-2}}} \right)^{1.4 \pm 0.15} \mathrm{M}_{\odot}\ \mathrm{year}^{-1} \mathrm{kpc}^{-2}$$

where $\Sigma_{\rm gas} = \Sigma_{HI} + \Sigma_{H2}$ is obtained averaging over the entire star-forming disk. Recently, it has become possible to study the star formation law by fitting Schmidt laws to individual galaxies for which $\Sigma_{\rm gas}$ and $\Sigma_{\rm star}$ are measured in azimuthally averaged rings or even on a pixel-by-pixel basis. ❯ *Figure 10-8* is based on a recent study by Bigiel et al. 2008 and shows the local star formation rate per unit area measured on a scale of ~750 pc as a function of the local atomic gas density (left panel), molecular gas density (middle panel), and total gas density (right panel). The atomic hydrogen distribution saturates at about $10\,\mathrm{M}_{\odot} pc^{-2}$, and the figure shows that it

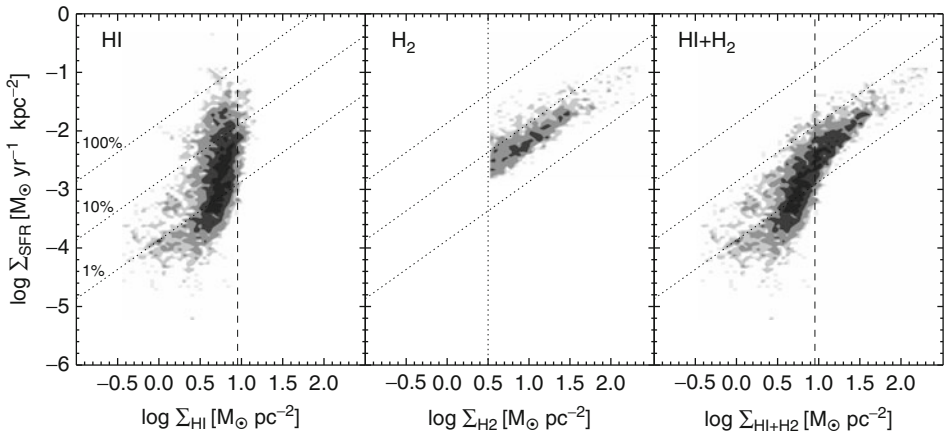

Fig. 10-8

From Bigiel et al. (2010), based on data published in Bigiel et al. (2008). Local star formation rate per unit area (measured on a scale of ~750 pc from seven nearby spirals) as a function of the local atomic gas density (*left panel*), molecular gas density (*middle panel*), and total gas density (*right panel*). The *diagonal dotted lines* show lines of constant star formation efficiency, indicating the level of star formation needed to consume 1%, 10%, and 100% of the gas in 10^8 years. The *dashed vertical lines* in the *left* and *right* panels indicate the surface density at which HI saturates

poorly correlates with the star formation rate measured. Gas in excess of this value is predominantly molecular, and the middle panel of ❯ *Fig. 10-8* shows that there is a well-defined law for this gas component, which is well described by a power law with slope $N \sim 1$. This implies a constant molecular hydrogen depletion time of ~2 Gyr. As argued in Bigiel et al. (2008), the star formation law can be interpreted as a combination of two laws that regulate the conversion of atomic to molecular hydrogen and the formation of stars from molecular gas, respectively. Variants of ❯ 10.4 are commonly adopted in galaxy formation models where the formation of molecular gas is not usually followed explicitly. While this allows us to bypass the question of how stars form, it should be noted that such an empirical relation is then applied also beyond the regimes where it has been originally measured.

3.3.3 The Initial Mass Function

Galaxy properties depend not only on the rate and efficiency of star formation but also on the mass spectrum with which stars form, that is, the initial mass function (IMF). Observational results for our Milky Way suggest that the IMF has roughly the same form, independent of the location in the galaxy. The first determination of the IMF in the solar neighborhood was obtained by Salpeter (1955) who found that it is well described by a power law:

$$\phi(m)dm \propto m^{-b}dm$$

with $b = 2.35$, for stars in the mass range $0.4\,M_\odot \lesssim m \lesssim 10 M_\odot$. $\phi(m)dm$ provides the relative number of stars born with masses in the range $m \pm dm/2$. Different measurements have been

made more recently, and they suggest that the IMF deviates from a pure power law, becoming flatter at the lowest mass end and steeper at the highest mass end. All subsequent determinations do not deviate significantly from the Salpeter IMF for masses $\gtrsim 1\,M_\odot$, while for lower masses, there are significant differences among different determinations. One of the most recent measurements that is widely used in current years has been made by Chabrier (2003):

$$\xi(m) \propto \begin{cases} m^{-1.35} \text{ for } m > 1M_\odot \\ \exp\left(-[\log(m/0.2M_\odot)]^2/0.6\right) \text{ for } m < 1M_\odot \end{cases}$$

where $\xi(m)$ is the logarithmic IMF and is defined as $\xi(m)d\log m = \phi(m)dm$.

The question of the "universality" of the IMF is a heavily debated one, particularly in recent years. From the observational point of view, there are large uncertainties, and only a small number of local star-forming clouds can be studied in detail (for a critical review of observational measurements, see Bastian et al. 2010). From the theoretical point of view, it is worth reminding that the ability of a gas cloud to collapse and fragment depends on the local Jeans mass:

$$M_J \simeq 700\,M_\odot (T/200\,\mathrm{K})^{3/2} (n/10^4\,\mathrm{cm}^{-3})^{-1/2} (\mu/2)^{-2},$$

where T, n, and μ are the temperature, number density, and mean molecular weight at the halt of fragmentation. For gas of primordial composition, a minimum temperature of ~200 K is reached when molecular hydrogen cooling becomes inefficient. This gives a Jeans mass $M_J \simeq 10^3\,M_\odot$. If metals are present, cooling can proceed to lower temperatures allowing the collapsing gas cloud to undergo fragmentation and form smaller clumps. The IMF can therefore depend on the metallicity: the hydrodynamical simulations by Smith and Sigurdsson (2007) show that above a critical metallicity, of about $10^{-3}\,Z_\odot$, clouds can fragment to form low-mass stars, while for gas of lower metallicities, stars form following a more top-heavy IMF. The critical metallicity defined above is well below that of the observed galaxies and, therefore, this effect might not be significant for galaxy formation studies. Further work is, however, needed to clarify if the IMF depends on the metallicity also above this critical value.

3.4 Feedback

The importance and the need of physical mechanisms that are able to modulate the efficiency of galaxy formation as a function of halo mass was recognized early on: Larson (1975, 1976) noted that supernovae-driven winds could remove most of the gas and heavy elements from low-mass galaxies. White and Rees (1978) argued that feedback is required to explain the overall low efficiency of galaxy formation. If dark matter haloes represent the birthplaces of luminous galaxies, evidence for the need of feedback comes from the observation that the shape of the halo mass function is very different from the shape of the observed luminosity function of galaxies. Thus, a simple model that assumes a fixed mass-to-light ratio would overpredict by order of magnitudes both the number of faint galaxies and that of bright ones (see left panel of *Fig. 10-9*). By matching the observed galaxy groups to dark matter haloes that are predicted to have the same space density, it is possible to derive the mass-to-light ratio that guarantees a match between theoretical predictions and the observed luminosity function (see right panel of *Fig. 10-9*). The required mass-to-light ratio is lowest for haloes of mass $\sim 10^{12}\,M_\odot$ which are, in other words, those in which galaxy formation is most efficient. In addition, this simple exercise confirms that the overall efficiency of galaxy formation must be low since most baryons do not end up as stars.

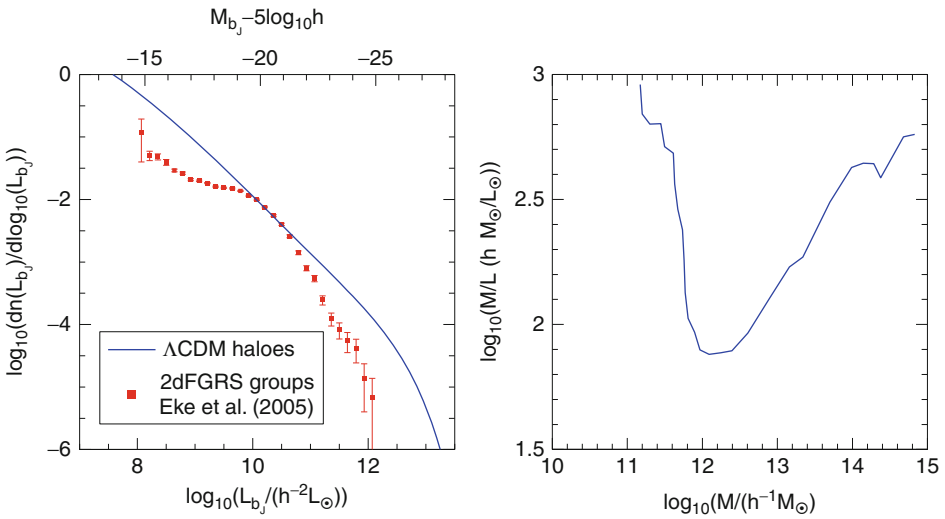

Fig. 10-9

From Baugh (2006). *Left panel*: the *solid line* has been obtained converting the dark matter halo mass function into a galaxy luminosity function by assuming a fixed mass-to-light ratio chosen to match the knee of the luminosity function. Data points with error bars show observational measurements. *Right panel*: the mass-to-light ratio required to match the observed luminosity function

We now know that feedback processes are those that have arguably the strongest influence on the observed galaxy properties but also those that are the most difficult to model. Broadly speaking, galaxy formation models consider three different forms of feedback: photoionization heating, supernovae feedback, and feedback from active galactic nuclei (AGN). The first two are believed to play an important role in shaping the faint end of the luminosity function, while the latter is believed to play a crucial role in regulating the condensation of gas in relatively massive haloes, thereby reducing the number of bright galaxies that would be predicted in the simple model outlined above. The following sections describe in more details these three modes of feedback and comment on recent theoretical results.

3.4.1 Photoionization Heating

It is believed that the hydrogen in the intergalactic medium must have been reionized somewhere between $z \sim 6$ and $z \sim 30$. Although it remains uncertain which energy sources were responsible for reionization, it was soon realized that the photoionizing background responsible for reionizing the intergalactic medium may also act to inhibit galaxy formation. In particular, this process acts in two different ways: (i) it heats the gas increasing its thermal pressure and therefore inhibiting its accretion onto dark matter haloes, and (ii) it also heats the gas that has already collapsed in haloes, therefore reducing the abundance of neutral atoms which can be collisionally excited, which in turn reduces the rate of radiative cooling of gas inside haloes (Doroshkevich et al. 1967; Efstathiou 1992). Both these mechanisms can effectively suppress galaxy formation in small haloes. ➤ *Figure 10-10* shows examples of net

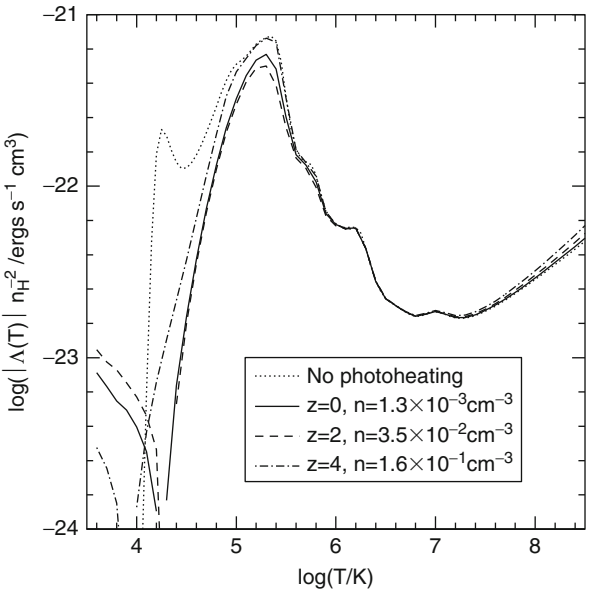

Fig. 10-10

From Benson et al. (2002). The net cooling/heating function for gas at different redshifts in the presence of the photoionizing background predicted in a fiducial model of galaxy formation. The absolute value of the cooling/heating rate is plotted per unit volume, for a gas with metallicity 0.3 $Z_\odot$

heating/cooling functions in the presence of a photoionizing background. These were calculated coupling the photoionization background computed self-consistently from a galaxy formation model with a photoionization code (for details, see Benson et al. 2002) and are computed for the typical densities of gas at each redshift indicated in the legend. The figure shows that photoionization can significantly suppress cooling in haloes with virial temperature in the range 10^4–10^5 K and therefore inhibit the formation of low-mass galaxies.

The value of the characteristic mass, M_c, below which galaxies are strongly affected by photoionization was calculated by Gnedin (2000) who argued that $M_c = M_F$, i.e., the filtering mass that corresponds to the scale over which baryonic perturbations are smoothed in linear perturbation theory. This relation has often been used in galaxy formation models to explain the low number of satellites observed in the Local Group. Recent numerical work has shown that the fitting function provided by Gnedin overestimates the characteristic mass by large factors (Okamoto et al. 2008).

3.4.2 Supernovae Feedback

The mechanical energy supplied by massive stars in the form of supernovae and stellar winds represents the engine that drives the galactic-scale outflows that are observed in the most actively star-forming galaxies both in the local universe and at high redshift. Observations suggest that outflows are ubiquitous in galaxies in which the global star formation rate per

unit area exceeds roughly $10^{-1} M_{\odot}\,\mathrm{year}^{-1}\,\mathrm{kpc}^{-2}$ and that the material is multiphase containing cold, warm, and hot gas, plus dust and magnetized relativistic plasma. Different techniques and datasets can be used to estimate the mass and energy outflow rates. The available data suggest that the outflow rates are comparable to the star formation rates and that radiative losses in superwinds are not significant (Heckman 2002, and references therein). The estimated outflow speeds can be significant (in the range from hundreds to thousands $\mathrm{km\,s}^{-1}$), and recent studies suggest that there is a weak trend with the galaxy star formation rate (Weiner et al. 2009, see also Martin 2005). This trend seems to support a picture in which winds are momentum driven through radiation pressure (Murray et al. 2005) rather than by the kinetic energy of supernova ejecta by entrainment in the hot wind (Strickland and Stevens 2000, and references therein). Unfortunately, the observational measurements available refer to material that is still relatively deep within the gravitational potential of the halo. So the estimated outflow rates are difficult to translate into rates at which mass, metals, and energy escape from galaxies and are eventually transported into the intergalactic medium. The fate of the winds (or superwinds depending on their velocity) will depend critically on a number of unknowns and on the multiphase nature of the outflowing material.

The dynamical evolution of a starburst-driven outflow has been studied using hydrodynamical simulations. The deposition of mechanical energy by supernovae and stellar winds creates an overpressured cavity of hot gas inside the starburst. This cavity expands, sweeping up ambient material and developing a bubble-like structure. If the ambient medium is stratified (like in a disk), the bubble expands most rapidly in the direction of the vertical pressure gradient. Numerical simulations show that when the bubble size reaches several disk vertical scale heights, it is fragmented because of Raleigh-Taylor instabilities which allow the hot gas to blow out of the disk into the halo in a weakly collimated bipolar outflow (a "wind"). *Figure 10-11* shows snapshots of the density distribution for models of galaxies with different initial masses and mechanical luminosities (as indicated in the legend). These simulations, carried out by Mac Low and Ferrara (1999), also showed that it is relatively difficult for supernovae-driven outflows to remove the gas from low-mass galaxies, while they appear to be fairly efficient at ejecting metals. This "metal loading" effect is just another manifestation of the importance of the multiphase nature of the outflow material and is often neglected in models of galaxy formation.

As mentioned above, supernovae feedback is believed to play a very important role in regulating the number of faint galaxies (Benson et al. 2003) but also in shaping the mass-metallicity relation that is observed for star-forming galaxies (Tremonti et al. 2004) and in enriching the intergalactic and intracluster medium (De Lucia et al. 2004b). Given the uncertainties discussed above, this process is included in galaxy formation models using a number of different prescriptions that are based on observations and/or theoretical arguments. Currently, it is difficult to argue that one specific model is and/or works better than another.

3.4.3 AGN Feedback

AGN can release huge amounts of energy during their lifetimes. Assuming an energy conversion efficiency of ϵc^2 per unit of accreted mass, one finds that an accreting black hole liberates $\sim 10^{19}(\epsilon/0.01)$ erg per gram, and it is easy to compute that this energy input can easily exceed the binding energy of the host galaxy (Begelman 2004). Broadly speaking, there are two different forms of feedback.

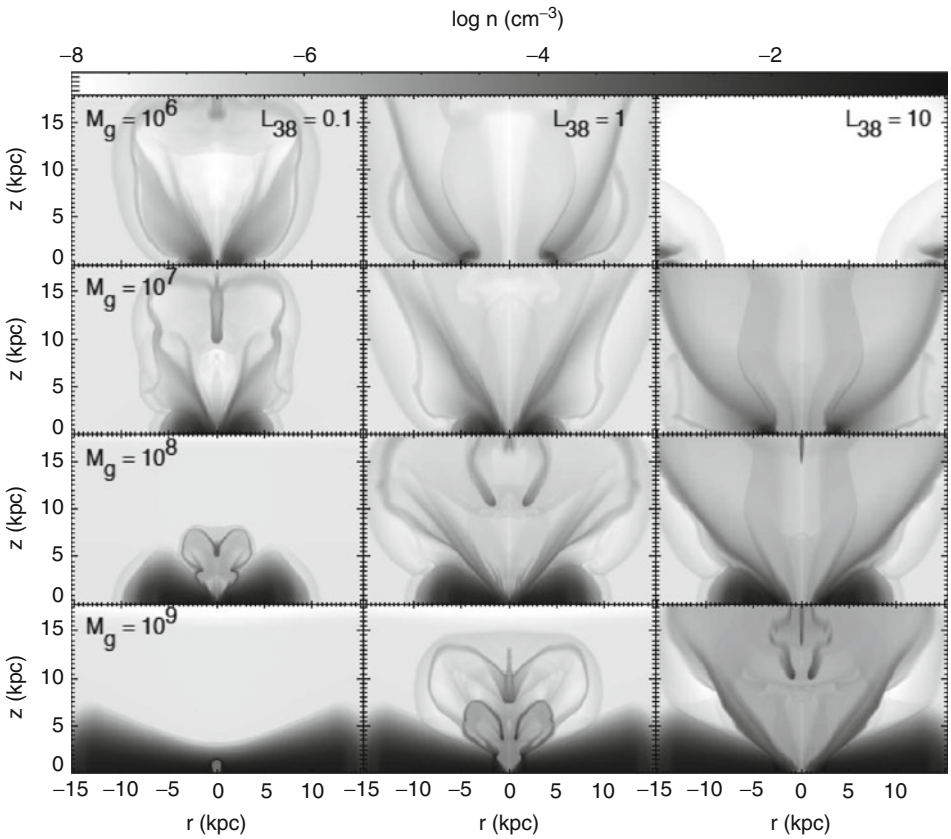

Fig. 10-11

From Mac Low and Ferrara (1999). Density distributions for models with different initial masses M_g and mechanical luminosities L in units of 10^{38} ergs s^{-1} after 75 Myr (energy input ended at 50 Myr)

The photons generated by AGN can ionize and heat the surrounding gas. If the host galaxy contains significant amounts of dust, the radiation pressure on the dust grains can overcome the gravitational force of the halo, generating a momentum-driven wind (Murray et al. 2005). This is a "radiative" feedback mode, and in the literature, it is sometimes referred to as "quasar mode." This energy injection mechanism is effective in broad-line quasars for which the presence of high velocity winds has been confirmed in a number of cases. Galaxy formation models incorporate this energy injection channel in mergers of gas-rich galaxies that can funnel copious amounts of cold gas toward the central regions of galaxies and feed the central black holes with high gas accretion rates. This particular form of feedback is therefore believed to be important at higher redshift where the activity of quasars peaks. Numerical simulations show that it can affect significantly the physical properties of the host galaxy, by expelling large amounts of gas and therefore suppressing significantly subsequent star formation (Springel et al. 2005a).

When the accretion rate onto the central black hole is low, AGN can drive highly collimated and powerful jets which can reach out well into the surrounding halo. This is a "mechanical" feedback mode, sometimes referred to as "radio mode." Evidence for this form of feedback can

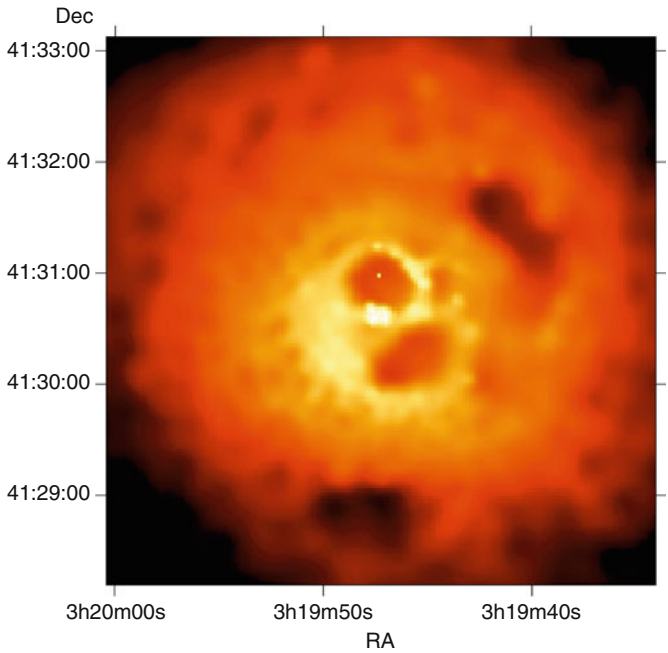

Fig. 10-12
From Fabian et al. (2000). Adaptively smoothed 0.5–7 keV Chandra image of the core of the Perseus cluster

be seen in the central regions of galaxy clusters: X-ray observations show that these often contain cavities that are filled with relativistic gas and are believed to be inflated by the jet launched from the central black hole. *Figure 10-12* shows, for example, an X-ray image (adaptively smoothed) of the central region of the Perseus cluster that contains a bright radio source at its center. It has long been realized that this form of feedback can provide a solution to the "cooling flow" problem, i.e., the observation that the gas at the center of most galaxy clusters is apparently not condensing and turning into stars, although the observed X-ray emission implies a cooling time that is much shorter than the age of the system (Tabor and Binney 1993). The ensemble-averaged power from radio galaxies seems sufficient to offset the mean level of cooling, and a large fraction of central cluster galaxies are radio loud (Best et al. 2007). The steep dependence of the radiative cooling function on density makes, however, difficult to stabilize cooling flows so that heating approximately balances cooling at all radii. Numerical simulations show that the efficiency of this feedback in suppressing gas condensation depends strongly on a number of unknown parameters, e.g., the duty cycle (i.e., the frequency of the energy injection), the geometry, and gas viscosity (e.g., Sijacki and Springel 2006).

In addition to the two modes of AGN feedback described above, significant outputs of energetic particles (cosmic rays) or exotic particles consisting, for example, of relativistic neutrons and neutrinos can contribute to inject energy into the surrounding gas. The precise composition of the bubbles, in these terms is, however, not well known, both from an observational and a theoretical point of view. Thus, it is currently unclear if the energy released via this channel is significant.

3.5 Chemical Enrichment

As explained above, the first generation of stars (the Pop III stars) formed from gas with primordial composition. Stellar nucleosynthesis and the subsequent pollution of the interstellar and intergalactic medium (through, for example, supernovae-driven winds) affect the formation of later stellar populations. In particular, the presence of heavy elements increases significantly the rate at which gas can cool and affects the luminosity and colors of the stellar populations. In addition, star formation also leads to the formation of dust which attenuates the optical and ultraviolet light of galaxies and re-emits at longer wavelengths.

The final stages of stellar evolution and metal production depend on the stellar mass. Stars with masses $\lesssim 8 M_\odot$ end up their life as C/O white dwarfs, after an asymptotic giant branch (AGB) phase during which the star is burning helium in an inner shell and hydrogen in an outer shell. Unfortunately, there are still large theoretical uncertainties in the treatment of convection and mass loss from AGB stars. If the C/O white dwarf is part of a close binary system, it can accrete material from the companion. When the star reaches the Chandrasekhar limit ($1.4\,M_\odot$), it explodes as supernovae type Ia which dominate the production of elements in the iron peak. Massive stars (with masses $\gtrsim 8 M_\odot$) enrich the interstellar medium with metals via both stellar winds and their final explosions as core-collapse supernovae (type II SNe). These are primarily responsible for the production of α elements (among which oxygen, magnesium, silicon, calcium) but also of other elements like nitrogen and sodium. Since the progenitors of type II SNe are massive stars with lifetimes shorter than $\sim 10^7$ year, while the progenitors of type Ia SNe are less massive stars with lifetimes $\gtrsim 10^8$ year, the relative proportions of the metal species they contribute (often quantified in the $[\alpha/\mathrm{Fe}]$ ratio) provide information on the time scale of star formation. So studying the metallicity and $[\alpha/\mathrm{Fe}]$ ratio of galaxies, it is possible to constrain the time scale over which star formation took place. The only complication is given by the fact that the $[\alpha/\mathrm{Fe}]$ ratio does not depend only on the star formation history but also on the shape of the IMF.

In the framework of galaxy formation models, chemical evolution has often been (and still largely is) included using the instantaneous recycling approximation (i.e., the models neglect the finite lifetime of stars so that both chemical enrichment and gas recycling are assumed to take place at the same time of star formation) and a constant yield that is usually treated as a free parameter. More recent studies have included a more accurate treatment of type Ia supernovae and are able to follow the evolution of individual elements (e.g., Nagashima et al. 2005; Arrigoni et al. 2010).

3.6 Galaxy-Galaxy Interactions

In the hierarchical scenario, dark matter haloes (and therefore the galaxies that reside in them) undergo frequent interactions with each other. These interactions have dramatic influence on the morphologies and star formation histories of the galaxies involved. Numerical simulations have shown that close interactions can lead to a strong internal dynamical response driving the formation of spiral arms and, depending on the structural properties of the disks, of strong bar modes. The developing nonaxisymmetric structures (spiral arms and/or central bars) lead to a compression of the gas that can fuel starburst/AGN activity (see Mihos 2004, and references therein). Simulations have also shown that in sufficiently close encounters between

galaxies of similar mass, violent relaxation completely destroys the disk and leaves a kinematically hot remnant with photometric and structural properties that resemble those of elliptical galaxies.

The merger hypothesis for the formation of elliptical galaxies was suggested early on by Toomre and Toomre (1972) and later confirmed by many numerical simulations (Mihos 2004; Cox et al. 2008, and references therein). In recent years, a large body of observational evidence has been collected that demonstrates that a relatively large fraction of early-type systems show clear evidence of interactions, mergers, and recent star formation, in particular at high redshift. However, the data also seem to indicate that only a small fraction of the final mass is involved in these episodes. This observational result has often been interpreted as strong evidence against the somewhat extended star formation history naively predicted from hierarchical models. A related issue concerns the α-element enhancements observed in elliptical galaxies. As explained above, the [α/Fe] ratio is believed to encode important information on the time scale of star formation, and it is a well-established result that massive ellipticals have supersolar [α/Fe] ratios, suggesting that they formed on relatively short time scales and/or have an initial mass function that is skewed toward massive stars. The inability of early models of the hierarchical merger paradigm to reproduce this observed trend has been pointed out as a serious problem for these models (Thomas 1999).

In order to model galaxy interactions and mergers, one needs to know what determines the structural and physical properties of a merger remnant. Numerical simulations have shown that these depend mainly on the following two factors:

1. The progenitor mass ratio. As mentioned above, during "major" mergers, violent relaxation plays an important role, and as a consequence, the merger remnant has little resemblance to its progenitors. On the other hand, during minor mergers, the interaction is less destructive so that the merger remnant often resembles its most massive progenitor. The exact value at which one distinguishes between minor and major mergers is somewhat arbitrary but is usually chosen to be of the order of $M_2/M_1 \sim 0.3$.
2. The physical properties of the progenitors. The structure of the galaxies involved in a merger plays an important role in determining the response to interactions: disks that are stable against the growth of instabilities (e.g., because of a central bulge or a lowered disk surface density) will be less "damaged" than disk-dominated systems that are prone to strong instabilities. In addition, in a merger between two gas-rich progenitors, a significant fraction of the gas content can be fuelled toward the center, triggering a starburst and/or accretion of gas onto the central black hole. Merger-driven starbursts are instead suppressed if the two merging systems are gas poor. These purely stellar mergers are often referred to as a "dry" or "red," and as will be discussed below, they are believed to contribute significantly to the recent assembly of elliptical galaxies.

Figures 10-13 and *10-14* show the projected stellar and gas mass density, respectively, during a merger with baryonic mass ratio 2.3:1. The figures show that the satellite galaxy first makes a fast, direct approach toward the primary galaxy. The tidal interaction between the two merging disks generates symmetric tails in both of them. Due to the initial orbital energy, the two galaxies separate again for several orbital periods (~1 Gyr in the particular case shown) before getting closer again. After the first or second passage, the initial angular momentum is lost, and the orbit becomes almost entirely radial. This limits the coupling between orbital and spin angular momentum and therefore the tidal response during the final coalescence.

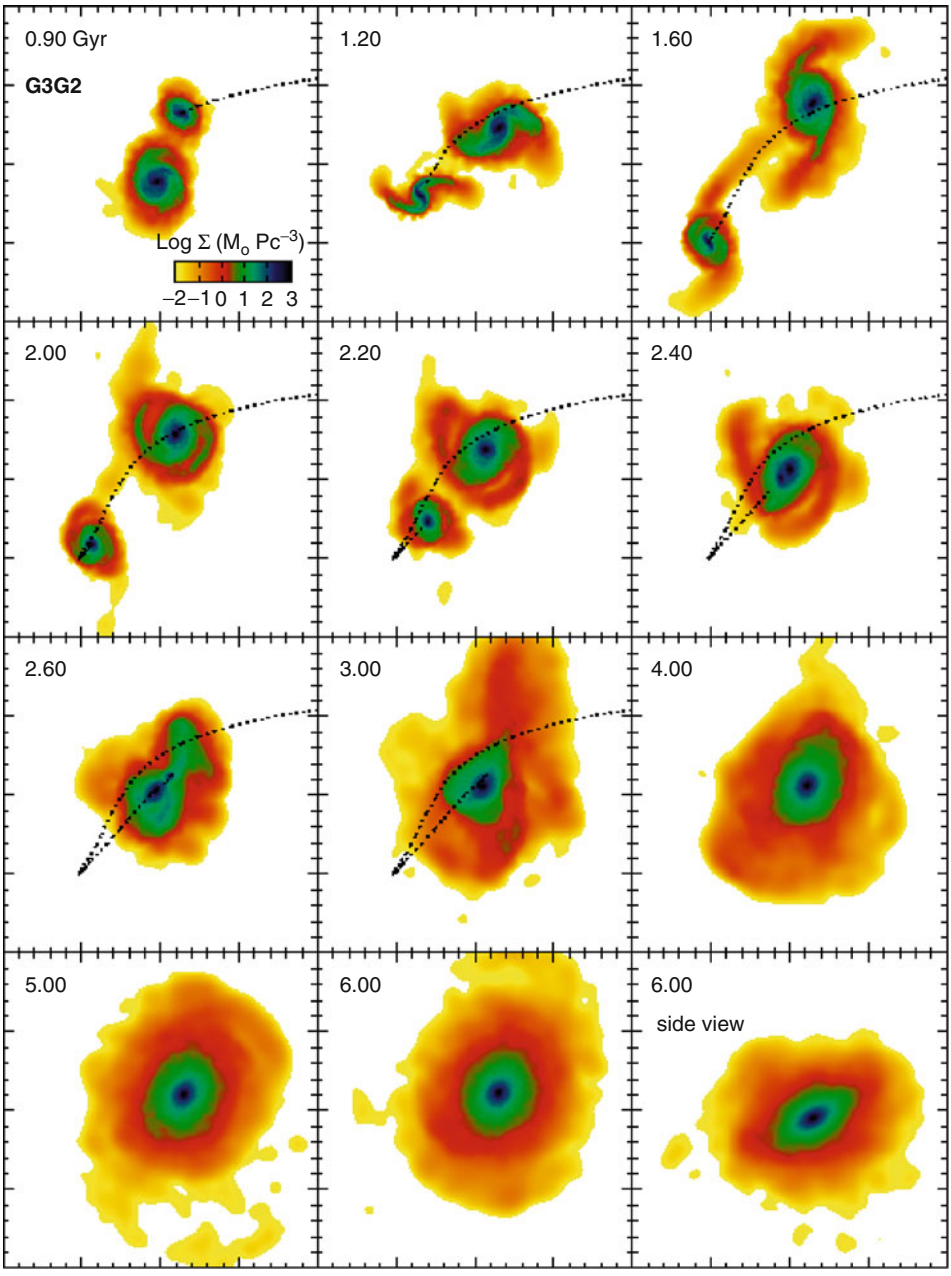

Fig. 10-13

From Cox et al. (2008). Projected stellar mass density during a merger simulation with mass ratio 2.3:1, as viewed in the orbital plane. Each panel measures 200 kpc on a side. The time (in Gyr) is displayed in the *upper left-hand* side of each panel. The orbit of the satellite galaxy G2 is denoted by a *dotted line* until it has completely merged with the primary galaxy. The *bottom right-hand panel* shows a side view of the final merger remnant

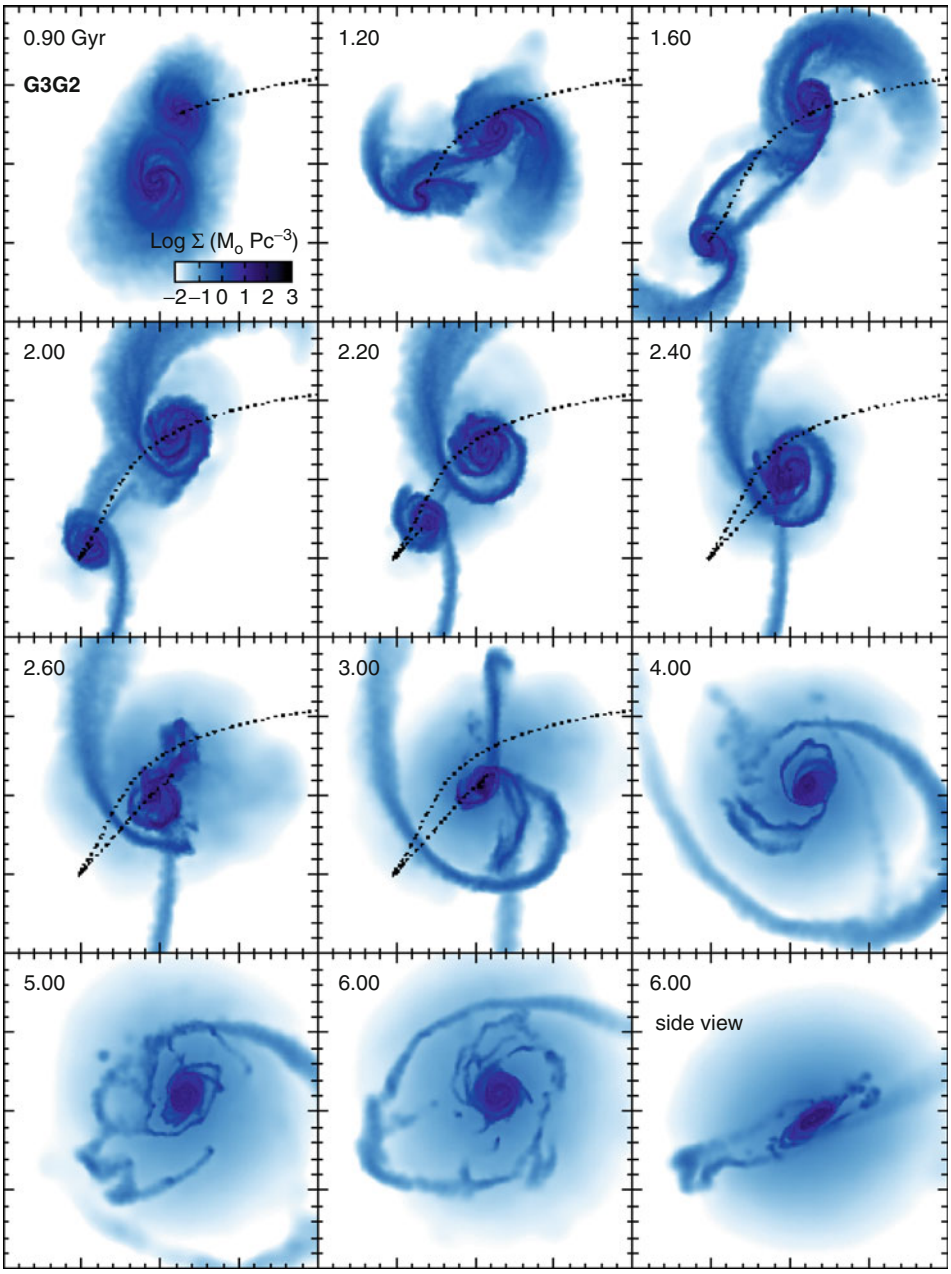

Fig. 10-14
From Cox et al. (2008). Similar to *Fig. 10-13* but for the gaseous component

The time scales of the galaxy mergers depend significantly on the orbital parameters that, to some extent, also affect the structural properties of the remnant. For example, the relative orientation of the orbital spin with respect to the intrinsic spin of the progenitors influences significantly the prominence of tidal tails. A good approximation of the merging times of galaxies is provided by the classical Chandrasekhar (1943) dynamical friction formula:

$$\frac{d}{dt}\vec{v}_{\rm orb} = -4\pi G^2 \ln(\Lambda)\, M_{\rm sat}\, \rho_{\rm host}(< v_{\rm orb})\, \frac{\vec{v}_{\rm orb}}{v^3_{\rm orb}},$$

where $\rho_{\rm host}(<v_{\rm orb})$ is the density of background particles with velocities less than the orbital velocity $v_{\rm orb}$ of the satellite, $M_{\rm sat}$ is the mass of the satellite, and Λ is the Coulomb logarithm that depends on the mass ratio between the two merging galaxies. The formula given above, that is valid in the approximation of a point mass satellite and a uniform background mass distribution, is often adopted in analytic models of galaxy formation to estimate the time scale for an orbiting satellite to lose its energy and angular momentum and merge with the central galaxy of its host halo. Recent work has, however, shown that the classical dynamical friction formula tends to underestimate merging times computed from controlled numerical experiments and high-resolution cosmological simulations (e.g., Boylan-Kolchin et al. 2008). In addition, it should be noted that different models usually assume different variations of the classical formula (e.g., adopt a different "fudge" factor and/or a different expression for the Coulomb logarithm) that can lead to significant differences in the estimated merger times (see Sect. 8 and Fig. 14 in De Lucia et al. 2010).

3.7 The Environment

The distribution of galaxies on the sky is not uniform: galaxies appear to be arranged in a complex web of filaments and sheets that surround empty "voids" and intersect in dense nodes that can contain up to thousands galaxies, the rich clusters of galaxies. It has been known for a long time that the local and large-scale environments play an important role in determining many galaxy properties. First indications of a correlation between the galaxy type and the environment can be found in the *The Realm of Nebulae* by E. Hubble (1936), but the milestone paper in the subject is probably the work by Dressler (1980), who showed the existence of a well-defined relationship between local galaxy density and galaxy type for a sample of ~20 massive nearby clusters.

Disentangling the processes responsible for the observed correlations has proved difficult, and it remains unclear whether the observed relations are imprinted during formation or by physical processes at work preferentially in dense environments. The difficulty is in part intrinsic: according to the current paradigm for structure formation, dark matter collapses into haloes in a bottom-up fashion. Small systems form first and subsequently merge to form progressively larger systems. As structure grows, galaxies join more and more massive systems, therefore experiencing different "environments" during their lifetimes. In this context, it is clear that both "heredity" (i.e., the initial conditions) and "environment" (i.e., subsequent physical processes that galaxies experience during their lifetimes) do play a role in shaping the observed galaxy properties and in determining the observed environmental trends.

A number of different physical mechanisms have been early identified that can influence significantly the physical properties of galaxies in a cluster environment. Broadly speaking, they can be grouped in two big families: (i) interactions with other cluster members and with the

cluster potential well, and (ii) interactions with the hot gas that is known to permeate galaxy clusters. In the following, the specific mechanisms often considered when trying to assess the influence of the environment on galaxy evolution are discussed in more detail.

3.7.1 Galaxy Harassment

Galaxy mergers and more generally strong galaxy–galaxy interactions are commonly viewed as a rarity in massive clusters because of the large velocity dispersion of the system. It should be noted, however, that they are still important in the outskirts of galaxy clusters, and they were certainly more efficient in the infalling group environment. Therefore, they may represent an important "preprocessing" step in the evolution of cluster galaxies. In rich clusters, the encounters between galaxies will be generally high-speed interactions. The colliding galaxy is impulsively heated and becomes less bound and more vulnerable to disruptions by further encounters and by tidal interactions with the global cluster potential.

The cumulative effect of repeated, numerous fast encounters is usually referred to as "harassment." This process has been discussed in early work on the dynamical evolution of cluster galaxies (Richstone 1976) and has been explored in detail using numerical simulations (Farouki and Shapiro 1981; Moore et al. 1998). These have confirmed that repeated high-speed collisions, coupled with the effects of the global tidal field of the cluster, can drive a strong response in cluster galaxies. The efficiency of the process is, however, largely limited to low-luminosity hosts due to their slowly rising rotation curves and their low-density cores. Therefore, it is believed that harassment might have an important role in the formation of dwarf ellipticals or in the destruction of low-surface brightness galaxies in clusters, but it is less able to explain the evolution of luminous cluster galaxies.

Figure 10-15 shows the evolution of the stellar surface density of a galaxy that is orbiting close to the center of a galaxy cluster. The first stages of the evolution are characterized by the formation of a strong bar and of an open spiral pattern that is, however, easily stripped by tidal interactions. In contrast, the bar appears to be quite stable. It undergoes strong "buckling" instabilities that make the central part of the galaxy more spherical. In the final stages, the galaxy resembles a drawf spheroidal system.

3.7.2 Cannibalism

Early theoretical studies have discussed the role of cannibalism due to dynamical friction in the formation of brightest cluster galaxies (e.g., Ostriker and Tremaine 1975). This early work, however, significantly overestimated the efficiency of the process due to different simplified assumptions adopted. In the now standard paradigm of structure formation, clusters assemble quite late, through the merging of smaller systems. In this perspective, cooling flows are the main fuel for galaxy formation at high redshift, in dense and lower-mass haloes. This source is removed at lower redshift possibly due to feedback from AGN. Galaxy-galaxy mergers, as discussed above, are most efficient within small haloes with low-velocity dispersions. These mergers are indeed driven by dynamical friction, but it is the accretion rate of the galaxies onto the protocluster, along with the cluster growth itself, that regulate and set the conditions for galaxy merging.

This is illustrated very nicely in *Fig. 10-16* which shows the merger tree of the central galaxy of a cluster-sized halo (for details, see De Lucia and Blaizot 2007). The brightest cluster

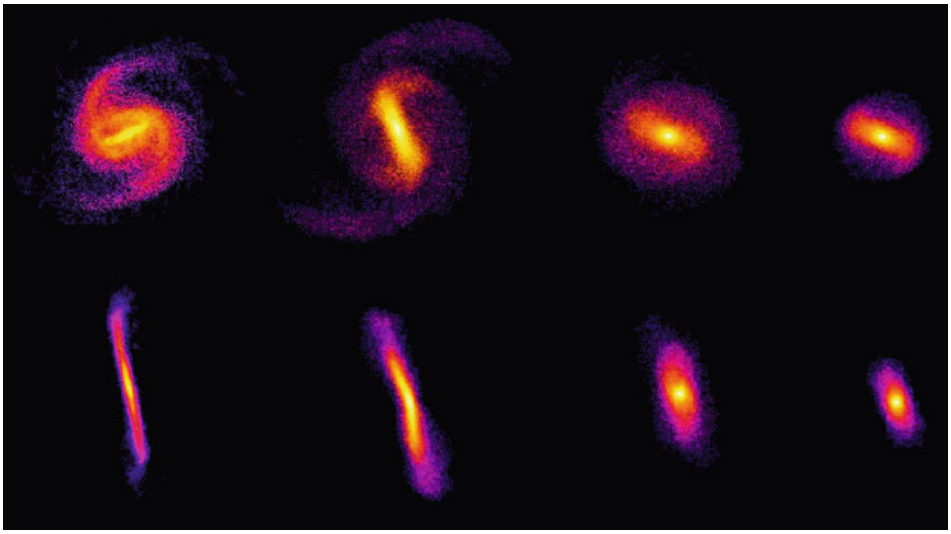

Fig. 10-15
From Mastropietro et al. (2005). Evolution of the stellar surface density of a galaxy that is orbiting close to the center of a galaxy cluster at $z = 0$. The *top panels* represent the face on projections, while the edge on projections are shown in the *bottom panels*

galaxy (BCG) itself lies at the top of the plot (at $z = 0$), and all its progenitors (and their histories) are plotted downward going back in time recursively. Galaxies with stellar mass larger (resp. smaller) than $10^{10}\ h^{-1}\ \mathrm{M}_{\odot}$ are shown as symbols (resp. lines) and are color-coded according to their rest-frame B-V color. The leftmost branch in *Fig. 10-16* represents the "main branch," obtained by connecting the galaxy at each time step to the progenitor with the largest stellar mass at the immediately preceding time step (the "main progenitor").

Figure 10-16 shows another important point: in the context of the hierarchical paradigm for structure formation, the full history of a galaxy is described by its complete merger tree. Whereas in the "monolithic" approximation, the history of a galaxy can be described by a set of functions of time, hierarchical histories are difficult to summarize in a simple form because even the identity of a galaxy is ill-defined. A galaxy is no more a single object when viewed at different times but the ensemble of its progenitors, all of which need to be taken into account for a correct characterization of the stellar population of the final object. It is also interesting to note that although the merger trees of these central galaxies have a very large number of branches, only a small fraction of these contribute significantly to the buildup of their stellar mass: in this particular example, ~70% of the mass comes from the accretion of 12 galaxies more massive than $10^{10}\ h^{-1}\ \mathrm{M}_{\odot}$ (see also De Lucia et al. 2006).

3.7.3 Ram-Pressure Stripping

Galaxies travelling through a dense intracluster medium suffer a strong ram-pressure stripping that can sweep cold gas out of the stellar disk (Gunn and Gott 1972). Depending on the binding energy of the gas in the galaxy, the intracluster medium will either blow through the galaxy, removing some of the diffuse interstellar medium, or will be forced to flow around the galaxy

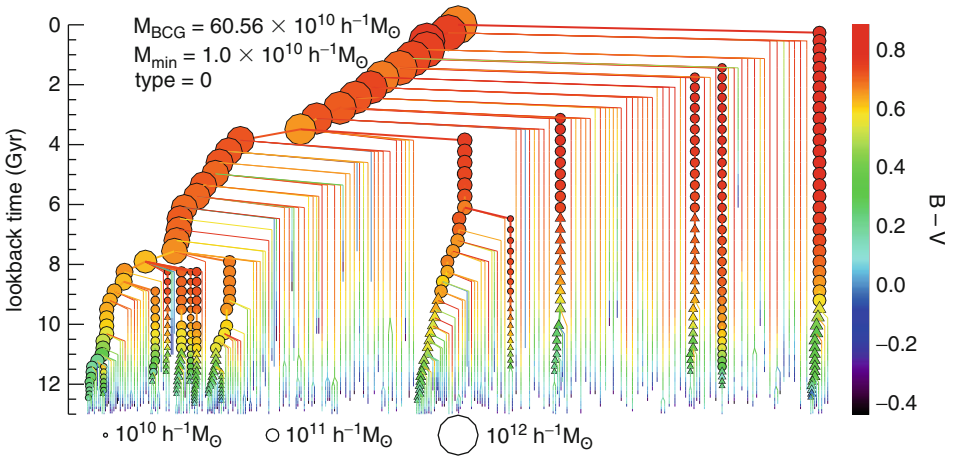

Fig. 10-16

From De Lucia and Blaizot (2007). BCG merger tree. Symbols are color-coded as a function of B-V color and their area scales with the stellar mass. Only progenitors more massive than $10^{10}\ M_\odot\ h^{-1}$ are shown with *symbols*. *Circles* are used for galaxies that reside in the FOF group inhabited by the main branch. *Triangles* show galaxies that have not yet joined this FOF group

(Cowie and Songaila 1977; Nulsen 1982). Ram-pressure stripping is expected to be more important at the center of massive systems because of the large relative velocities and higher densities of the intracluster medium. By considering the distribution and history of ram-pressure experienced by galaxies in clusters, Brüggen and De Lucia (2008) estimated that strong episodes of ram-pressure are indeed predominant in the inner core of the clusters. They also showed, however, that virtually all cluster galaxies suffered weaker episodes of ram-pressure, suggesting that this physical process might have a significant role in shaping the observed properties of the entire cluster galaxy population. In addition, Brüggen and De Lucia found that ram-pressure fluctuates strongly so that episodes of strong ram-pressure alternate to episodes of weaker ram pressure, possibly allowing the gas reservoir to be replenished and intermittent episodes of star formation to occur.

Ample observational evidence that ram-pressure is occurring is available, and the process has been extensively studied using hydrodynamical simulations. *Figure 10-17* shows snapshots from the simulations carried out by Quilis et al. (2000). The figure shows a galaxy moving face on and nearly edge on through the core of a rich cluster at a velocity of ~2,000 km s^{-1}. These simulations showed that the time scale for stripping is very short compared to the orbital time scale, and that the multiphase structure of the interstellar medium and the presence of bubbles and holes make the disk more susceptible to viscous stripping. A simple estimate of the efficiency of ram-pressure was obtained by Gunn and Gott (1972), by comparing the ram pressure with the galactic gravitational restoring force per unit area. This leads to the following condition:

$$\rho_{\rm ICM} > \frac{2\pi G \Sigma_\star \Sigma_{\rm ISM}}{V^2}$$

where $\rho_{\rm ICM}$ is the density of the intracluster medium, V is the velocity of the galaxy, and $\Sigma_{\rm star}$ and $\Sigma_{\rm ISM}$ are the mean stellar and gaseous surface density of the disk. Early numerical

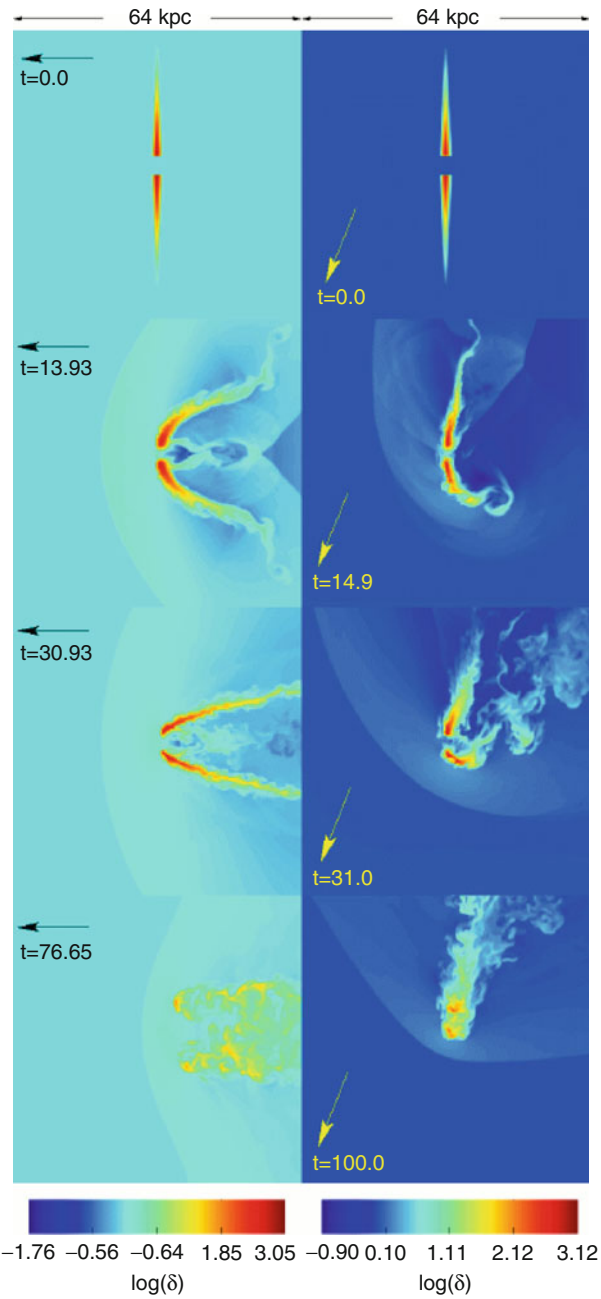

Fig. 10-17

From Quilis et al. (2000). The evolution of the gaseous disk of a spiral galaxy moving face on (*left column*) and inclined 20° to the direction of motion (*right column*) through a diffuse hot intracluster medium. Each *snapshot* shows the density of gas ($\delta = \rho/\rho_{ICM}$) within a 0.2-kpc slice through the center of the galaxy, and each frame is 64 kpc on a side

simulations showed that this analytical estimate fares fairly well, as long as the galaxies are not moving close to edge on. More recent numerical work (e.g., Roediger and Brüggen 2007) has shown that this formulation often yields incorrect mass-loss rates. In addition, simple models based on the Gunn and Gott formula usually do not consider the possibility that ram-pressure stripping could temporarily enhance star formation.

3.7.4 Strangulation

Current theories of galaxy formation assume that when a galaxy is accreted onto a larger structure, the gas supply can no longer be replenished by cooling that is suppressed because of the removal (by tides and ram-pressure) of the hot-gas halo associated with the infalling galaxy (Larson et al. 1980). This process is usually referred to as "strangulation" (or "starvation" or "suffocation"). It is common to read in discussions related to these physical mechanisms that strangulation is expected to affect the star formation of cluster galaxies on relatively long time scales, and therefore to cause a slow decline of the star formation activity. As we will see below, however, in recent galaxy formation models, this process is usually associated to a strong supernovae feedback. As a consequence, galaxies that fall onto a larger system consume their cold gas rapidly, moving onto the red sequence on very short time scales.

Traditionally, in galaxy formation models, the stripping of the hot gaseous reservoir has been assumed to be complete and instantaneous. Using a suite of controlled full hydrodynamic simulations, however, McCarthy et al. (2008) have found that a fraction (about 30%) of the initial hot galactic halo gas can remain in place even after 10 Gyr. Saro et al. (2010) have confirmed that cooling can occur on satellite galaxies, but this seems to be limited to the most massive ones. In these satellites, the star formation can last for up to ~1 Gyr after accretion, albeit significantly suppressed with respect to the average value before accretion.

3.8 Stellar Populations

Observational studies of galaxies make use of the radiation emitted by them to infer their physical properties. In order to make a close link between model predictions and observational data, it is therefore necessary to compute the luminosity emitted by the galaxy as a function of wavelength or frequency. Analytically, the spectral energy distribution of a galaxy can be expressed as the superposition of numerous "single-stellar populations" (SSPs) that are populations of stars with the same age, initial mass function, and chemical composition. The luminosity of each of these SSPs can be written as

$$L_\nu^{(\mathrm{SSP})}(t, Z, \phi) = \int_{M_{\min}}^{M_{\max}} \phi(M') L_\nu^{(\mathrm{star})}(t, Z) dM'$$

where $M_{\min}$ and $M_{\max}$ are the minimum and maximum mass for stars, respectively, $\phi(M)$ is the initial mass function, and $L_\nu^{(\mathrm{star})}(t, Z)$ is the spectrum of a single star of metallicity Z and age t. If the luminosity of the SSPs is known as a function of age and metallicity, then the luminosity of a galaxy can be written as

$$L_\nu^{(\mathrm{galaxy})} = \int_0^t dt' \int_0^\infty dZ' \dot{M}_\star(t', Z') \times L_\nu^{(\mathrm{SSP})}(t - t', Z', \phi) \tag{10.5}$$

where $\dot{M}_\star(t, Z)$ gives the rate at which stars of metallicity Z form at the time t inside the galaxy.

Several libraries are available in the literature which provide $L_\nu^{(\mathrm{SSP})}(t, Z, \phi)$ for different ages, metallicities, and initial mass functions (e.g., Bruzual and Charlot 2003; Maraston et al. 2009; Conroy et al. 2009). These libraries are constructed using a combination of theoretical stellar evolution models, direct observations of stars for which age and metallicity can be measured, and theoretical models of stellar atmospheres where no observations are available. In the framework of galaxy formation, these population synthesis models are usually treated as "black boxes." It is important, however, to remember that significant uncertainties remain in many of their ingredients. The AGB regime, for example, is very difficult to treat because of the pulsational regime, the double-shell burning, and especially the strong mass losses affecting this phase. This leads to large uncertainties in the evolution of the spectral continuum in rest-frame near infrared (for a review, see Maraston 2011).

Real galaxies are not made only of stars but also contain gas (both hot and cold) and dust. This can significantly affect the observed luminosities in the ultraviolet and in the optical and even dominate the luminosity in the far-infrared portion of the spectrum. Indeed dust, that is believed to be produced in the envelops of AGB stars and from supernovae, absorbs light emitted by stars particularly at short wavelengths, is heated by this light, and re-emits it at longer wavelengths (infrared and sub-mm). It is clear that in order to accurately model the dust extinction and emission, one needs to know how dust grains and stars are distributed and which is the composition of the dust grains.

For a population of galaxies that are assumed to have the same composition and distribution of dust, one can derive an "effective" extinction law that can then be used without modeling in detail the dust distribution. For example, one can assume that an "obscuting screen" or a "slab" geometry of dust is sitting between the galaxy and the observer. A simple estimate of the amount of extinction can then be obtained by adopting the measured effective law (e.g., the law found for local starburst galaxies by Calzetti et al. 1994) and by scaling the depth at optical wavelengths on the basis of the physical properties of the galaxy under consideration (e.g., gas content and metallicity). Alternatively, the propagation of light through the interstellar medium can be studied using radiative transfer calculations which take into account the geometry of the galaxy, as well as the distribution and mix of dust (Silva et al. 1998; Jonsson 2006). Recent work by Fontanot et al. (2009b) has compared the two approaches and has shown that the former can predict quite accurately results from the full radiative transfer calculation, with a small scatter. However, there is a large galaxy-to-galaxy variation, likely due to different geometries, that the simple approach cannot capture. It should be noted that also radiative transfer codes often need to make a number of assumptions about the physical state of the dust and about its distribution relative to stars and the interstellar medium. Finally, given the uncertainties involved, often these properties are assumed not to vary as a function of cosmic time.

4 Putting It Together: Models of Galaxy Formation in a Cosmological Context

As discussed above, the process of galaxy formation involves complex and nonlinear physical processes that cover many orders of magnitude in physical size (from the scale of black holes to that of massive galaxy clusters) and in time scales. In addition, as we have seen in the previous section, many (if not all) of the physical processes at play cannot be treated from first principles.

In the past decades, however, three major approaches have been used and further developed in order to circumvent these difficulties and improve our understanding of the physical processes that drive galaxy formation and evolution. The following provides a brief review of these techniques and discusses the most recent successes and problems of one particular class of models that is widely used to make detailed predictions of galaxy properties at different cosmic epochs and environments.

This section will not provide a detailed description of the implementations used in different models. Given the complexity involved, such a description would rapidly become out of date as models are continuously being improved and developed. Rather, this section is aimed at discussing the weaknesses and the strengths of each of the methods that can be used to model galaxy formation in a cosmological context.

4.1 The Halo Occupation Distribution Method

This method essentially bypasses any explicit modeling of the physical processes driving galaxy formation and evolution and specifies the link between dark matter haloes and galaxies in a purely statistical fashion. The halo occupation framework has a long history: a first description can be found in Neyman and Scott (1952) who discussed an analytic model that described galaxy clustering as the superposition of randomly distributed clusters with given profiles. The method has become very popular in more recent years, after it was realized that it provides a powerful formalism for understanding the clustering of galaxies (e.g., Benson et al. 2000; Berlind and Weinberg 2002, and references therein). A classical halo occupation distribution (HOD) model can be constructed by specifying the probability that a halo of mass M contains N galaxies of a particular class (the halo occupation distribution – $P(N|M)$) and by assuming a spatial distribution of galaxies inside dark matter haloes (the most common assumption is that the distribution of galaxies follows that of the dark matter). The halo occupation distribution can then be constrained using galaxy clustering data. For example, a simple model that is often employed in the literature assumes that the mean number of galaxies above a certain luminosity threshold changes with halo mass as

$$N_{\rm avg} = \begin{cases} 0 & \text{if } M < M_{\rm min} \\ (M/M_1)^{\alpha} & \text{otherwise,} \end{cases} \tag{10.6}$$

where $M_{\rm min}$ is a cutoff halo mass below which haloes cannot contain galaxies. In this model, M_1 corresponds to the mass of haloes that contain, on average, one galaxy. ❯ *Figure 10-18* shows the influence of $M_{\rm min}$ (left panel) and α (right panel) on the galaxy correlation function and demonstrates that these observational data can be used to constrain the HOD parameters.

The same approach can be extended to constrain the halo occupation as a function of some galaxy physical property (e.g., luminosity, color, type, etc.). For example, one can define a "conditional luminosity function" $\Phi(L|M)dL$ that specifies the average number of galaxies with luminosities in the range $L \pm dL/2$ that reside in a halo of mass M. This provides a direct link between the observed galaxy luminosity function and the halo mass function:

$$\Phi(L) = \int_0^{\infty} \Phi(L|M) n(M) dM$$

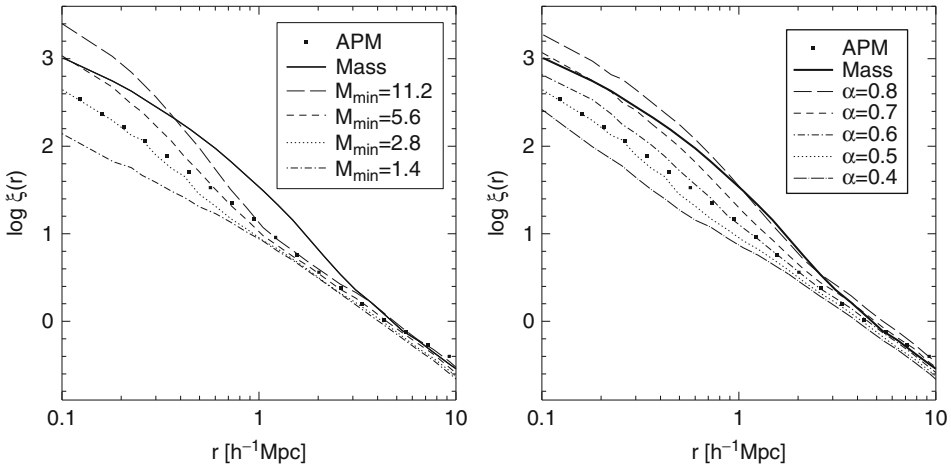

Fig. 10-18

From Berlind and Weinberg (2002). Influence of M_{min} and α on the predicted galaxy correlation function. *Curves* show galaxy correlation functions for HOD models constructed assuming the distribution described in (10.6) with different values of M_{min} and α as indicated in the legend. *Data points* show the correlation function measured from the APM galaxy survey (Baugh 1996)

In addition, one can express the total luminosity of a halo of a given mass as a function of the conditional luminosity function:

$$< L(M) > = \int_0^\infty \Phi(L|M) L dL$$

It has been shown that by adopting this formalism, it is possible to constrain both galaxy formation and cosmology by using the following observational data: the observed luminosity function, the luminosity dependence of the galaxy-galaxy two-point correlation function, and the average mass-to-light ratios as function of halo mass (van den Bosch et al. 2003).

The method described above is conceptually simple and relatively easy to implement. As shown, it can be constrained using the increasing amount of available information on the clustering properties of galaxies at different cosmic epochs, and it provides important statistical constraints for galaxy formation models. It remains difficult, however, to move from this purely statistical characterization of the link between dark matter haloes and galaxies to a more physical understanding of the galaxy formation process itself. In addition, the method described above implicitly assumes that the number of galaxies of a given type populating a dark matter halo, as well as the clustering properties of dark matter haloes, depend only on the halo mass. That is, the method neglects the assembly bias that has been discussed in Sect. 2.4. Recent studies show that this might not be a significant problem, at least for relatively bright galaxies (Tinker et al. 2008, and references therein). Further investigations are, however, needed, particularly in light of the statistical power and redshift coverage of forthcoming observational surveys.

A variant of the HOD approach is provided by the subhalo abundance matching (SHAM) technique. The method consists in assigning observable galaxy properties to the subhalo population of an N-body simulation, assuming a monotonic relation between these properties and

some property of the substructure (e.g., the mass or the maximum circular velocity of dark matter halos) at the time of "accretion," i.e., when the halo was accreted onto a larger structure becoming a subhalo (Conroy et al. 2006; Wang et al. 2006). This method offers some advantages with respect to the simple HOD approach described above. For example, it explicitly accounts for the dependence of the halo history on the environment. It could, however, depending on the resolution of the simulation, miss a significant fraction of the galaxy population: the "orphan" galaxies (i.e., those whose parent substructures were destroyed below the resolution by tidal stripping).

4.2 Hydrodynamical Simulations

Two different approaches can be used to include gas physics in N-body simulations. The most straightforward technique is adopted in smoothed-particle hydrodynamics (SPH) codes. These are based on a Lagrangian method which essentially works dividing the fluid into a set of discrete elements (particles). These have a spatial distance ("smoothing length"), over which their properties are smoothed by a kernel function. Any physical quantity of a particle (for example, density, temperature, and chemical composition) can then be obtained by summing the relevant properties of all the particles which lie within the range of the kernel. The contributions of each particle to a physical property are weighted according to their distance from the particle of interest and their density. Because of the smoothing, SPH codes have problems in resolving and treating dynamical instabilities developing at sharp interfaces in a multiphase fluid (e.g., shocks, Agertz et al. 2007). The Lagrangian nature of the method, however, means that regions of high density are automatically better resolved than regions of low density so that it is possible to study many orders of magnitude in the fluid properties. An alternative scheme is adopted in Eulerian codes in which the fluid equations are solved on a grid which is fixed in time and that can be "refined" several times to increase the resolution in regions of interest. This method is thus well adapted for capturing shocks and discontinuities. The resolution of the simulation can be increased by using "adaptive mesh refinements," but it can become quite time consuming.

As a tool for studying galaxy formation and evolution, hydrodynamical simulations offer the great advantage of providing an explicit description of the gas dynamics. They are, however, quite demanding in terms of computational time and memory consumption so that it is often necessary to limit the resolution range and the size of the volume being simulated. Additionally, and perhaps more importantly, complex physical processes such as star formation, feedback, etc., still have to be included as "subgrid physics." This is the case either because the resolution of the simulation is inadequate to treat a specific problem or simply because (and this is true almost always) we do not have a complete theory for the physical problem under consideration.

Current state-of-the-art full hydrodynamic cosmological simulations include the Galaxies-Intergalactic Medium Interaction Calculation project (GIMIC, Crain et al. see 2009) and the OverWhelmingly Large Simulations project (OWLS, Schaye et al. 2010). In the GIMIC project, five different regions with different mean overdensities have been selected from the Millennium simulation. These have been resimulated at high resolution using a SPH code that takes into account metal-dependent cooling in the presence of an ionizing UV background and includes a model for star formation and supernova feedback. Since each of the simulations is a considerable investment of computational time, they have been run using a unique set of parameters and concentrating on the environmental effects of the physical processes considered. The OWLS project represents a complimentary approach as it is based on a suite of over 50 cosmological

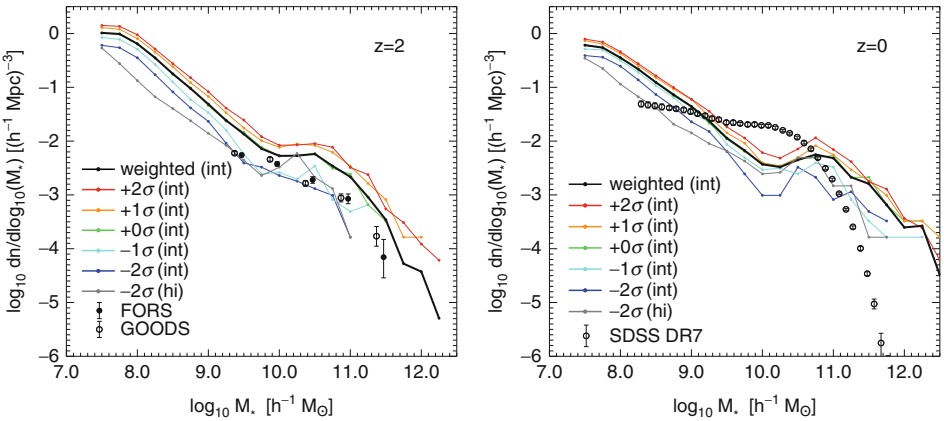

Fig. 10-19

From Crain et al. (2009). The stellar mass function of galaxies at $z = 2$ (*left panel*) and $z = 0$ (*right panel*). Results are shown for all five intermediate-resolution simulations considered in the GIMIC project (*colored curves*) and their weighted average (*black curve*). The stellar mass function of the region with largest overdensity at high resolution is also shown (*gray curve*) to illustrate the degree of convergence. Symbols with error bars show observational measurements from Drory et al. (2005) at $z = 2$ and from Li and White (2009) at $z = 0$

simulations (typically much smaller than the GIMIC high-resolution regions) that investigate the effects of different implementations of subgrid physics.

Figure 10-19 shows how the stellar mass function resulting from the GIMIC simulations compares with observational measurements at $z \sim 2$ in the left panel and at $z = 0$ in the right panel. Colored lines show results from each simulation while the black lines show their weighted average. The figure shows that the shape of the predicted stellar mass function differs significantly from that measured: the simulations predict an excess of low- and high-mass galaxies with respect to the observations and a "dip" in correspondence of the "knee" of the observed stellar mass function (that is where most of the galaxy mass is). At higher redshift, where the observations span only a limited mass range, the agreement looks better, but the shape of the predicted galaxy mass function does not vary with respect to the $z = 0$ predictions. The excess at large masses in the overdense regions originates mainly from the fact that these simulations do not include any modeling of the heating processes that can quench cooling flows in clusters (see also next sections). At low and intermediate masses, the disagreement with observational data is likely due to the simple wind model that has been adopted (for details, see Crain et al. 2009). A feedback model that follows the scalings expected for momentum-driven winds can give a better match with observational data around the knee of the luminosity function but still fails at higher and lower masses (Davé et al. 2011).

Much work has been done using direct simulations of the baryonic physics to study the formation and evolution of individual haloes at high resolution. These simulations take advantage of the zoom technique: first, a cosmological simulation of a large region is used to select a suitable target halo. The particles in the selected halo and its surroundings (usually all the particles within two times the virial radius) are then traced back to their initial Lagrangian region and are replaced by a larger number of lower mass particles. These are perturbed using the

same fluctuation distribution as in the parent simulation but now extended to smaller scales to account for the increase in resolution. This resampling of the initial conditions thus allows a localized increase in mass and force resolution. Outside the high-resolution region, particles of variable mass (increasing with distance) are used, so that the computational effort is concentrated on the region of interest while still maintaining a faithful representation of the large-scale density and velocity field of the parent simulation.

On the cluster scale, significant disagreements with the observational data are still found in terms of the statistical description of the cluster galaxy population. For example, Saro et al. (2006) analyze a set of 19 cluster resimulations carried out using a SPH code that includes gas cooling, star formation, a detailed treatment of stellar evolution and chemical enrichment, as well as supernova feedback. They find that the total number of galaxies in their simulated clusters falls short of the observational measurements by a factor 2–3. The problem does not have an obvious numerical origin (e.g., lack of mass and force resolution). In addition, the BCGs of the simulated clusters are always predicted to be too massive and too blue when compared to data, stressing the need for the inclusion of a physical process that suppresses gas condensation at the center of relatively massive haloes.

At galaxy scales, simulations have generally had problems reproducing disk-dominated galaxies in typical dark matter haloes, when taking into account the cosmological setting. One major problem is known as the "angular momentum catastrophe": baryons condense early in clumps that then fall into larger haloes and merge via dynamical friction. This produces a net and significant transfer of angular momentum from the baryons to the dark matter. As a result, simulated disks are generally too small with up to ten times less angular momentum than real disk galaxies. The formation of a realistic rotationally supported disk galaxy in a fully cosmological simulation is still an open problem. Recent numerical work shows that it is in part due to limited resolution and related numerical effects that cause artificial angular momentum loss and spurious bulge formation (for a detailed discussion, see Mayer et al. 2008). The physics of galaxy formation during the merger of the most massive protogalactic lumps at high redshift and, in particular, the feedback due to supernovae is, howevelr, also playing a very important role (e.g., Scannapieco et al. 2008, and references therein).

4.3 Semianalytic Models of Galaxy Formation

The backbone of any semianalytic model is a statistical representation of the growth of dark matter haloes, i.e., a merger tree. Once the backbone of the model is constructed, galaxy formation and evolution can be coupled to the merger trees using a set of analytic laws that are based on theoretical and/or observational arguments to describe complex physical processes like star formation, supernovae and AGN feedback processes, etc. Adopting this formalism, it is possible to express the full galaxy formation process through a set of differential equations that describe the variation in mass of the different galactic components (e.g., gas, stars, metals). Given our limited understanding of the physical processes at play, these equations contain "free parameters," whose values are typically chosen in order to provide a reasonably good agreement with observational data in the local universe. These techniques find their seeds in the pioneering work by White and Rees (1978), have been laid out in a more detailed form in the early 1990s (White and Frenk 1991; Cole 1991), and have been substantially extended and refined in the last years by a number of different groups. For a detailed review of these techniques, the interested reader is referred to Baugh (2006).

In their first renditions, semianalytic models relied on Monte Carlo realizations of merging histories of individual objects, generated using the extended Press-Schechter theory (e.g., Kauffmann et al. 1993). An important advance of later years came from the coupling of semianalytic techniques with large-resolution N-body simulations that are used to specify the location and evolution of dark matter haloes – the birthplaces of luminous galaxies (Kauffmann et al. 1999; Benson et al. 2000). On a next level of complexity, some more recent implementations of these techniques have explicitly taken into account dark matter substructures, i.e., the haloes within which galaxies form are still followed when they are accreted onto a more massive system (Springel et al. 2001; De Lucia et al. 2004b). There is one important caveat to bear in mind regarding these methods: dark matter substructures are fragile systems that are rapidly and efficiently destroyed below the resolution limit of the simulation (see Sect. 2.4). Depending on the resolution of the simulations used, this can happen well before the actual merger can take place. This treatment introduces a complication due to the presence of "orphan galaxies," i.e., galaxies whose parent substructure mass has been reduced below the resolution limit of the simulation. In most of the available semianalytic models, these galaxies are assumed to merge onto the corresponding central galaxies after a residual merging time which is given by some variation of the classical dynamical friction formula.

One great advantage of these hybrid methods, with respect to classical techniques based on the extended Press-Schechter formalism, is that they provide full dynamical information about model galaxies. Using realistic mock catalogs generated with these methods, accurate and straightforward comparisons with observational data can be carried out. Since N-body simulations can handle large numbers of particles, the hybrid approach can access a very large dynamic range of mass and spatial resolution, at small computational costs. In addition, since the computational times are limited, these methods also allow a fast exploration of the parameter space and an efficient investigation of the influence of specific physical assumptions. This comes at the expenses, however, of loosing an explicit description of the gas dynamics.

One common criticism to semianalytic models is that there are "too many" free parameters. It should be noted, however, that the number of these parameters is not larger than the number of published comparisons with different and independent sets of observational data, for any of the semianalytic models discussed in the recent literature. In addition, these are not "statistical" parameters but, as explained above, they are due to our lack of understanding of the physical processes considered. Therefore, a change in any of these parameters has consequences on a number of different predictable properties so that often there is little parameter degeneracy for a given set of prescriptions. Finally, observations and theoretical arguments often provide important constraints on the range of values that different parameters can assume.

4.4 Successes and Problems of Semianalytic Models of Galaxy Formation

Clearly, each of the methods described above has its own advantages and weaknesses, and they should be viewed as complementary rather than competitive. In the framework of galaxy formation, semianalytic models certainly represent the most developed theoretical tool for interpreting observations of galaxy formation and evolution. This section provides a brief discussion of some of the most recent successes and problems of current models.

It is interesting to start this discussion from what can be considered the most fundamental description of the galaxy population: the galaxy luminosity function. As mentioned above, since

early implementations of semianalytic techniques, it was clear that a relatively strong supernovae feedback was needed in order to suppress the large excess of faint galaxies due to the steep increase of low-mass dark matter haloes (White and Frenk 1991; Benson et al. 2003). The left panel of *Fig. 10-20* shows results from different models: the simplest one is obtained converting the dark matter halo mass function into a galaxy luminosity function by assuming a fixed mass-to-light ratio (this is the same model shown in *Fig. 10-9*). As discussed in Sect. 3.4, this model overpredicts the faint and the bright ends of the luminosity function by orders of magnitude. The other lines shown in the left panel of *Fig. 10-20* correspond to different models where different ingredients have been switched on, as indicated in the legend. None of these models reproduces the observational measurements. The right panel of the figure shows how the predicted K-band luminosity function compares with observational measurements, for increasing efficiency of supernovae feedback. Adopting a relatively strong feedback (see also Guo et al. 2011), the agreement with the observational data becomes satisfactory at the faint end. It is interesting to note, however, that matching the faint end of the luminosity function comes at the expenses of exacerbating the excess of luminous bright objects. This is due to the fact that the material reheated and/or ejected by low-mass galaxies in this model (but this is generally true for most of the models that can be found in the literature) ends up in the hot gas that is associated with the corresponding central galaxies. At later times, this material cools efficiently onto the corresponding central galaxies increasing their luminosities and star formation rates, at odds with observational data.

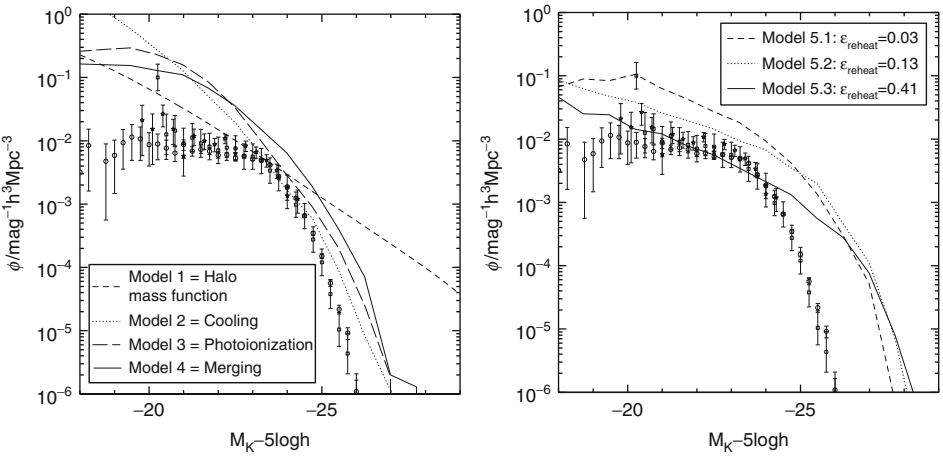

Fig. 10-20

From Benson et al. (2003). *Points* show observational determinations of the observed K-band luminosity function. *Left panel*: lines show results from different models: model 1 is obtained converting the dark matter halo mass function into a galaxy luminosity function by assuming a fixed mass-to-light ratio chosen to match the knee of the luminosity function; model 2 shows results from a model with no feedback from supernovae, no photoionization suppression, and no mergers; model 3 is similar to model 2 but includes photoionization; model 4 includes also galaxy mergers. *Right panel*: lines show predictions from a model with increasing efficiency of supernovae feedback, as indicated in the legend

Matching the bright end of the luminosity function has proved much more difficult than matching the faint end, and a reasonable success has been achieved only recently by means of a relatively strong form of "radio-mode" AGN feedback (see Sect. 3.4.3). Different prescriptions of AGN feedback have been proposed in recently published models (e.g., Croton et al. 2006; Bower et al. 2006; Monaco et al. 2007), and still much work remains to be done in order to understand if and how the energy injected by intermittent radio activity at the cluster center is able to efficiently suppress the cooling flows. In addition, recent work has pointed out that most models assume a strong dependence of radio-mode feedback on the parent halo mass. As a consequence of this assumption, these models predict that essentially all massive galaxies should be associated with a bright radio source, while observational data suggest that faint and bright radio sources are found in similar environments in equal numbers (Fontanot et al. 2011). Finally, it should be noted that although AGN feedback has received much attention in recent years, the necessity of introducing a physical process to suppress the condensation of gas at the center of massive haloes and the hypothesis that this might be due to feedback from AGN was noted in earlier work (e.g., Kauffmann et al. 1999).

The main reason for the success of the "radio-mode" AGN feedback is that it does not require star formation to be effective. As a consequence, this mode of feedback permits to suppress the luminosity of massive galaxies and, at the same time, to keep their stellar populations old, in qualitative agreement with observational data (see, e.g., De Lucia et al. 2006). The models also seem to reproduce, at least qualitatively, the observed trend for more massive ellipticals to have shorter star formation time scales. A good quantitative agreement has not been shown yet, and a detailed comparison between models and observations is complicated by large uncertainties associated to star formation histories extracted from observed galaxy spectra (see, e.g., Fontanot et al. 2009a).

In these models, ellipticals and bulges form mainly through mergers (for a detailed analysis of the contribution of different channels, see De Lucia et al. 2011). Naively, one expects very large numbers of mergers in the hierarchical scenario, where more massive systems form through the mergers of smaller units, and larger systems are expected to be made up by a larger number of progenitors. It is therefore interesting to ask how the number of progenitors varies as a function of galaxy mass. The left panel of *Fig. 10-21* shows the "effective number of stellar progenitors" of elliptical galaxies of different mass. This quantity represents a mass-weighted counting of the stellar systems that make up the final galaxy, and therefore provides a good proxy for the number of significant mergers required to assemble a galaxy of given mass. The figure shows results from a model where only mergers contribute to the formation of bulges (empty circles) and those from a model where bulges can also form through disk instability (filled symbols). The vertical dashed line indicates the threshold above which the morphology classification can be considered robust (the limit is set by the resolution of the parent N-body simulation). As expected, more massive galaxies are made up of more pieces. The number of effective progenitors is, however, less than two up to stellar masses of $\simeq 10^{11}\,M_{\odot}$, indicating that the formation of these systems typically involves only a small number of major mergers. Only more massive galaxies are built through a larger number of mergers, reaching up to $\simeq 5$ for the most massive systems. The right panel of *Fig. 10-21* shows the distribution of "formation" (top panel) and "assembly" redshifts (bottom panel) of model ellipticals. The formation redshift is defined here as the redshift when 50% of the stars that end up in ellipticals today are already formed, while the assembly redshift is defined as the redshift when 50% of the stars that end up in ellipticals today are already assembled in a single object. The right panel of *Fig. 10-21* shows that more massive galaxies are "older," albeit with a large scatter, but

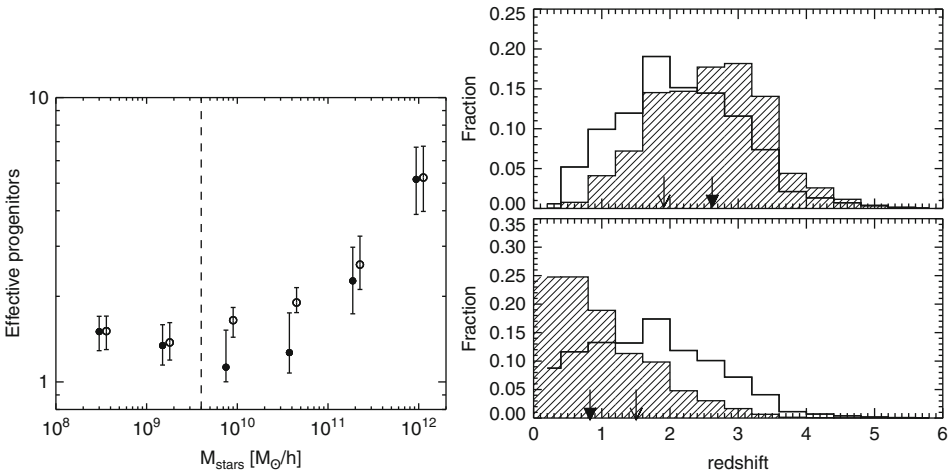

Fig. 10-21

From De Lucia et al. (2006). *Right panel*: distribution of formation (*top panel*) and assembly redshifts (*bottom panel*). The *shaded histogram* is for elliptical galaxies with stellar mass larger than 10^{11} $M_\odot$, while the open histogram is for all ellipticals with mass larger than 4×10^9 $M_\odot$. *Arrows* indicate the medians of the distributions, with the *thick arrows* referring to the *shaded histograms*. *Left panel*: effective number of progenitors as a function of galaxy stellar mass for model elliptical galaxies. *Symbols* show the median of the distribution, while error bars indicate the *upper* and *lower* quartiles. *Filled* and *empty symbols* refer to a model with and without a disk instability channel for the formation of the bulge

assemble "later" than their lower-mass counterparts. The assembly history of ellipticals hence parallels the hierarchical growth of dark matter haloes, in contrast to the formation history of the stars themselves. Data shown in the right panel of *Fig. 10-21* imply that a significant fraction of present elliptical galaxies have assembled relatively recently, through purely stellar mergers. This finding appears to be supported by recent observational results (e.g., van Dokkum 2005).

Models predict an increase in stellar mass by a factor 2–4 since $z \sim 1$, depending on stellar mass (De Lucia et al. 2006; De Lucia and Blaizot 2007). This creates a certain tension with the observation that the massive end of the galaxy mass function does not appear to evolve significantly over the same redshift interval. A large part of this tension is removed when taking into account observational errors and uncertainties on galaxy mass estimates (see Fontanot et al. 2009a, and references therein). For the mass assembly of the BCGs, the situation is worse: while observations seem consistent with no mass growth since $z \sim 1$, models predict an increase in mass by a factor about 4 (De Lucia and Blaizot 2007; Whiley et al. 2008). One major caveat in this comparison, however, is given by the fact that observational studies usually adopt fixed metric aperture magnitudes (which account for about 25–50% of the total light contained in the BCG and intracluster light), while models compute total magnitudes. Semianalytic models do not provide information regarding the spatial distribution of the BCG light, so aperture magnitudes cannot be calculated. In addition, most of the available models do not take into

account the stripping of stars from other cluster galaxies due to tidal and harassment effects (Monaco et al. 2006; Conroy et al. 2007).

Most of the models currently available exhibit a remarkable degree of agreement with a large number of observations for the galaxy population in the local universe (e.g., the observed relations between stellar mass, gas mass, and metallicity; the observed luminosity, color, and morphology distribution; the observed two-point correlation functions). When analyzed in detail, however, some of these comparisons show important and systematic (i.e., common to most of the semianalytic models discussed in the literature) disagreements. A few of the problems on which the community is focusing in current years are discussed in the following.

Although models are not usually tuned to match observations of galaxy clustering, they generally reproduce the observed dependence of clustering on magnitude and color. The agreement appears particularly good for the dependence on luminosity, while the amplitude difference on color appears greater in the models than observed (Springel et al. 2005b). This problem might be (at least in part) related to the excess of small red satellite galaxies which plagues all models discussed in the recent literature (e.g., see Fig. 11 in Croton et al. 2006 and discussion in Fontanot et al. 2009a). At low redshift, this excess is largely due to satellite galaxies that were formed and accreted early on and that are dominated by old stellar populations. As explained in Sect. 3.7.4, semianalytic models assume that when a galaxy is accreted onto a larger structure, the gas supply can no longer be replenished by cooling that is suppressed by an instantaneous and complete stripping of the hot-gas reservoir. Since this process is usually combined with a relatively efficient supernovae feedback, galaxies that are accreted onto a larger system consume their gas very rapidly, moving onto the red sequence on quite short time scales (Weinmann et al. 2006; Wang et al. 2007). This contributes to produce an excess of faint and red satellites and a transition region (sometimes referred to as "green valley") which does not appear to be as well populated as observed. Much effort has been recently devoted to this problem, and many models have implemented a noninstantaneous stripping of the hot halo around satellites (e.g., Font et al. 2008; Weinmann et al. 2010; Guo et al. 2011). With these modifications, a larger fraction of the model satellites have bluer colors, resulting in a color distribution that is in better (but not perfect) agreement with the observational data. These models, however, still appear to overestimate the number of low- and intermediate-mass galaxies at higher redshift and the clustering signal on small scales (see, e.g., Figs. 20 and 23 in Guo et al. 2011).

The completion of new high-redshift surveys has recently pushed comparisons between model results and observational data to higher redshift (Stringer et al. 2009; de la Torre et al. 2011). This currently still rather unexplored regime for models of galaxy formation is very interesting because it is at high redshift that predictions from different models differ more dramatically.

To close this section, it is worth reminding that a long-standing problem for hierarchical models has been to match the zero point of the Tully-Fisher relation (the observed correlation between the rotation speed and the luminosity of spiral disks, Tully and Fisher 1977) while reproducing, at the same time, the observed luminosity function. As discussed in Baugh (2006), no model with a realistic calculation of galaxy size has been able to match the zero point of the Tully-Fisher relation using the circular velocity of the disk measured at the half-mass radius. It remains unclear if this difficulty is related to some approximation in the size calculation, or if it is related to more fundamental shortcomings of the cold dark matter model.

5 Concluding Remarks

This chapter has assumed that the cosmogony of our universe is well described by the ΛCDM paradigm. As a matter of fact, the CDM paradigm is not without its problems. The most discussed ones are related to the central mass distribution of low-surface brightness galaxies, to the existing substructure in galaxy-size haloes, and to the angular momentum of the galaxy disks (e.g., Tasitsiomi 2003; Benson 2005, and references therein). If there is a CDM "crisis" then, it is on the galactic and subgalactic scale, where the influence of the baryonic component is expected to play an important role. Indeed, many (all?) of the problems listed above might have an astrophysical solution, so rather than problems of the CDM scenario they might be problems with the way we model the evolution of the baryons in the cosmological context. It is clear then that in order to really test CDM, we need to improve our galaxy formation models so as to make more accurate predictions on small scales.

As discussed in Sect. 3 of this chapter, galaxy formation is a very difficult physical problem as one should account for a variety of phenomena that act on different scales and at different times and that interact in many possible ways. In addition, both theoretically and observationally, we have a very limited understanding of most of the physical processes that should be taken into account. Given the complexity of the problem, it is clear that we are not yet (and perhaps we will never be) in the position of being able to model galaxy formation "from first principles." We can, however, use a number of different techniques that can help us improving our understanding of the physical processes at play.

The theoretical tools that can be employed to study galaxy formation and evolution are many and complementary. None of them will ever provide "the model" that reproduces the observed universe. Indeed, it would be perhaps naive to believe that it is possible to summarize all the complexity that we observe in a set of analytic or semianalytic or seminumerical equations. The way to proceed then is to take advantage of the complementarity between different approaches and use the observational data to falsify the hypotheses that have been made. In going into this loop, one has to remember that the models generally include a number of free parameters. However, more than to the exact value of the parameters, attention should be given to the "parameterizations," i.e., specific assumptions on the physical processes considered. It is clear that the larger the number of processes considered is, the larger will the number of parameters/parametrizations be. There will be some degeneracy that can, however, be limited by considering a larger set of observational constraints.

These are exciting times to study galaxy formation. More and better data are becoming available. Theoretical models that try to reproduce the ever more detailed observational picture of the universe, will also require ever more complex modeling. Only by keeping the close link between theoretical predictions and observational data discussed, will it be possible to shed light on the physical processes governing galaxy formation and evolution.

Acknowledgments

The author acknowledges financial support from the European Research Council under the European Community's Seventh Framework Programme (FP7/2007-2013)/ERC grant agreement n. 202781.

This work is dedicated to Luigi Cuomo.

Cross-References

- Active Galactic Nuclei
- Cosmology
- Galaxy Interactions
- Large Scale Structure of the Universe
- Star Formation
- The Influence of Environment on Galaxy Evolution

References

Abel, T., Bryan, G. L., & Norman, M. L. 2002, Science, 295, 93

Agertz, O., Moore, B., Stadel, J., Potter, D., Miniati, F., Read, J., Mayer, L., Gawryszczak, A., Kravtsov, A., Nordlund, Å., Pearce, F., Quilis, V., Rudd, D., Springel, V., Stone, J., Tasker, E., Teyssier, R., Wadsley, J., & Walder, R. 2007, MNRAS, 380, 963

Arrigoni, M., Trager, S. C., Somerville, R. S., & Gibson, B. K. 2010, MNRAS, 402, 173

Bastian, N., Covey, K. R., & Meyer, M. R. 2010, ARA&A, 48, 339

Baugh, C. M. 1996, MNRAS, 280, 267

Baugh, C. M. 2006, Rep. Prog. Phys., 69, 3101

Begelman, M. C. 2004, Coevolution of Black Holes and Galaxies (Cambridge University Press), 374

Benson, A. J. 2005, Small Scales, Big Issues for Cold Dark Matter (London: Imperial College Press), 59–75

Benson, A. J., & Bower, R. 2011, MNRAS, 410, 2653

Benson, A. J., Cole, S., Frenk, C. S., Baugh, C. M., & Lacey, C. G. 2000, MNRAS, 311, 793

Benson, A. J., Pearce, F. R., Frenk, C. S., Baugh, C. M., & Jenkins, A. 2001, MNRAS, 320, 261

Benson, A. J., Lacey, C. G., Baugh, C. M., Cole, S., & Frenk, C. S. 2002, MNRAS, 333, 156

Benson, A. J., Bower, R. G., Frenk, C. S., Lacey, C. G., Baugh, C. M., & Cole, S. 2003, ApJ, 599, 38

Berlind, A. A., & Weinberg, D. H. 2002, ApJ, 575, 587

Bertschinger, E. 1985, ApJS, 58, 39

Best, P. N., von der Linden, A., Kauffmann, G., Heckman, T. M., & Kaiser, C. R. 2007, MNRAS, 379, 894

Bett, P., Eke, V., Frenk, C. S., Jenkins, A., Helly, J., & Navarro, J. 2007, MNRAS, 376, 215

Bigiel, F., Leroy, A., Walter, F., Brinks, E., de Blok, W. J. G., Madore, B., & Thornley, M. D. 2008, AJ, 136, 2846

Bigiel, F., Leroy, A., & Walter, F. 2010, Computational Star Formation, Proceedings of the International Astronomical Union, IAU Symposium, 270, 327–334

Binney, J. 1977, ApJ, 215, 483

Binney, J., & Merrifield, M. 1998, Galactic Astronomy (Princeton: Princeton University Press)

Birnboim, Y., & Dekel, A. 2003, MNRAS, 345, 349

Blumenthal, G. R., Faber, S. M., Primack, J. R., & Rees, M. J. 1984, Nature, 311, 517

Bond, J. R., Cole, S., Efstathiou, G., & Kaiser, N. 1991, ApJ, 379, 440

Bower, R. G., Benson, A. J., Malbon, R., Helly, J. C., Frenk, C. S., Baugh, C. M., Cole, S., & Lacey, C. G. 2006, MNRAS, 370, 645

Boylan-Kolchin, M., Ma, C.-P., & Quataert, E. 2008, MNRAS, 383, 93

Brüggen, M., & De Lucia, G. 2008, MNRAS, 383, 1336

Bruzual, G., & Charlot, S. 2003, MNRAS, 344, 1000

Calzetti, D., Kinney, A. L., & Storchi-Bergmann, T. 1994, ApJ, 429, 582

Chabrier, G. 2003, PASP, 115, 763

Chandrasekhar, S. 1943, ApJ, 97, 255

Ciardi, B., & Ferrara, A. 2005, Space Sci. Rev., 116, 625

Clark, P. C., Glover, S. C. O., Klessen, R. S., & Bromm, V. 2011, ApJ, 727, 110

Cole, S. 1991, ApJ, 367, 45

Conroy, C., Wechsler, R. H., & Kravtsov, A. V. 2006, ApJ, 647, 201

Conroy, C., Wechsler, R. H., & Kravtsov, A. V. 2007, ApJ, 668, 826

Conroy, C., Gunn, J. E., & White, M. 2009, ApJ, 699, 486

Cowie, L. L., & Songaila, A. 1977, Nature, 266, 501

Cox, T. J., Jonsson, P., Somerville, R. S., Primack, J. R., & Dekel, A. 2008, MNRAS, 384, 386

Crain, R. A., Theuns, T., Dalla Vecchia, C., Eke, V. R., Frenk, C. S., Jenkins, A., Kay, S. T., Peacock, J. A., Pearce, F. R., Schaye, J., Springel, V., Thomas, P. A., White, S. D. M., & Wiersma, R. P. C. 2009, MNRAS, 399, 1773

Croton, D. J., Springel, V., White, S. D. M., De Lucia, G., Frenk, C. S., Gao, L., Jenkins, A., Kauffmann, G., Navarro, J. F., & Yoshida, N. 2006, MNRAS, 365, 11

Croton, D. J., Gao, L., & White, S. D. M. 2007, MNRAS, 374, 1303

Davé, R., Oppenheimer, B. D., & Finlator, K. 2011, MNRAS, 415, 11

Davis, M., Efstathiou, G., Frenk, C. S., & White, S. D. M. 1985, ApJ, 292, 371

de la Torre, S., Meneux, B., De Lucia, G., Blaizot, J., Le Fèvre, O., Garilli, B., Cucciati, O., Mellier, Y., et al. 2011, A&A, 525, A125+

De Lucia, G., & Blaizot, J. 2007, MNRAS, 375, 2

De Lucia, G., Kauffmann, G., Springel, V., White, S. D. M., Lanzoni, B., Stoehr, F., Tormen, G., & Yoshida, N. 2004a, MNRAS, 348, 333

De Lucia, G., Kauffmann, G., & White, S. D. M. 2004b, MNRAS, 349, 1101

De Lucia, G., Springel, V., White, S. D. M., Croton, D., & Kauffmann, G. 2006, MNRAS, 366, 499

De Lucia, G., Boylan-Kolchin, M., Benson, A. J., Fontanot, F., & Monaco, P. 2010, MNRAS, 406, 1533

De Lucia, G., Fontanot, F., Wilman, D., & Monaco, P. 2011, MNRAS, 517–+

Doroshkevich, A. G., Zel'Dovich, Y. B., & Novikov, I. D. 1967, Sov. Astron., 11, 233

Dressler, A. 1980, ApJ, 236, 351

Drory, N., Salvato, M., Gabasch, A., Bender, R., Hopp, U., Feulner, G., & Pannella, M. 2005, ApJ, 619, L131

Efstathiou, G. 1992, MNRAS, 256, 43P

Eggen, O. J., Lynden-Bell, D., & Sandage, A. R. 1962, ApJ, 136, 748

Einasto, J. 1965, Trudy Inst. Astrofiz. Alma-Ata, 51, 87

Einasto, J., Saar, E., Kaasik, A., & Chernin, A. D. 1974, Nature, 252, 111

Evrard, A. E. 1990, ApJ, 363, 349

Fabian, A. C., Sanders, J. S., Ettori, S., Taylor, G. B., Allen, S. W., Crawford, C. S., Iwasawa, K., Johnstone, R. M., & Ogle, P. M. 2000, MNRAS, 318, L65

Farouki, R., & Shapiro, S. L. 1981, ApJ, 243, 32

Font, A. S., Bower, R. G., McCarthy, I. G., Benson, A. J., Frenk, C. S., Helly, J. C., Lacey, C. G., Baugh, C. M., & Cole, S. 2008, MNRAS, 389, 1619

Fontanot, F., De Lucia, G., Monaco, P., Somerville, R. S., & Santini, P. 2009a, MNRAS, 397, 1776

Fontanot, F., Somerville, R. S., Silva, L., Monaco, P., & Skibba, R. 2009b, MNRAS, 392, 553

Fontanot, F., Pasquali, A., De Lucia, G., van den Bosch, F. C., Somerville, R. S., & Kang, X. 2011, MNRAS, 413, 957

Forcada-Miro, M. I., & White, S. D. M. 1997, ArXiv Astrophysics e-prints

Gao, L., White, S. D. M., Jenkins, A., Stoehr, F., & Springel, V. 2004, MNRAS, 355, 819

Gao, L., Springel, V., & White, S. D. M. 2005, MNRAS, 363, L66

Gao, L., Frenk, C. S., Boylan-Kolchin, M., Jenkins, A., Springel, V., & White, S. D. M. 2011, MNRAS, 410, 2309

Gnedin, N. Y. 2000, ApJ, 542, 535

Greif, T., Springel, V., White, S., Glover, S., Clark, P., Smith, R., Klessen, R., & Bromm, V. 2011, ApJ, 737, 75

Gunn, J. E., & Gott, J. R., III 1972, ApJ, 176, 1

Guo, Q., White, S., Boylan-Kolchin, M., De Lucia, G., Kauffmann, G., Lemson, G., Li, C., Springel, V., & Weinmann, S. 2011, MNRAS, 413, 101

Guth, A. H. 1981, Phys. Rev. D, 23, 347

Hayashi, E., & White, S. D. M. 2008, MNRAS, 388, 2

Hayashi, E., Navarro, J. F., & Springel, V. 2007, MNRAS, 377, 50

Heckman, T. M. 2002, in ASP Conf. Ser. 254, Extragalactic Gas at Low Redshift, ed. J. S. Mulchaey & J. T. Stocke (San Francisco: ASP). 292–+, Galactic Superwinds Circa 2001

Hubble, E. P. 1936, Realm of the Nebulae (New Haven: Yale University Press)

Jenkins, A., Frenk, C. S., White, S. D. M., Colberg, J. M., Cole, S., Evrard, A. E., Couchman, H. M. P., & Yoshida, N. 2001, MNRAS, 321, 372

Jing, Y. P., & Suto, Y. 2002, ApJ, 574, 538

Jonsson, P. 2006, MNRAS, 372, 2

Katz, N., Hernquist, L., & Weinberg, D. H. 1992, ApJ, 399, L109

Kauffmann, G., White, S. D. M., & Guiderdoni, B. 1993, MNRAS, 264, 201

Kauffmann, G., Colberg, J. M., Diaferio, A., & White, S. D. M. 1999, MNRAS, 303, 188

Kennicutt, R. C., Jr. 1998, ApJ, 498, 541

Kereš, D., Katz, N., Weinberg, D. H., & Davé, R. 2005, MNRAS, 363, 2

Klypin, A., Gottlöber, S., Kravtsov, A. V., & Khokhlov, A. M. 1999, ApJ, 516, 530

Knebe, A., Knollmann, S. R., Muldrew, S. I., Pearce, F. R., Aragon-Calvo, M. A., Ascasibar, Y., Behroozi, P. S., Ceverino, D., et al. 2011, MNRAS, 415, 2293

Komatsu, E., Smith, K. M., Dunkley, J., Bennett, C. L., Gold, B., Hinshaw, G., Jarosik, N., Larson, D., et al. 2011, ApJS, 192, 18

Lacey, C., & Cole, S. 1993, MNRAS, 262, 627

Larson, R. B. 1975, MNRAS, 173, 671

Larson, R. B. 1976, MNRAS, 176, 31

Larson, R. B., Tinsley, B. M., & Caldwell, C. N. 1980, ApJ, 237, 692

Lemson, G., & Kauffmann, G. 1999, MNRAS, 302, 111
Li, C., & White, S. D. M. 2009, MNRAS, 398, 2177
Mac Low, M.-M., & Ferrara, A. 1999, ApJ, 513, 142
Macciò, A. V., Dutton, A. A., van den Bosch, F. C., Moore, B., Potter, D., & Stadel, J. 2007, MNRAS, 378, 55
Maraston, C. 2011, Why Galaxies Care about AGB Stars II: Shining Examples and Common Inhabitants. Astronomical Society of the Pacific, 391
Maraston, C., Strömbäck, G., Thomas, D., Wake, D. A., & Nichol, R. C. 2009, MNRAS, 394, L107
Martin, C. L. 2005, ApJ, 621, 227
Mastropietro, C., Moore, B., Mayer, L., Debattista, V. P., Piffaretti, R., & Stadel, J. 2005, MNRAS, 364, 607
Mayer, L., Governato, F., & Kaufmann, T. 2008, Adv. Sci. Lett., 1, 7
McCarthy, I. G., Frenk, C. S., Font, A. S., Lacey, C. G., Bower, R. G., Mitchell, N. L., Balogh, M. L., & Theuns, T. 2008, MNRAS, 383, 593
McKee, C. F., & Ostriker, E. C. 2007, ARA&A, 45, 565
Mihos, J. C. 2004, Clusters of Galaxies: Probes of Cosmological Structure and Galaxy Evolution (Cambridge/New York: Cambridge University Press), 277
Mo, H., van den Bosch, F. C., & White, S. 2010, Galaxy Formation and Evolution (Cambridge/New York: Cambridge University Press)
Monaco, P., Murante, G., Borgani, S., & Fontanot, F. 2006, ApJ, 652, L89
Monaco, P., Fontanot, F., & Taffoni, G. 2007, MNRAS, 375, 1189
Moore, B., Lake, G., & Katz, N. 1998, ApJ, 495, 139
Murray, N., Quataert, E., & Thompson, T. A. 2005, ApJ, 618, 569
Nagashima, M., Lacey, C. G., Okamoto, T., Baugh, C. M., Frenk, C. S., & Cole, S. 2005, MNRAS, 363, L31
Navarro, J. F., Frenk, C. S., & White, S. D. M. 1997, ApJ, 490, 493
Navarro, J. F., Hayashi, E., Power, C., Jenkins, A. R., Frenk, C. S., White, S. D. M., Springel, V., Stadel, J., & Quinn, T. R. 2004, MNRAS, 349, 1039
Neto, A. F., Gao, L., Bett, P., Cole, S., Navarro, J. F., Frenk, C. S., White, S. D. M., Springel, V., & Jenkins, A. 2007, MNRAS, 381, 1450
Neyman, J., & Scott, E. L. 1952, ApJ, 116, 144
Nulsen, P. E. J. 1982, MNRAS, 198, 1007
Ocvirk, P., Pichon, C., & Teyssier, R. 2008, MNRAS, 390, 1326
Okamoto, T., Gao, L., & Theuns, T. 2008, MNRAS, 390, 920
Ostriker, J. P., & Tremaine, S. D. 1975, ApJ, 202, L113
Ostriker, J. P., Peebles, P. J. E., & Yahil, A. 1974, ApJ, 193, L1
Padmanabhan, T. 1993, Structure Formation in the Universe (Cambridge/New York: Cambridge University Press)
Parkinson, H., Cole, S., & Helly, J. 2008, MNRAS, 383, 557
Peacock, J. A. 1999, Cosmological Physics (Cambridge/New York: Cambridge University Press)
Peebles, P. J. E. 1982, ApJ, 263, L1
Percival, W. J., Scott, D., Peacock, J. A., & Dunlop, J. S. 2003, MNRAS, 338, L31
Press, W. H., & Schechter, P. 1974, ApJ, 187, 425
Quilis, V., Moore, B., & Bower, R. 2000, Science, 288, 1617
Rees, M. J., & Ostriker, J. P. 1977, MNRAS, 179, 541
Richstone, D. O. 1976, ApJ, 204, 642
Roediger, E., & Brüggen, M. 2007, MNRAS, 380, 1399
Rubin, V. C., & Ford, W. K., Jr. 1970, ApJ, 159, 379
Salpeter, E. E. 1955, ApJ, 121, 161
Saro, A., Borgani, S., Tornatore, L., Dolag, K., Murante, G., Biviano, A., Calura, F., & Charlot, S. 2006, MNRAS, 373, 397
Saro, A., De Lucia, G., Borgani, S., & Dolag, K. 2010, MNRAS, 406, 729
Scannapieco, C., Tissera, P. B., White, S. D. M., & Springel, V. 2008, MNRAS, 389, 1137
Schaye, J., Dalla Vecchia, C., Booth, C. M., Wiersma, R. P. C., Theuns, T., Haas, M. R., Bertone, S., Duffy, A. R., McCarthy, I. G., & van de Voort, F. 2010, MNRAS, 402, 1536
Schmidt, M. 1959, ApJ, 129, 243
Sheth, R. K., & Tormen, G. 2004, MNRAS, 350, 1385
Sijacki, D., & Springel, V. 2006, MNRAS, 366, 397
Silva, L., Granato, G. L., Bressan, A., & Danese, L. 1998, ApJ, 509, 103
Smith, B. D., & Sigurdsson, S. 2007, ApJ, 661, L5
Springel, V., White, S. D. M., Tormen, G., & Kauffmann, G. 2001, MNRAS, 328, 726
Springel, V., Di Matteo, T., & Hernquist, L. 2005a, MNRAS, 361, 776
Springel, V., White, S. D. M., Jenkins, A., Frenk, C. S., Yoshida, N., Gao, L., Navarro, J., Thacker, R., Croton, D., Helly, J., Peacock, J. A., Cole, S., Thomas, P., Couchman, H., Evrard, A., Colberg, J., & Pearce, F. 2005b, Nature, 435, 629
Strickland, D. K., & Stevens, I. R. 2000, MNRAS, 314, 511
Stringer, M. J., Benson, A. J., Bundy, K., Ellis, R. S., & Quetin, E. L. 2009, MNRAS, 393, 1127
Summers, F. J., Davis, M., & Evrard, A. E. 1995, ApJ, 454, 1
Sutherland, R. S., & Dopita, M. A. 1993, ApJS, 88, 253
Tabor, G., & Binney, J. 1993, MNRAS, 263, 323

Tasitsiomi, A. 2003, Int. J. Mod. Phys. D, 12, 1157
Thomas, D. 1999, MNRAS, 306, 655
Tinker, J. L., Conroy, C., Norberg, P., Patiri, S. G., Weinberg, D. H., & Warren, M. S. 2008, ApJ, 686, 53
Toomre, A., & Toomre, J. 1972, ApJ, 178, 623
Tremonti, C. A., Heckman, T. M., Kauffmann, G., Brinchmann, J., Charlot, S., White, S. D. M., Seibert, M., Peng, E. W., Schlegel, D. J., Uomoto, A., Fukugita, M., & Brinkmann, J. 2004, ApJ, 613, 898
Tully, R. B., & Fisher, J. R. 1977, A&A, 54, 661
Tweed, D., Devriendt, J., Blaizot, J., Colombi, S., & Slyz, A. 2009, A&A, 506, 647
van de Voort, F., Schaye, J., Booth, C. M., Haas, M. R., & Dalla Vecchia, C. 2011, MNRAS, 554–+
van den Bosch, F. C., Mo, H. J., & Yang, X. 2003, MNRAS, 345, 923
van Dokkum, P. G. 2005, AJ, 130, 2647
Viola, M., Monaco, P., Borgani, S., Murante, G., & Tornatore, L. 2008, MNRAS, 383, 777
Wang, L., Li, C., Kauffmann, G., & De Lucia, G. 2006, MNRAS, 371, 537
Wang, L., Li, C., Kauffmann, G., & De Lucia, G. 2007, MNRAS, 377, 1419
Weiner, B. J., Coil, A. L., Prochaska, J. X., Newman, J. A., Cooper, M. C., Bundy, K., Conselice, C. J., Dutton, A. A., et al. 2009, ApJ, 692, 187
Weinmann, S. M., van den Bosch, F. C., Yang, X., & Mo, H. J. 2006, MNRAS, 366, 2
Weinmann, S. M., Kauffmann, G., von der Linden, A., & De Lucia, G. 2010, MNRAS, 406, 2249
Whiley, I. M., Aragón-Salamanca, A., De Lucia, G., von der Linden, A., Bamford, S. P., Best, P., Bremer, M. N., Jablonka, P., Johnson, O., Milvang-Jensen, B., Noll, S., Poggianti, B. M., Rudnick, G., Saglia, R., White, S., & Zaritsky, D. 2008, MNRAS, 387, 1253
White, S. D. M. 1976, MNRAS, 177, 717
White, S. D. M. 1994, ArXiv Astrophysics e-prints
White, S. D. M., & Frenk, C. S. 1991, ApJ, 379, 52
White, S. D. M., & Rees, M. J. 1978, MNRAS, 183, 341
White, S. D. M., Frenk, C. S., & Davis, M. 1983, ApJ, 274, L1
Wojtak, R., Łokas, E. L., Gottlöber, S., & Mamon, G. A. 2005, MNRAS, 361, L1
Yoshida, N., Stoehr, F., Springel, V., & White, S. D. M. 2002, MNRAS, 335, 762
Yoshida, N., Omukai, K., Hernquist, L., & Abel, T. 2006, ApJ, 652, 6
Zeldovich, I. B., Einasto, J., & Shandarin, S. F. 1982, Nature, 300, 407
Zwicky, F. 1937, ApJ, 86, 217