to $d\mathbf{u} = \dot{\mathbf{u}}\ dt$, with $\dot{\mathbf{u}}$ given by the equation of motion (in practice, more sophisticated time integration schemes are used). The hard part is finding the gravitational force, since this involves summation over $(N-1)$ other particles each time we need a force for one particle. All the craft in the field involves finding clever ways in which all the forces can be evaluated in less than the raw $O(N^2)$ computations per timestep. We will have to omit the details of this, unfortunately, but one obvious way of proceeding is to solve Poisson's equation on a mesh using a Fast Fourier Transform. This can convert the $O(N^2)$ time scaling to $O(N \ln N)$, which is a qualitative difference given that N can be as large as 10^{10}.

These non-linear effects boost the amplitude of the power spectrum at small physical scales (large k scales) as can be seen in Figure 17 . For cosmological observations we need to understand these non-linear effects to high precision. This is one of the issues facing modern day cosmology and non-linear effects can only be calculated through large scale suites of HPC N-body simulations.

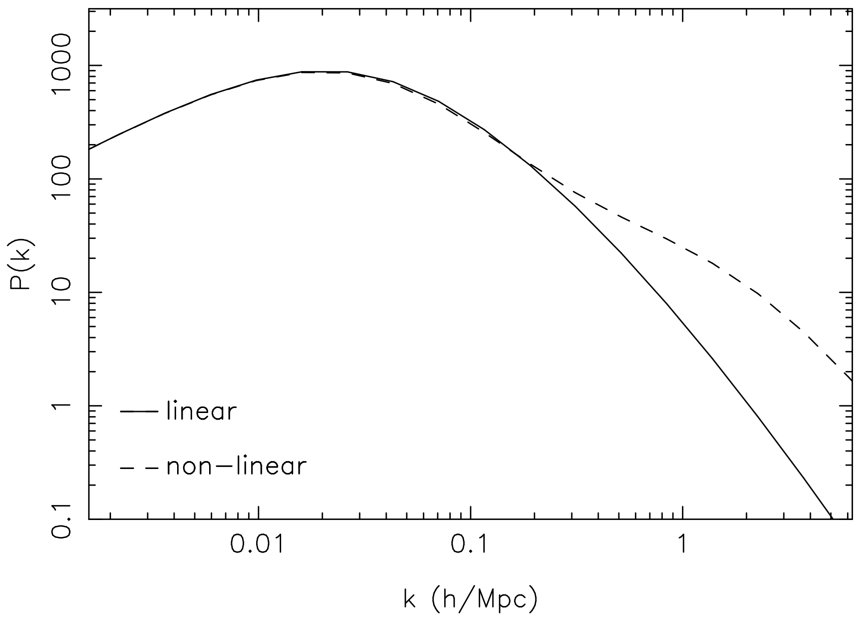

Figure 17. ΛCDM power spectrum normalised by $\sigma_8 = 0.9$. The linear power spectrum is show solid and the non-linear power spectrum is shown dashed using the fitting formula from Smith et al 2003.

PRESS–SCHECHTER AND THE HALO MASS FUNCTION N-body models can yield evolved density fields that are nearly exact solutions to the equations of motion, but working out what the results mean is then more a question of data analysis than of deep insight. Where possible, it is important to have analytic models that guide the interpretation of the numerical results. Press & Schechter (1974) is a key example of a theory which produces results that only slightly differ from full numerical simulations.

Press-Schechter theory assumes that if we smooth the linear density perturbations on some mass scale M, then the fraction of space in which the smoothed density field exceeds some critical threshold δ_c (the **critical overdensity** for collapse) is in collapsed objects of mass greater than M. If the density field is Gaussian, the probability that a given point lies in a region with $\delta > \delta_c$ is

$$p(\delta > \delta_c \mid R) = \frac{1}{\sqrt{2\pi}\,\sigma(R)} \int_{\delta_c}^{\infty} \exp\left(-\delta^2/2\sigma^2(R)\right)\, d\delta, \tag{201}$$

where $\sigma(R)$ is the linear rms in the filtered version of δ. The PS argument now takes this probability to be proportional to the probability that a given point has ever been part of a collapsed object of scale $> R$. This is really assuming that the only objects that exist at a given epoch are those that have only just reached the $\delta = \delta_c$ collapse threshold; if a point has $\delta > \delta_c$ for a given R, then it will have $\delta = \delta_c$ when filtered on some larger scale and will be counted as an object of the larger scale. The problem with this argument is that half the mass remains unaccounted for: PS therefore simply multiplying the probability by a factor 2. This fudge can be given some justification, but we just accept it for now. The fraction of the universe condensed into objects with mass $> M$ can then be written in the universal form

$$F(> M) = \sqrt{\frac{2}{\pi}} \int_{\nu_c}^{\infty} \exp(-\nu^2/2)\, d\nu, \tag{202}$$

where $\nu_c = \delta_c/\sigma(M)$ is the threshold in units of the rms density fluctuation and M is the mass contained in a sphere of comoving radius R in a homogeneous universe

$$M = \frac{4\pi}{3}\,\bar{\rho}\,R^3. \tag{203}$$

This is the linear-theory view, before the object has collapsed. We define the mass function $f(M)$ where $f(M)\,dM$ is the comoving number density of objects in the range dM. The probability of a point in space forming as mass between M and $M + dM$ is dF/dM, therefore;

$$M f(M)/\rho_0 = |dF/dM|, \tag{204}$$

where ρ_0 is the total comoving density. We can write this result in terms of the **multiplicity function**, $M^2 f(M)/\rho_0$,

$$\frac{M^2 f(M)}{\rho_0} = \frac{dF}{d\ln M} = \left|\frac{d\,\ln\sigma}{d\,\ln M}\right| \sqrt{\frac{2}{\pi}}\,\nu\,\exp\left(-\frac{\nu^2}{2}\right). \tag{205}$$

which is the fraction of the mass carried by objects in a unit range of $\ln M$.

Remarkably, given the dubious assumptions, this expression matches very well to what is found in direct N-body calculations, when these are analysed in order to pick out candidate haloes: connected groups of particles with density about 200 times the mean. The PS form is imperfect in detail, but the idea of a mass function that is universal in terms of ν seems to hold, and a good approximation is

$$F(> \nu) = (1 + a\,\nu^b)^{-1} \exp(-c\,\nu^2), \tag{206}$$

where $(a, b, c) = (1.529, 0.704, 0.412)$. Empirically, one can use $\delta_c = 1.686$ independent of the density parameter (see Section 15.8 in Peacock 1999 for the spherical model argument for the value of

δ_c). A plot of the mass function according to this prescription is given in figure 18, assuming what we believe to be the best values for the cosmological parameters. This shows that the Press-Schechter formula captures the main features of the evolution, even though it is inaccurate in detail. We see that the richest clusters of galaxies, with $M \simeq 10^{15}\, h^{-1} M_\odot$, are just coming into existence now, whereas at $z = 5$ even a halo with the mass of the Milky Way, $M \simeq 10^{12}\, h^{-1} M_\odot$ was similarly rare. It can be seen that the abundance of low-mass haloes declines with redshift, reflecting their destruction in the merging processes that build up the large haloes.

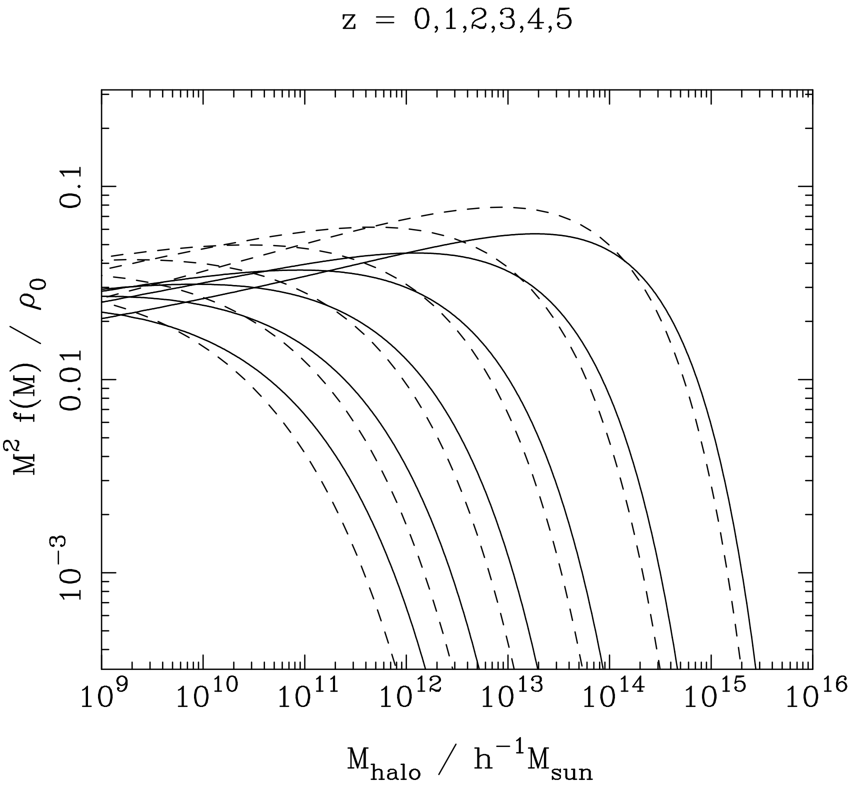

Figure 18. The mass function in the form of the **multiplicity function**: fraction of mass in the universe found in virialized haloes per unit range in $\ln M$. The solid lines show a fitting formula to N-body data and the dashed lines contrast the original Press-Schechter formula.

7.1 A recipe for galaxy formation

Press-Schechter theory has an interesting consequence in the context of the CDM model, where there is power on all scales: the sequence of structure formation must be **hierarchical**. This means that we expect the universe to fragment into low-mass clumps at high redshift, following which a number of clumps **merge** into larger units at later times. This process is controlled by the density variance as a function of smoothing scale, $\sigma^2(R)$. In a hierarchical model, this increases without limit as $R \to 0$, so there is always a critical scale at which $\sigma \simeq 1$. As the density fluctuations grow, this critical scale grows also. These collapsed systems are known as **dark-matter haloes**. The largest such haloes, forming today, are the rich clusters of galaxies.

Since the appreciation in the 1970s that galaxies seemed to be embedded in haloes of dark matter, it has been clear that one should be able to construct an approximate theory for the assembly of galaxies based on the assumption that everything is dominated by the dark matter. Therefore, once we understand the history of the haloes, we should be able to make plausible guesses about how the baryonic material will behave. Over the years, this route has been followed to the point where there now exists an elaborate apparatus known as **semianalytic galaxy formation**. This is not yet a fully satisfactory theory, in that it is not able to make robust predictions of the properties that the galaxy population should have. However, it has succeeded in illuminating the main issues that need to be understood in a complete theory. In essence, semianalytic models include the following elements:

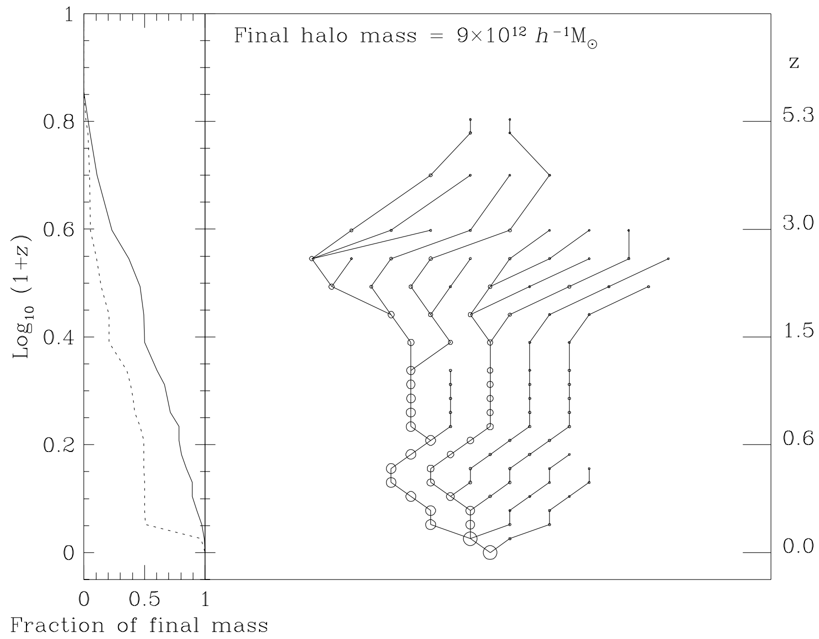

Figure 19. An example of a merger tree for a halo of $M \simeq 10^{13}\, M_\odot$ at $z = 0$, from Helly et al. (2002). The size of circle is proportional to halo mass, and the leftmost panel shows the fraction of the total mass in resolved progenitors (solid) and the mass of the largest progenitor (dashed).

Merger trees. A halo that exists at a given time will have been constructed by the merging of smaller fragments over time. We need to be able to predict this history.

Fate of subhaloes. When haloes merge, they do not instantly lose their identity. Their cores survive as distinct subhaloes for some time. In group/cluster scale haloes, these will mark the locations of the galaxies. In general, subhaloes will eventually merge within the parent halo, and sink to the centre. Thus there is always a tendency to have a dominant central galaxy (e.g. the Milky Way is surrounded by the much smaller Magellanic Cloud dwarfs).

Accounting of gas and stars. The first generation of haloes is assumed to start life with gas distributed along with the dark matter in the universal ratio Ω_b/Ω_{dm}. From the density of the gas, the cooling rate can be calculated. Whatever gas reaches a temperature below 10^4 K is deemed to be a reservoir of cold gas suitable for star formation. Some empirical relation based on the amount and density of this gas is then used to predict a star-formation rate. When haloes merge, their contents of stars, cold gas, and hot gas are added.

Feedback. As we will show below, the above recipe fails to match observation, as it predicts that stars should form most efficiently in the smallest galaxies – so that a system of the size of the Milky Way should be just a collection of globular clusters, rather than predominately a giant gaseous disk. Therefore, the critical (and so far unsolved) problem in galaxy formation is to make gas cool less efficiently. The idea here is that energy is put back into the hot gas as a result of the nonlinear events that happen inside galaxies. These are principally of two kinds: supernova explosions and nuclear activity around a central black hole.

VIRIAL TEMPERATURE There is no time here to dig very deeply into the component parts of this recipe, but a few points are worth making. First consider the characteristic density of a virialized halo. We have argued that this is some multiple $f_c \simeq 200$ of the background density at virialization (or 'collapse'):

$$\rho_c = f_c \, \rho_0 \, (1 + z_c)^3. \tag{207}$$

The virialized potential energy for constant density is $3GM^2/(5r)$, where the radius satisfies $4\pi\rho_c r^3/3 = M$. This energy must equal $3MkT/(\mu m_p)$, where $\mu = 0.59$ for a plasma with 75% hydrogen by mass. Hence, using $\rho_0 = 2.78 \times 10^{11} \Omega_m h^2 \, M_\odot \mathrm{Mpc}^{-3}$, we obtain the **virial temperature**:

$$T_{\rm virial}/\mathrm{K} = 10^{5.1} (M/10^{12} M_\odot)^{2/3} \, (f_c \Omega_m h^2)^{1/3} \, (1 + z_c). \tag{208}$$

This is an illuminating expression. It tells us that the most massive systems forming today, with $M \simeq 10^{15} M_\odot$, will have temperatures of 10^7 – 10^8 K. The intergalactic medium in clusters is thus very hot, and emits in X-rays. It also cools very inefficiently, since such hot plasmas emit only bremsstrahlung. Conversely, pregalactic systems with $M \lesssim 10^9 M_\odot$ at $z \simeq 10$ have a virial temperature that is barely at the level of 10^4 K required for ionization. Their gas is thus predominately neutral, and should form stars with maximum efficiency. This is the cooling paradox referred to above.

But the same formula allows us to see how to escape from the paradox. The virial temperature is equivalent to a velocity dispersion, which is essentially the velocity at which particles orbit in the dark-matter potential well. This velocity therefore also gives the order of magnitude of the escape velocity for the system. Haloes with a virial temperature of only $\sim 10^4$ K thus constitute very shallow potential wells and will lose any of their gas that becomes heated to $\gtrsim 10^5$ K. This is liable to happen as soon as any supernovae from the first generation of star formation explode. For type II supernovae associated with massive stars, this can be virtually instantaneous ($\lesssim 10^7$ years). Star formation in these early dwarf galaxies might well be expected to be self-quenching. Indications that this process did happen can be found when measuring HI rotation curves of dwarfs: the typical baryon fraction is only about 1% (as opposed to something close to the global 20% in clusters).