Approximating non-linear structure: Spherical collapse and the halo model

Barnaby Rowe

March 9, 2006

The formation of mass structure in the Universe by gravitational instability is a highly favoured model in modern cosmology, underpinning the so-called ΛCDM paradigm. A study of the dynamics of linear perturbation theory, as outlined by Caroline Francis in the preceding seminar, can tell us much about large-scale structure but breaks down wherever the local self-gravitation of matter becomes an important consideration. This we know as the **non-linear** regime, encompassing much of the most interesting observed structure in the Universe; indeed dark matter halos are *defined* using the results of non-linear theory.

Although the full development of this self-gravitational collapse cannot be solved without employing large-scale N-body simulations, the problem can be treated approximately. The aim of this seminar is to outline one of the most recent of such treatments, known as the **halo model**, which attempts to reproduce the statistics of matter structure in all regimes. In the next hour I will attempt to give a feeling of the satisfying theoretical simplicity of its component parts, and try to describe some of its more important successes and limitations.

As Caroline discussed previously in her seminar, we commonly describe the patterns we see in the matter distribution in terms of a dimensionless density perturbation:

$$1 + \delta(\boldsymbol{x}) \equiv \frac{\rho(\boldsymbol{x})}{\langle \rho \rangle}. \tag{1}$$

We will use this notation throughout the seminar. We will also only tend to consider results for a flat, matter-dominated Universe with density parameter

$$\Omega_0 = \Omega_{\rm tot} = 1, \tag{2}$$

otherwise referred to as the Einstein-de Sitter model. Although this does not accurately represent the vacuum-energy dominated Universe we believe to exist around us, many of the most basic results for such a cosmology cannot be obtained without numerical integration. We thus choose the simpler case of equation (2) as an attempt to utilise the extra insight available from analytically derived results. It turns out that these results are usually sufficiently accurate in any case – not to mention the fact that numerical integration is the subject of *Anita's* pedagogy...

The spherical collapse model of halo formation

Our first step in modelling the large-scale evolution of matter into non-linear structure is a description of the 'microscopic' case: the collapse of a single overdense region of the Universe into a self-gravitating halo, via a model known as spherical "top hat" collapse. This model assumes spherical symmetry for our overdensity, but is able to fully describe the early, intermediate and late stages of its evolution and collapse.

It is a highly useful result within General Relativity that **Friedmann's Equation**, written in its dimensional form as

$$\dot{R}^2 = \frac{8\pi G}{3}\rho R^2 - kc^2, \tag{3}$$

which describes the expansion of an isotropic and homogeneous Universe, applies equally to a spherical Newtonian mass (see Peacock 1999). The radius R of an overdense sphere behaves in the same way as the expansion factor for a closed sub-universe, and we are therefore able to model the general growth of a spherically symmetric density perturbation using the same equations as classical cosmology.

Equation (3) can be most easily solved with a re-parameterisation in terms of the variable η, where $\mathrm{d}\eta = c\mathrm{d}t/R(t)$. Assuming that $\rho(t)$ can be written as $\rho(t) = \rho_0 R_0^3/R^3$ we then have

$$\left(\frac{\mathrm{d}R}{\mathrm{d}\eta}\right)^2 = \frac{8\pi G\rho_0 R_0^3}{3c^2}R - kR^2. \tag{4}$$

Defining the useful scaling constant

$$R_* = \frac{4\pi G\rho_0 R_0^3}{3c^2} = \frac{GM}{c^2} \tag{5}$$

where M is the mass initially (and subsequently) enclosed in our overdense sphere, we can rewrite equation (4) in the following simple form:

$$\left[\frac{\mathrm{d}}{\mathrm{d}\eta}\left(\frac{R}{R_*}\right)\right]^2 = 2\left(\frac{R}{R_*}\right) - k\left(\frac{R}{R_*}\right)^2. \tag{6}$$

For an overdense region $k = +1$, and this equation has the solution

$$\frac{R(\eta)}{R_*} = 1 - \cos\eta. \tag{7}$$

The cosmic time can be found by recalling our definition of η, from which

$$t(\eta) = \int_0^\eta \mathrm{d}\eta' \frac{R(\eta')}{c}. \tag{8}$$

The two equations which completely describe the evolution of our spherically symmetric overdensity are thus

$$R(\eta) = R_*(1 - \cos\eta) \tag{9}$$

$$t(\eta) = \frac{R_*}{c}(\eta - \sin\eta) \tag{10}$$

where R_* is given by equation (5). This solution will form the basis of our approximate theory of non-linear collapse, but to understand its consequences we need to examine its behaviour at some important epochs in the history of the overdensity.

Evolution for $\eta \ll 1$

In order to describe the early expansion of the sphere as a function of t, we must expand the solutions given in (9) and (10) as power series in η. Doing so around $\eta = 0$, and ignoring terms of order greater than η^5, we find

$$R(t) \simeq \frac{R_*}{2}\left(\frac{6ct}{R_*}\right)^{2/3}\left[1 - \frac{1}{20}\left(\frac{6ct}{R_*}\right)^{2/3}\right]. \tag{11}$$

Here we recover, as we should rightly hope so to do, two important results from both the underlying cosmological expansion and linear theory. The leading-order term

$$R(t \to 0) = \frac{R_*}{2}\left(\frac{6ct}{R_*}\right)^{2/3} = \left(\frac{9GMt^2}{2}\right)^{1/3} \tag{12}$$

can be rearranged to give the early-epoch density dependence,

$$\rho(t \to 0) = \frac{3M}{4\pi R^3(t \to 0)} = \frac{1}{6\pi G t^2} = \rho_0(t), \tag{13}$$

recovering the standard result for the evolution of the cosmological (critical) density of an Einstein-de Sitter Universe (see e.g. Coles & Lucchin 1995).

The next order term in the expansion can also be examined. Since the density of the sphere is given by $\rho \propto 1/R^3$, the fractional overdensity is thus

$$\frac{\delta\rho}{\rho} = -3\frac{\delta R}{R} = \frac{3}{20}\left(\frac{6ct}{R_*}\right)^{2/3} \tag{14}$$

to first order. Remembering that for an Einstein-de Sitter Universe $a(t) \propto t^{2/3}$, we are clearly recovering a correctly-scaled version of the linear-theory result $\delta_{\rm lin} \propto a$, as found by Caroline in the previous seminar.

Turnaround

At later times, the evolution of the sphere will differ significantly from that of the cosmological background. Examination of equation (9) shows immediately that a maximum radius $R_{\rm max} = 2R_*$ is reached when $\eta = \pi$, occurring at a time $t_{\rm max} = \pi R_*/c$. Using our result from equations (12) and (13), we calculate the density enhancement of our evolving sphere relative to the cosmological background as

$$1 + \delta_{\rm max} = \frac{\rho_{\rm max}}{\rho_0(t_{\rm max})} = \frac{(R_*/2)^3(6ct_{\rm max}/R_*)^2}{R^3_{\rm max}} = \frac{9\pi^2}{16} \simeq 5.55. \tag{15}$$

Using equation (14) we can also calculate what linear theory would have predicted for the perturbation at this time:

$$\delta_{\rm lin} = \frac{3}{20}(6\pi)^{2/3} \simeq 1.08. \tag{16}$$

The utility of the top hat model is becoming apparent: not only can we now describe the full non-linear evolution of individual density perturbations, but the results for these objects can be very easily related to those of linear theory.

Collapse and virialization

We now turn to what is in some ways the most interesting epoch in the history of a matter overdensity: its collapse into the self-gravitating object commonly known as a 'halo'. Ubiquitous in modern N-body simulations of collisionless dark matter, the concept of a halo has existed since the advent of accurate measurements of galaxy rotation curves, and needless to say forms an important part of the halo model too.

If only gravity operates on a perfectly spherically symmetric perturbation, the region will collapse to a black hole when $\eta = 2\pi$, at a time we label $t_{\rm coll} = 2\pi R_*/c$. When this occurs the corresponding linear prediction for the density perturbation in this region is $\delta_{\rm lin} = (3/20)(12\pi)^{2/3} \simeq 1.69$.

However, total collapse will never occur in practice; slight departures from pure spherical symmetry will cause the kinetic energy of collapse to be converted into random motions. The perturbation will thus eventually reach some form of thermalized, bound, equilibrium state, which if we invoke the **virial theorem** will have total kinetic energy K related to potential energy V by

$$V = -2K. \tag{17}$$

Assuming energy conservation during the transition to this equilibrium state, equation (17) can be shown to imply that the radius $R_{\rm vir}$ of such a virialized perturbation, *which we define as being a halo*, is given by

$$R_{\rm vir} = \frac{1}{2} R_{\rm max} = R_*. \tag{18}$$

In our simple top hat model this will occur at $\eta = 3\pi/2$, and the epoch this defines is often chosen to roughly estimate the density contrast we should expect for collapsed, relaxed objects.

It could be expected however that the process of virialization take a little longer than predicted by our symmetric model, and it is common for authors to assume that this $R_{\rm vir}$ is in fact only reached after $t_{\rm coll}$. The full non-linear density contrast is then

$$1 + \delta_{\rm vir} = \frac{(R_*/2)^3 (6ct_{\rm coll}/R_*)^2}{R_{\rm vir}^3} = \frac{(6\pi)^2}{2} \simeq 178. \tag{19}$$

As before, linear theory predicts a perturbation of $\delta_{\rm lin} \simeq 1.69$ at $t_{\rm coll}$ for the region corresponding to this collapsed halo.

This all leads to an approximate but usefully simple theoretical model of non-linear structure formation whereby one assumes that overdense regions evolve as prescribed by linear theory until $\delta_{\rm lin} \gtrsim 1$, after which point we expect to find a collapsed halo with an average density contrast of order a few hundred.

Questions should be asked of the validity of the assumption of spherical symmetry in this analysis, especially as we should expect in general that overdense (or underdense) regions in our Universe will be quite aspherical. It is found, however, that a density contrast $\sim 100 - 200$ turns out to be quite successful at defining and hence finding halos in N-body simulations, which is cause for some encouragement (see e.g., Navarro, Frenk & White 1997; Peacock 1999). In such simulations a commonly used definition of the mass of a halo is M_{200}, defined as the mass encompassed by a sphere centred on the halo within which $1 + \delta = 200$.

We are left therefore in a pretty strong position: we can use the relatively simple results of linear theory to predict the regions in which we might expect halos to form, and can make a reasonable guess at how dense we might expect the final collapsed halo to be. Surely therefore, if we know or can make assumptions about the statistics of the initial density fluctuations in the Universe, we should be able to evolve these fluctuations using linear theory... Then, using the results of the top hat model, we should be able to populate sufficiently dense regions in our model linear Universe with appropriate collapsed halos.

We can certainly try, and the remainder of this seminar will look briefly at some of these attempts to describe the properties of non-linear structure without resort to sheer numerical brute-force.

Press-Schechter theory

The Press & Schechter (1974; PS hereafter) formalism is an attempt to analytically estimate the **mass function** $n(M)$ of bound objects in the Universe, defined such that $n(M){\rm d}M$ is the comoving number density of objects with mass between M and $M + {\rm d}M$. One success of non-linear theory has been its abiltity to generate a workable approximate form for this statistic, in good agreement with simulation data, allowing the rapid calculation of a vast number of cosmological predictions requiring estimated numbers of galaxies, clusters and halos in the Universe.

I must apologise in advance for the fact that a little knowledge of power spectrum analysis is necessary to reproduce some of the results in this section: time does not permit me to enter into the details, so I instead refer the interested reader to Peacock (1999) or Coles & Lucchin (1995).

In general, we assume that the density fluctuations predicted by linear theory and associated with a given mass scale M can be described by a density fluctuation field $\delta(\boldsymbol{x}; R_{\rm f}) \equiv \delta_M$, filtered on a spacial scale $R_{\rm f}$ corresponding to a mass $M \propto \rho_0 R_{\rm f}^3$. Obviously, some form of artifical smoothing (or filtering) of the density field over a given spatial scale is going to be needed to assign a mass to extended regions in space. This filtering function can be of arbitrary form, sometimes being taken to be a top hat function so as to sharply define the volume being considered, and sometimes a Gaussian function for computational convenience (see Peacock 1999).

One of the basic assumptions of the PS approach is that the distribution of values of this linear density field δ_M will be Gaussian. If so, the probability that a point lies within a region of scale $R_{\rm f}$ exhibiting a density fluctuation between δ_M and $\delta_M + {\rm d}\delta_M$ is simply

$$p(\delta_M; R_{\rm f}){\rm d}\delta_M = \frac{1}{\sqrt{2\pi\sigma^2(R_{\rm f})}} \exp\left[-\frac{\delta_M^2}{2\sigma^2(R_{\rm f})}\right]{\rm d}\delta_M. \quad (20)$$

We have here introduced $\sigma^2(R_{\rm f})$, defined to be the dimensionless **mass variance** we would expect to see due to linear perturbation growth in a region of scale $R_{\rm f}$;

$$\sigma^2 = \frac{\left\langle (M - \langle M \rangle)^2 \right\rangle}{\langle M \rangle^2}. \quad (21)$$

How we determine the functional form of this mass variance could easily occupy another seminar such as this one. It can be expressed in term of the linear mass **power spectrum**, which tells us about the degree of 'clumpiness' (or correlation) we expect to find in the Universe at a given scale. The power spectrum will be a function of the wavenumber $k = 2\pi/L$, where L is the scale in question, and is conveniently described as the dimensionless contribution to $\left\langle \delta_k^2 \right\rangle$ per unit interval of $\ln k$, i.e.

$$\Delta^2_{\rm lin}(k) \equiv \frac{{\rm d}\left\langle \delta_k^2 \right\rangle}{{\rm d}\ln k} \quad (22)$$

The linear theory variance over a filtering scale $R_{\rm f}$ may then be expressed as

$$\sigma^2(R_{\rm f}) = \int_0^\infty \Delta^2_{\rm lin}(k) W_{\rm f}^2(kR_{\rm f}) \frac{{\rm d}k}{k}, \quad (23)$$

where $W_{\rm f}(kR_{\rm f})$ is the Fourier transform of the filtering function we are using to define spatial scales. $\Delta^2_{\rm lin}(k)$ can be calculated from theoretical models and expressed in terms of the primordial scale-free power spectrum and objects known as **transfer functions** $T(k)$. These are commonly approximated by fitting formulae such as those of Bardeen et al. (1986) or more recently Eisenstein & Hu (1998), who accurately describe the so-called "baryon wiggles". The normalisation of $\Delta^2_{\rm lin}$ is set by the value of the well-known cosmological parameter σ_8, which is simply the value of σ measured for a smoothing radius $R_{\rm f} = 8h^{-1}$Mpc, roughly matching the typical scales of rich clusters.

Returning to equation (20), the total probability of a given point lying within a region of scale $R_{\rm f}$ in which the linear density perturbation δ_M is greater than some critical density $\delta_{\rm c}$ is thus:

$$F(>\delta_{\rm c}; R_{\rm f}) = \int_{\delta_{\rm c}}^\infty \frac{1}{\sqrt{2\pi\sigma^2(R_{\rm f})}} \exp\left[-\frac{\delta_M^2}{2\sigma^2(R_{\rm f})}\right]{\rm d}\delta_M. \quad (24)$$

Crucially, what the PS formalism now argues is that with a value of $\delta_{\rm c}$ chosen to be $\simeq 1.69$ (motivated, of course, by the results of the spherical collapse model), this probability is *proportional* to the probability that a given point has ever been part of a collapsed object of scale greater than R_{f}.

Examination of this argument reveals its underlying implication that the only *isolated* objects that exist at a given epoch are those that have just collapsed ($\delta_M = \delta_{\mathrm{c}}$). Regions with $\delta_M > \delta_{\mathrm{c}}$ for a given R_{f} are simply counted as part of larger collapsed object on a scale R'_{f} for which $\delta_{M'} = \delta_{\mathrm{c}}$.

If we ask the obvious next question, i.e. what about all the mass in regions where $\delta_M < 0$, we are led to conclude that it remains uncounted. This fact led PS to the further conclusion that the correct constant of proportionality they sought, relating (24) to the probability of a region of space having ever been contained within a halo of mass M, was simply the integer 2. Most surprisingly, it turns that this assignment of a single multiplicative factor of 2 can actually be justified, although this result is dependent on the choice of spatial filtering: see Peacock & Heavens (1990) and Bond et al. (1991) for a full discussion of this topic, known as the "cloud-in-cloud" problem.

Accepting all this, the integral probability (24) is related to the mass function $n(M)$ by

$$n(M) = \frac{2}{M}\rho_0 \left| \frac{\mathrm{d}F(M)}{\mathrm{d}M} \right|, \tag{25}$$

since the mass derivative of F gives the fraction of independent regions which have become self-bound in lumps with mass between M and $M + \mathrm{d}M$. Written explicitly, the PS mass function is therefore

$$n(M) = \frac{\delta_{\mathrm{c}}}{\sqrt{2\pi\sigma^2(M)}} \frac{1}{M} \left| \frac{\mathrm{d}\sigma}{\mathrm{d}M} \right| \exp\left[-\frac{\delta_{\mathrm{c}}^2}{2\sigma^2(M)} \right]. \tag{26}$$

We can examine the approximate properties of this function if we express σ as a negative power-law in mass: in this case the $n(M)$ has itself power-law behaviour for low M followed by an exponential cut-off on larger mass scales where $\sigma \lesssim 1$.

The Press-Schechter formalism has proven successful and influential because it offers a quick method of describing in a more analytic fashion the behaviour of fully numeric N-body simulations (and of doing so fairly well). It could be *rightly* argued however that it constitutes little more than a glorified recipe for labelling points in the original linear density field, failing to describe the motion of density peaks away from their original positions during the process of non-linear evolution. This is brought into sharp relief by the results of the Zeldovich approximation (see Peacock 1999, or another seminar!) and will clearly cause problems in the description of cosmological voids, an area which we've already seen causes PS some conceptual difficulties.

Notwithstanding the above, the PS approach works to within a reasonable degree of accuracy, and has proven an invaluable starting point for more detailed analyses (see e.g., Yano, Nagashima & Gouda 1996, Sheth & Tormen 1999). It therefore fulfils part of the brief we set ourselves earlier: we can now populate a model Universe (with a linear theory motivated density field) with roughly the right numbers of collapsed halos. One further condition we might impose for the

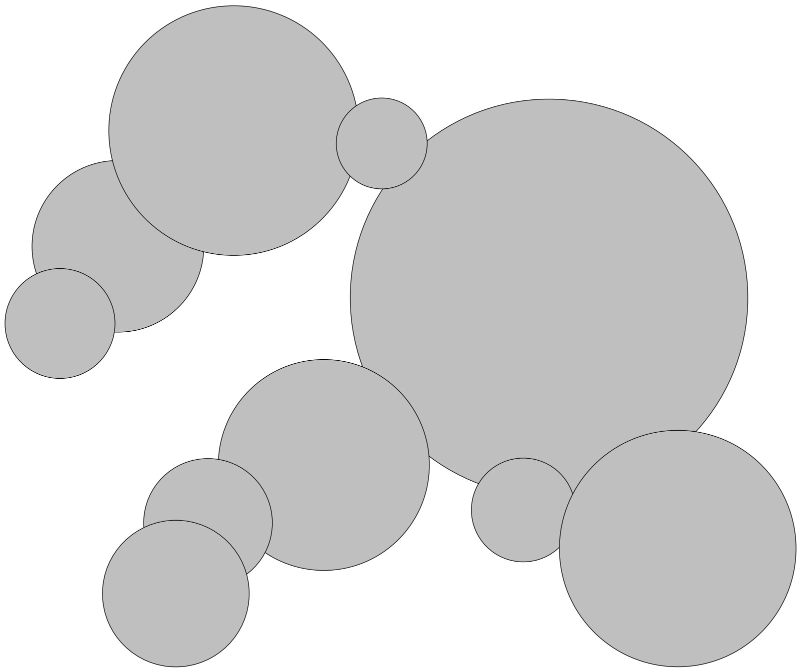

Figure 1: Schematic of the halo model picture of cosmological mass distributions, i.e. divided entirely into virialized halos of varying masses and mass-dependent profiles.

success of our model would be that it approximately reproduce the correlation statistics, or clumpiness, of matter at a variety of scales. This is what the halo model seeks to do, as we shall now discuss, by building on all we've found above *and* borrowing a little from simulation data.

The halo model: predicting the power spectrum on all scales

The PS model of the Universe can be interpreted as dividing all matter at a given time into virialized halos of different masses, and this picture is also the starting point of the **halo model**, as recently developed by several authors (Seljak 2000; Peacock 2000; Peacock & Smith 2000; Ma & Fry 2000a, 2000b).

The model then takes on the old idea (see Neyman, Scott & Shane 1953; Peebles 1974) that on small scales (large k) the statistics of matter correlations are dominated by the density profiles of the halos themselves – the small scale correlations are assumed to be *intra*-halo correlations. This is where recent simulation work lends a helping hand, producing a number of contenders for a 'universal' density profile: analytic functions which accurately fit numerical dark matter halos and can be correctly scaled over multiple decades in mass. The most famous example is that of Navarro, Frenk & White (1997), while more recent profiles such as those of Moore et al. (1999) and Navarro et al. (2004)

attempt to better fit the inner regions of CDM halos as resolved by up-to-date simulations.

It will again be convenient to express the power spectrum as the dimensionless contribution to the $\left\langle {\delta_k}^2 \right\rangle$ per unit interval of $\ln k$, i.e.

$$\Delta^2(k) \equiv \frac{\mathrm{d}\left\langle {\delta_k}^2 \right\rangle}{\mathrm{d}\ln k}, \tag{27}$$

I refer the reader to Peacock & Smith (2000) for the details, but for an assumed mass-scaling universal profile, the overall power spectrum $\Delta^2_{\mathrm{halos}}$ due to the internal structure of halos can be expressed by summing the different contributions from halos of all masses using the PS mass function $n(M)$. We can write this as the integral

$$\Delta^2_{\mathrm{halos}}(k) = \int_0^\infty f(k, M) n(M) \mathrm{d}M \tag{28}$$

where $f(k, M)$ is simply the contribution (in units of volume) to the power spectrum of a single given halo profile of mass M.

On larger scales the halos are assumed to cluster according to linear theory, and so the *inter*-halo correlations are described by the linear power spectrum Δ^2_{lin}: as mentioned earlier this can be derived theoretically and calculated using transfer functions. These separate contributions to the variance are then just added, so that our model Universe has

$$\Delta^2_{\mathrm{tot}} = \Delta^2_{\mathrm{halos}} + \Delta^2_{\mathrm{lin}}. \tag{29}$$

This *is* the halo model, and it matches the dark matter power spectra of N-body simulations very well, despite its apparent simplicity; see Figure 2. While the model requires ingredients to be fixed by numerical experiments, it provides an extremely useful framework for accurately describing the statistics of mass clustering analytically, and therefore quickly.

There are slightly different flavours of the halo model: Seljak (2000), for example, attempts a slightly more accurate model of the correlations between halos by convolving the linear theory density field with a biased distribution of halo profiles. Additionally, it should be noted for completeness that in practice the Sheth & Tormen (1999) mass function is used to sum the contributions of halos, in place of the PS, as it also allows better for the spatial biasing of large-scale structure. It turns out however that both these details have only minor impact on the results (see Seljak 2000).

Finally, there are also issues about the conservation of mass in the halo model, as it could be argued that we should really be *shifting* the linear power rather than simply adding to it. A full treatment – like so many of the things that I have mentioned only in passing – lies outside the scope of a seminar such as this, but thanks for turning up anyway!

References

Bardeen, J. M., Bond, J. R., Kaiser, N., Szalay, A. S. 1986, 304, 15.

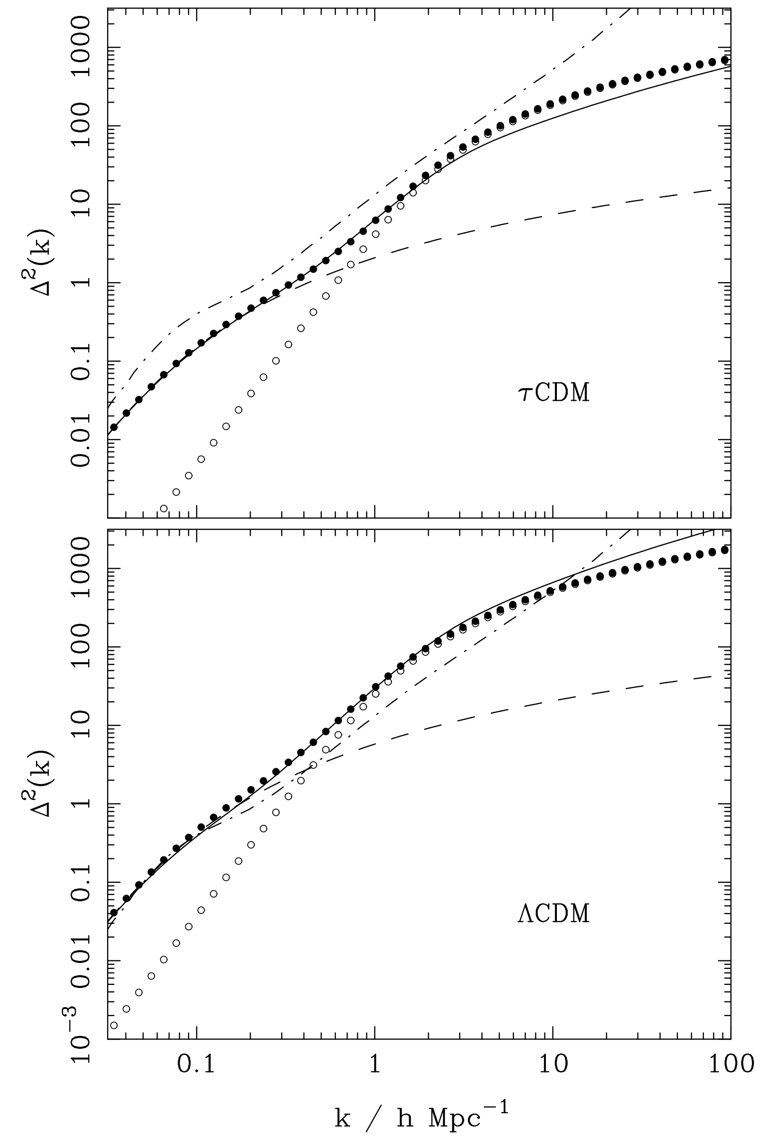

Figure 2: Density power spectrum computed for the τCDM (top) and ΛCDM (bottom) models. The linear spectrum is shown dashed; the solid line shows the non-linear spectrum, calculated and accurately fitted to CDM simulations according to the approximation of Peacock & Dodds (1996). The spectrum according to randomly placed halos is denoted by open circles; if the linear power spectrum is added, the main features of the non-linear spectrum are well reproduced (filled circles). For reference, the dot-dashed lines shows the APM power spectrum (Maddox, Efstathiou & Sutherland 1996). Figure taken from Peacock & Smith (2000).

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

Coles, P., Lucchin, F. 1995, Cosmology. Publisher: Chichester, UK: John Wiley & Sons Ltd., 1995. ISBN: 047195473.

Eisenstein, D. J., Hu, W. 1998, ApJ, 496, 605.

Ma, C.-P., Fry, J. N. 2000a, ApJ, 531, L87.

Ma, C.-P., Fry, J. N. 2000b, ApJ, 543, 503.

Maddox, S., Efstathiou, G., Sutherland, W. J. 1996, MNRAS, 283, 1227.

Moore, B., Quinn, T., Governato, F., Stadel, J., Lake, G. 1999b, MNRAS, 310, 1147.

Navarro, J. F., Frenk, C. S., White, S. D. M. 1997, ApJ, 490, 493.

Navarro, J. F., et al. 2004, MNRAS, 349, 1039.

Neyman, J., Scott, E. L., Shane, C. D. 1953, ApJ, 117, 92.

Peacock, J. A. 1999, Cosmological physics. Publisher: Cambridge, UK: Cambridge University Press, 1999. ISBN: 0521422701.

Peacock, J. A. 2000, preprint (astro-ph/0002013).

Peacock, J. A., Dodds, S. J. 1996, MNRAS, 280, L19.

Peacock, J. A., Heavens, A. F. 1990, MNRAS, 243, 133.

Peacock, J. A., Smith, R. E. 2000, MNRAS, 318, 1114.

Peebles, P. J. E. 1974, A&A, 32, 197.

Press, W. H., Shechter, P. 1974, ApJ, 187, 425.

Seljak, U. 2000, MNRAS, 318, 203.

Sheth, R. K., Tormen, G. 1999, MNRAS, 308, 119.

Taylor, A. N. 2005, Course notes on Astrophysical Cosmology, *available at* http://www.roe.ac.uk/~ant/.

White, M., Hernquist, L., Springel, V. 2001, ApJL, 550, L129.

Yano, T., Nagashima, M., Gouda, N. 1996, ApJ, 466, 1.