8

The nonlinear evolution

8.1 Introduction

e linear perturbation theory developed in chapters 4 and 5 fails when e density contrast becomes nearly unity. Since most of the observed uctures in the universe – like galaxies, clusters etc. – have density con- sts far in excess of unity, their structure can be understood only by a ly nonlinear theory. This chapter discusses the nonlinear evolution of rturbations, starting from where we left off in chapters 4 and 5. Nonlin- r evolution can be studied analytically if some simplifying assumptions e made. Such simplified analytic models are studied in sections 8.2, 5 and 8.6. The results of these calculations are used to understand the operties of the galaxies in sections 8.3, 8.4 and 8.7. Sections 8.8 and 9 attempt to provide a more detailed modelling of the properties of spi- l and elliptical galaxies and discuss the difficulties encountered in such tempts. Finally, section 8.10 reviews the results of N-body simulations ed in studying nonlinear evolution.

8.2 Spherical model for the nonlinear collapse

e evolution of density perturbations in the linear regime was analyzed chapters 4 and 5. The final result of this analysis was an expression for e processed power spectrum $P(k)$ at $t \gtrsim t_{\rm dec}$. The observed isotropy of e MBR guarantees that the density contrast δ_k must have been quite all ($\delta_k \lesssim 10^{-4}$ or so) at this epoch, implying that the evolution of the nsity contrast can be studied using linear theory at $t \gtrsim t_{\rm dec}$ and that grows in proportion to the scale factor $a(t)$. At some later time, $t_{\rm nl}(\lambda)$, e density contrast at a wavelength λ will become comparable to unity. er, $t > t_{\rm nl}(\lambda)$, the linear perturbation theory fails at this wavelength and e have to study the evolution using some other techniques.

The Fourier transform $\delta_k(t)$ of the density contrast $\delta(t, \mathbf{x})$ was useful in e linear regime because each mode was evolving independently. Since

this is no longer true in the nonlinear limit, there is no specific advantage *in using the Fourier components; it is better to study the evolution of* $\delta(\mathbf{x}, t)$ directly, in the $\mathbf{x}$ space.

This may be done as follows: Consider the density contrast $\delta(\mathbf{x}, t_i)$ in the universe at some time t_i. This density contrast will divide the universe into several overdense ($\delta > 0$) and under dense ($\delta < 0$) regions. It is reasonable to expect that regions which are significantly overdense will collapse and (eventually) form gravitationally bound objects. In these overdense regions, the self-gravity of the local mass concentration will work *against* the expansion of the universe; i.e., this region will expand at a progressively *slower* rate compared to the background universe. Such a slowing down will increase the density contrast between the overdense region and the background universe and – consequently – make the gravitational potential of the local mass concentration (in that region) more and *more dominant. Eventually, such a region will collapse under its own* self-gravity and will form a bound system.

The details of the above process will depend on the initial density profile. The simplest model which one can study analytically[1] is based *on the assumption that the overdense region is spherically symmetric* (about some point). Let us suppose that the overdense region we are interested in has an initial density distribution

$$\rho(r, t_i) = \rho_b(t_i) + \delta\rho(r, t_i) = \rho_b(t_i)[1 + \delta_i(r)] \tag{8.1}$$

where $\delta_i(r) = \delta(r, t_i)$ is the initial density contrast which is some specified, *non-increasing*, function of r. Since we are now interested in perturbations with $\lambda \ll d_H$, the size R of the overdense region (which may *be taken to be the scale over which* δ_i *is significant*) can be taken to be much smaller than the Hubble radius. In this case, we can study the dynamics of this region using the Newtonian approximation developed in chapter 4. In the Newtonian limit, it is convenient to use the proper radial *coordinate* $r = a(t)|\mathbf{x}|$ *where* $\mathbf{x}$ *is the comoving Friedmann coordinate.* The dynamics of the overdense region is determined by the gravitational potential

$$\begin{aligned}\phi_{\text{total}}(r, t) &= \phi_b(r, t) + \delta\phi(r, t) = -\frac{1}{2}\left(\frac{\ddot{a}}{a}\right) r^2 + \delta\phi(r, t) \\ &= \frac{2\pi}{3} G\rho_b r^2 + \delta\phi(r, t)\end{aligned} \tag{8.2}$$

where ϕ_b is the equivalent Newtonian potential of the Friedmann metric (see chapter 4) and $\delta\phi$ is the potential generated due to the excess density $\delta\rho(r, t)$. The motion of a thin shell of particles located at a distance r is governed by the equation

$$\frac{d^2\mathbf{r}}{dt^2} = -\nabla\phi_{\text{total}} = -\frac{4\pi G\rho_b(t)}{3}\mathbf{r} - \nabla(\delta\phi) = -\frac{GM_b}{r^3}\mathbf{r} - \frac{G\delta M(r, t)}{r^3}\mathbf{r}. \tag{8.3}$$

writing the second term, we have used the fact that, for a spherically mmetric density distribution, the gravitational force only depends on e mass δM contained inside the shell. Here M_b and $\delta M(r,t)$ stand for

$$M_b = \frac{4\pi}{3}\rho_b(t)r^3 = \frac{4\pi}{3}\rho_b(t)a^3(t)x^3 = \text{constant}; \tag{8.4}$$

$$\delta M(r,t) = 4\pi\int_0^r \delta\rho(q,t)q^2\,dq = 4\pi\rho_b(t)\int_0^r q^2\delta(q,t)\,dq. \tag{8.5}$$

simplify the analysis of the problem, we will assume that the spherical ells do not cross each other during the evolution. That is, if we initially bel the shells as $1, 2, \cdots$ etc. with the radii $r_1 < r_2 < r_3 \cdots$ etc., then the bsequent evolution is assumed to preserve the ordering $r_1 < r_2 \cdots$. In ch a case the mass contained within a shell of radius r does not change th time: $\delta M(r,t) = \delta M(r,t_i) = \text{constant}$. We can now combine the o terms in (8.3) to write

$$d^2r/dt^2 = -GM/r^2, \tag{8.6}$$

here

$$M = \rho_b\left(\frac{4\pi}{3}r_i^3\right)(1+\bar{\delta}_i), \quad \bar{\delta}_i = \left(\frac{3}{4\pi r_i^3}\right)\int_0^{r_i}\delta_i(r)4\pi r^2\,dr. \tag{8.7}$$

ere r_i is the initial radius of the shell with mass M and $\bar{\delta}_i$ is the average lue of δ within r_i at time t_i. The first integral of equation (8.6) is

$$\frac{1}{2}\left(\frac{dr}{dt}\right)^2 - \frac{GM}{r} = E \tag{8.8}$$

here E is a constant of integration. The sign of E determines whether given mass shell will expand forever or eventually decouple from the pansion and collapse. If $E > 0$, it follows from (8.8) that $\dot{r}^2$ will never come zero; the shell will expand for ever. On the other hand, if $E < 0$ en as r increases $\dot{r}$ will eventually become zero and later negative, plying a contraction and collapse.

This condition for the collapse of an overdense region can be expressed a more convenient form. To do this, let us consider the terms in (8.8) the initial instant $t = t_i$. It is convenient to choose t_i to be the time at hich δ is quite small so that the overdense region was expanding along th the background. That is, we shall assume that the peculiar velocities v_i are negligible at $t = t_i$ (a more general case is studied in exercise 8.1). hen, $\dot{r}_i = (\dot{a}/a)r_i = H(t_i)r_i \equiv H_i r_i$ at time t_i, and the initial kinetic nergy will be

$$K_i \equiv \left(\frac{\dot{r}^2}{2}\right)_{t=t_i} = \frac{H_i^2 r_i^2}{2}. \tag{8.9}$$

The potential energy at $t = t_i$ is $U = -|U|$ where

$$|U| = \left(\frac{GM}{r}\right)_{t=t_i} = G\frac{4\pi}{3}\rho_b(t_i)r_i^2(1+\bar{\delta}_i) = \frac{1}{2}H_i^2 r_i^2 \Omega_i(1+\bar{\delta}_i) \\ = K_i\Omega_i(1+\bar{\delta}_i) \tag{8.10}$$

with $\Omega_i = (\rho_b(t_i)/\rho_c(t_i))$ denoting the *initial* value of the density parameter Ω of the smooth background universe. The total energy of the shell is, therefore,

$$E = K_i - K_i\Omega_i(1+\bar{\delta}_i) = K_i\Omega_i[\Omega_i^{-1} - (1+\bar{\delta}_i)]. \tag{8.11}$$

The condition $E < 0$ for the shell to collapse (eventually), becomes $(1+\bar{\delta}_i) > \Omega_i^{-1}$, or

$$\bar{\delta}_i > [\Omega_i^{-1} - 1]. \tag{8.12}$$

In a closed or flat universe (with $\Omega_i^{-1} \leq 1$), this condition is satisfied by any overdense region with $\bar{\delta} > 0$. In this case, the overdense regions will always collapse although (as we will see) smaller overdensities will take longer times to turn-around and collapse. In an open universe with $\Omega_i < 1$, the overdensity has to be above a critical value for collapse to occur. For a general density distribution $\delta_i(r)$, only shells within a critical initial radius r_{cr}, such that $\bar{\delta}_i(r_{cr}) = \Omega_i^{-1} - 1$, will be able to collapse.

Let us now consider a shell with $E < 0$, which expands to a maximum radius r_m and then collapses. The maximum radius r_m which such a shell attains can be easily derived. To do this, note that at the instant of maximum expansion, we have $\dot{r} = 0$ giving

$$E = -GM/r_m = -(r_i/r_m)K_i\Omega_i(1+\bar{\delta}_i). \tag{8.13}$$

Equating this expression for E with the one in (8.11), we get

$$\frac{r_m}{r_i} = \frac{(1+\bar{\delta}_i)}{\bar{\delta}_i - (\Omega_i^{-1}-1)}. \tag{8.14}$$

Clearly, $r_m \gg r_i$ if $\bar{\delta}_i \gtrsim (\Omega_i^{-1}-1)$; shells which are only slightly overdense, compared to the critical value $(\Omega_i^{-1}-1)$, will expand much further and *can take a long time to collapse*.

The time evolution of the shell can be found by integrating the equations of motion. The solution to equation (8.8), for $E < 0$, is given in a *parametric form* by

$$r = A(1-\cos\theta), \quad t+T = B(\theta - \sin\theta); \quad A^3 = GMB^2 \tag{8.15}$$

where A and B are constants related to each other as shown. The parameter θ increases with increasing t, while r increases to a maximum value before decreasing to zero. The constant T allows us to set the

al condition that at $t = t_i$, $r = r_i$. A shell enclosing mass M and ally expanding with the background universe will progressively slow n, reach a maximum radius at $\theta = \pi$, 'turn-around' and collapse. The ch of maximum radius is also referred to as the epoch of 'turn-around'. the 'turn-around', $dr/dt = 0$ and $r = r_m$.

The constants A and B can be determined by using (8.14). At $\theta = \pi$, $) = r_m = 2A$; comparing with (8.14), we get

$$A = \frac{r_i}{2}\frac{(1+\bar{\delta}_i)}{[\bar{\delta}_i - (\Omega_i^{-1}-1)]}. \tag{8.16}$$

ng $A^3 = GMB^2$, and the expression for M from (8.7) we find B to

$$B = \frac{1+\bar{\delta}_i}{2H_i\Omega_i^{1/2}[\bar{\delta}_i - (\Omega_i^{-1}-1)]^{3/2}}. \tag{8.17}$$

e value of T can be fixed by setting $r = r_i$ at $t = t_i$. As an example, sider the case in which the background universe is flat ($\Omega_i = 1$). Then

$$A = \frac{r_i}{2}\left(\frac{1+\bar{\delta}_i}{\bar{\delta}_i}\right); \quad B = \frac{1}{2H_i}\frac{(1+\bar{\delta}_i)}{\bar{\delta}_i^{3/2}}. \tag{8.18}$$

$t = t_i$ we have to satisfy the conditions

$$r_i = \frac{r_i}{2}\left(\frac{1+\bar{\delta}_i}{\bar{\delta}_i}\right)(1-\cos\theta_i) \tag{8.19}$$

$$t_i + T = \frac{1}{2H_i}\left(\frac{1+\bar{\delta}_i}{\bar{\delta}_i^{3/2}}\right)(\theta_i - \sin\theta_i). \tag{8.20}$$

om (8.19), we get, $\cos\theta_i = (1-\bar{\delta}_i)\,(1+\bar{\delta}_i)^{-1}$. Since δ_i is expected be quite small, we can approximate this relation as $\cos\theta_i \simeq 1 - 2\delta_i$, taining $\theta_i^2 = 4\delta_i$. Substituting in (8.20), we get

$$H_i(t_i + T) = \frac{2}{3}(1+\delta_i). \tag{8.21}$$

r, since $H_i t_i = (2/3)$ for the $\Omega = 1$ universe, $H_i T = (2/3)\delta_i$. This shows at $(T/t_i) = \delta_i \ll 1$. Hence, we will ignore T in what follows. (Similar nclusions hold for models with $\Omega_i \neq 1$, as long as $\delta_i \ll 1$.) The equation (8.15), with the constants A and B fixed by (8.16) and (8.17), give complete information about how each perturbed mass shell evolves. hese equations can be used to work out all the characteristics of a spherical perturbation.

Consider, for example, the evolution of mean density within each mass ell. Since M is constant for each mass shell, the mean density within

a shell is

$$\overline{\rho}(t) = (3M/4\pi r^3) = \frac{3M}{4\pi A^3(1-\cos\theta)^3}. \tag{8.22}$$

In the special case in which the initial density enhancement is homogeneous, the average density calculated above is also the actual density. The density profile of such a constant density sphere is often referred to as the 'top-hat' profile. To work out the time evolution of the density contrast $\overline{\delta}(r,t)$, one also needs to know how the background density evolves. In the simplest case of a flat universe with $k=0$, the expansion factor $a(t)$ and density $\rho_b(t)$ of the background are given by:

$$a \propto t^{2/3}; \quad \rho_b(t) = \frac{1}{6\pi G t^2}. \tag{8.23}$$

Dividing the mean density $\overline{\rho}(r,t)$ in equation (8.22) by the background density, we get the mean density contrast:

$$\frac{\overline{\rho}(r,t)}{\rho_b(t)} = 1 + \overline{\delta}(r,t) = \frac{3M}{4\pi A^3}\,\frac{6\pi G B^2(\theta-\sin\theta)^2}{(1-\cos\theta)^2}, \tag{8.24}$$

where we have used the relation between t and θ given in equation (8.15) and set $T=0$. Since $A^3 = GMB^2$ it follows that

$$\overline{\delta} = \frac{9}{2}\frac{(\theta-\sin\theta)^2}{(1-\cos\theta)^3} - 1. \tag{8.25}$$

The linear evolution for the average density contrast is recovered in the limit of small t. In this limit, we have

$$\overline{\delta} \approx \frac{3\theta^2}{20}; \quad t \approx \frac{B\theta^3}{6} \tag{8.26}$$

so that

$$\overline{\delta} = \frac{3}{20}\left(\frac{6t}{B}\right)^{2/3}. \tag{8.27}$$

For a flat universe with $\Omega_i = 1$ and $H_i = 2/(3t_i)$

$$B = \frac{3}{4}\frac{t_i}{\overline{\delta}_i^{3/2}}(1+\overline{\delta}_i). \tag{8.28}$$

Using this value for B in (8.27) we find, to the leading order,

$$\overline{\delta} = \frac{3}{5}\overline{\delta}_i\left(\frac{t}{t_i}\right)^{2/3} \propto a(t). \tag{8.29}$$

This is the correct growth law ($\delta \propto t^{2/3}$) for the purely growing mode in the linear regime if the initial peculiar velocity is zero; the origin of the factor (3/5) is discussed in exercise 8.1.

For the $\Omega_i = 1$ model, A and B are given by (8.18). Assuming that $\bar{\delta}_i$ is small compared with unity, and retaining only the leading terms of $\bar{\delta}_i$ in A and B, we can write:

$$A \cong \frac{r_i}{2\bar{\delta}_i}; \quad B \cong \frac{3t_i}{4\bar{\delta}_i^{3/2}}. \tag{8.30}$$

For further discussion, it is convenient to use two other variables x and $\bar{\delta}_0$ in place of r_i and δ_i. The quantity x is the comoving radius: $x = r_i[a(t_0)/a(t_i)]$ corresponding to r_i; the parameter $\bar{\delta}_0$ is defined as: $\bar{\delta}_0 = [a(t_0)/a(t_i))\ (3\delta_i/5) = (3/5)\delta_i(1+z_i)$. This is the present value of the density contrast, as predicted by the linear theory, if the density contrast was δ_i at the redshift z_i. In terms of x and δ_0, we have:

$$A = \frac{3x}{10\bar{\delta}_0}; \quad B = \left(\frac{3}{5}\right)^{3/2} \frac{3t_0}{4\bar{\delta}_0^{3/2}}. \tag{8.31}$$

Hereafter we will omit the overbar on $\bar{\delta}$ when no confusion can arise. Collecting all our results together, the evolution of a spherical overdense region can be summarized by the following equations:

$$r(t) = \frac{r_i}{2\bar{\delta}_i}(1-\cos\theta) = \frac{3x}{10\bar{\delta}_0}(1-\cos\theta), \tag{8.32}$$

$$t = \frac{3t_i}{4\bar{\delta}_i^{3/2}}(\theta-\sin\theta) = \left(\frac{3}{5}\right)^{3/2} \frac{3t_0}{4\bar{\delta}_0^{3/2}}(\theta-\sin\theta), \tag{8.33}$$

$$\bar{\rho}(t) = \rho_b(t)\frac{9(\theta-\sin\theta)^2}{2(1-\cos\theta)^3}, \tag{8.34}$$

The density can be expressed in terms of the redshift by using the relation $(t/t_i)^{2/3} = (1+z_i)(1+z)^{-1}$. This gives

$$(1+z) = \left(\frac{4}{3}\right)^{2/3} \frac{\delta_i(1+z_i)}{(\theta-\sin\theta)^{2/3}} = \left(\frac{5}{3}\right)\left(\frac{4}{3}\right)^{2/3} \frac{\delta_0}{(\theta-\sin\theta)^{2/3}}; \tag{8.35}$$

$$\delta = \frac{9}{2}\frac{(\theta-\sin\theta)^2}{(1-\cos\theta)^3} - 1. \tag{8.36}$$

Given an initial density contrast δ_i at redshift z_i, these equations define (implicitly) the function $\delta(z)$ for $z > z_i$. Equation (8.35) defines θ in terms of z (implicitly); equation (8.36) gives the density contrast at that z. For comparison note that linear evolution gives the density contrast where

$$\delta_L = \frac{\bar{\rho}_L}{\rho_b} - 1 = \frac{3}{5}\frac{\delta_i(1+z_i)}{1+z} = \frac{3}{5}\left(\frac{3}{4}\right)^{2/3}(\theta-\sin\theta)^{2/3}. \tag{8.37}$$

We can estimate the accuracy of the linear theory by comparing $\delta(z)$ and $\delta_L(z)$. To begin with, for $z \gg 1$, we have $\theta \ll 1$ and we get $\delta(z) \simeq \delta_L(z)$. When $\theta = (\pi/2)$, $\delta_L = (3/5)(3/4)^{2/3}(\pi/2 - 1)^{2/3} = 0.341$ while $\delta = (9/2)(\pi/2-1)^2 - 1 = 0.466$; thus the actual density contrast is about 40 per cent higher. When $\theta = (2\pi/3)$, $\delta_L = 0.568$ and $\delta = 1.01 \simeq 1$. If we interpret $\delta = 1$ as the transition point to nonlinearity, then such a transition occurs at $\theta = (2\pi/3)$, $\delta_L \simeq 0.57$. From (8.35), we see that this occurs at the redshift $(1 + z_{nl}) = 1.06\delta_i(1 + z_i) = (\delta_0/0.57)$. The next important stage occurs at $\theta = \pi$ when the spherical region reaches the maximum radius of expansion. From our equations, we find that the redshift z_m, the proper radius of the shell r_m and the average density contrast δ_m at 'turn-around' are:

$$(1 + z_m) = \frac{\delta_i(1 + z_i)}{\pi^{2/3}(3/4)^{2/3}} = 0.57(1 + z_i)\delta_i = \frac{5}{3}\frac{\delta_0}{(3\pi/4)^{2/3}} \simeq \frac{\delta_0}{1.062},$$

$$r_m = \frac{3x}{5\delta_0},$$

$$\left(\frac{\bar{\rho}}{\rho_b}\right)_m = 1 + \bar{\delta}_m = \frac{9\pi^2}{16} \approx 5.6. \tag{8.38}$$

The first equation gives the redshift at turn-around for a region, parametrized by the (hypothetical) linear density contrast δ_0 at the present epoch. If, for example, $\delta_i \simeq 10^{-3}$ at $z_i \simeq 10^4$, such a perturbation would have turned around at $(1 + z_m) \simeq 5.7$ or when $z_m \simeq 4.7$. The second equation gives the maximum radius reached by the perturbation. The third equation shows that the region under consideration is nearly 6 times denser than the background universe, at turn-around. This corresponds to a density contrast of $\delta_m \approx 4.6$ which is definitely in the nonlinear regime. The linear evolution gives $\delta_L \simeq 1.063$ at $\theta = \pi$.

After the spherical overdense region turns around it will continue to contract. Equation (8.34) suggests that at $\theta = 2\pi$ all the mass will collapse to a point. However, long before this happens, the approximation that matter is distributed in spherical shells and that random velocities of the particles are small, will break down. The collisionless component of density, viz. the dark matter, will reach virial equilibrium by a process known as *'violent relaxation'. This process arises as follows*[2]*: During the* collapse there will be large fluctuations in the gravitational potential, *on a time scale of the order of the free-fall collapse time,* $t_{dyn} \simeq (G\rho)^{-1/2}$. Since the potential is changing with time, individual particles do not follow orbits which conserve the energy. Clearly, the change in the energy of a particle depends in a complex way on its initial position and velocity but the net effect will be to widen the range of energies available to the particles. Thus, a potential varying in time can provide a relaxation

mechanism for the particles which operates in a timescale $t_{\rm dyn}$ which is much smaller than the two-body-relaxation time t_R. This process has been termed 'violent relaxation' (for more details, see exercise 8.2).

The above process will relax the collisionless (dark matter) component to a configuration with radius $r_{\rm vir}$, velocity dispersion v and density $\rho_{\rm coll}$. (The behaviour of the baryonic component is a little more complicated and we will discuss it separately later.) Such a virialized system can be used to model the structures which we see in the universe. We shall now estimate the physical parameters of such a system.

After virialization of the collapsed shell, the potential energy U and the kinetic energy K will be related by $|U| = 2K$ so that the total energy $\mathcal{E} = U + K = -K$. At $t = t_m$ all the energy was in the form of potential energy. For a spherically symmetric system with constant density, $\mathcal{E} \approx -3GM^2/5r_m$. The 'virial velocity' v and the 'virial radius' $r_{\rm vir}$ for the collapsing mass can be estimated by the equations:

$$K \equiv \frac{Mv^2}{2} = -\mathcal{E} = \frac{3GM^2}{5r_m}; \quad |U| = \frac{3GM^2}{5r_{\rm vir}} = 2K = Mv^2. \tag{8.39}$$

We get:

$$v = (6GM/5r_m)^{1/2}; \quad r_{\rm vir} = r_m/2. \tag{8.40}$$

The time taken for the fluctuation to reach virial equilibrium, $t_{\rm coll}$, is essentially the time corresponding to $\theta = 2\pi$. From equation (8.35), we find that the redshift at collapse, $z_{\rm coll}$, is

$$(1 + z_{\rm coll}) = \frac{\delta_i(1 + z_i)}{(2\pi)^{2/3}(3/4)^{2/3}} = 0.36\delta_i(1 + z_i) = 0.63(1 + z_m) = \frac{\delta_0}{1.686}. \tag{8.41}$$

The density of the collapsed object can also be determined fairly easily. Since $r_{\rm vir} = (r_m/2)$, the mean density of the collapsed object is $\rho_{\rm coll} = 8\rho_m$ where ρ_m is the density of the object at turn-around. Further, $\rho_m \cong 5.6\rho_b(t_m)$ and $\rho_b(t_m) = (1 + z_m)^3\,(1 + z_{\rm coll})^{-3}\rho_b(t_{\rm coll})$. Combining these relations, we get

$$\rho_{\rm coll} \simeq 2^3\rho_m \simeq 44.8\rho_b(t_m) \simeq 170\rho_b(t_{\rm coll}) \simeq 170\rho_0(1 + z_{\rm coll})^3 \tag{8.42}$$

where ρ_0 is the present cosmological density. This result determines $\rho_{\rm coll}$ in terms of the redshift of formation of a bound object (see figure 8.1). (For comparison, it may be noted that linear theory predicts $\delta_L = 1.686$ at $\theta = 2\pi$.) Once the system has virialized, its density and size do not change. Since $\rho_b \propto a^{-3}$, the density contrast δ increases as a^3 for $t > t_{\rm coll}$.

Let us now consider the collapse of the baryonic component, for which a similar result holds. During the collapse, the gaseous mixture of hydrogen and helium develops shocks and gets reheated to a temperature

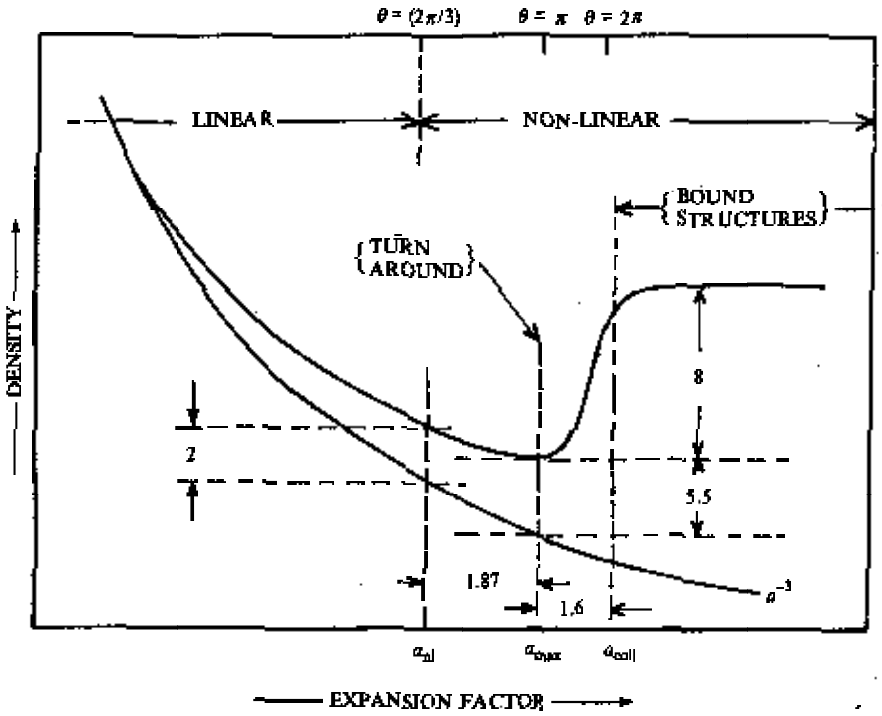

Fig. 8.1. This figure shows the growth of density in a spherical over-dense region. The lower curve shows the evolution of the background density in a matter-dominated case. The upper curve is the density of the spherical inhomogeneity. In the linear region, the contrast grows as a. Once the nonlinear stage is reached, the spherical region collapses faster, virializes and forms a bound structure. The density of the bound structure remains constant thereafter.

at which pressure balance can prevent further collapse. At this stage the thermal energy will be comparable to the gravitational potential energy. The temperature of the gas, $T_{\rm vir}$, is related to the velocity dispersion v^2 by $3\rho_{gas}T_{vir}/2\mu = \rho_{gas}v^2/2$, where ρ_{gas} is the gas density and μ is its mean molecular weight. This gives $T_{vir} = \mu v^2/3$. It is useful to express the above results with typical numbers for the various quantities shown explicitly. If the He fraction is Y by weight and the gas is fully ionized, then

$$\mu = \frac{(m_{\rm H}n_{\rm H} + m_{\rm He}n_{\rm He})}{(2n_{\rm H} + 3n_{\rm He})} = \frac{m_{\rm H}}{2}\left(\frac{1+Y}{1+0.375Y}\right) \cong 0.57m_{\rm H}, \qquad (8.43)$$

if $Y = 0.25$. Apart from the cosmological parameters, two parameters need to be specified. These may be chosen to be the mass M of the over-dense region and the redshift of formation $z_{\rm coll}$. Using the cosmological

arameters

$$\rho_0 = 1.88 \times 10^{-29} \Omega h^2 \,\mathrm{g\,cm^{-3}},$$
$$x = 0.92(\Omega h^2)^{-1/3}(M/10^{12}\,\mathrm{M}_\odot)^{1/3}\,\mathrm{Mpc}, \tag{8.44}$$
$$t_0 = 0.65 \times 10^{10} h^{-1}\,\mathrm{yr}.$$

d $\delta_0 = 1.686(1 + z_{\rm coll})$, we find

$$\begin{aligned} r_{\rm vir} &= 258(1+z_{\rm coll})^{-1}\left(\frac{M}{10^{12}\,\mathrm{M}_\odot}\right)^{1/3} h_{0.5}^{-2/3}\,\mathrm{kpc} \\ &= 434\delta_0^{-1}\, h_{0.5}^{-2/3}\, M_{12}^{1/3}\,\mathrm{kpc}, \\ v &= 100(1+z_{\rm coll})^{1/2}\left(\frac{M}{10^{12}\,\mathrm{M}_\odot}\right)^{1/3} h_{0.5}^{1/3}\,\mathrm{km\,s^{-1}} \\ &= 77\delta_0^{1/2}\, M_{12}^{1/3}\, h_{0.5}^{1/3}\,\mathrm{km\,s^{-1}}, \\ T_{\rm vir} &= 2.32 \times 10^5 (1+z_{\rm coll})\left(\frac{M}{10^{12}\,\mathrm{M}_\odot}\right)^{2/3} h_{0.5}^{2/3}\,\mathrm{K} \\ &= 1.36 \times 10^5 \delta_0 M_{12}^{2/3}\, h_{0.5}^{2/3}\,\mathrm{K}. \end{aligned} \tag{8.45}$$

lso note that

$$t_{\rm coll} = t_0(1+z_{\rm coll})^{-3/2}; \quad (1+z_m) = 1.59(1+z_{\rm coll}). \tag{8.46}$$

hese expressions use $h_{0.5}$, the Hubble constant in units of $50\,\mathrm{km\,s^{-1}\,M}$ $^{-1}$; we have also set $\Omega = 1$. The above results can be used to estimate he typical parameters of collapsed objects once we are given M and the ollapse redshift. For example, if objects with $M = 10^{12}\,\mathrm{M}_\odot$ (which is ypical of galaxies) collapse at a redshift of, say 2, then one gets $r_{\rm vir} \approx$ 6 kpc, $t_{\rm coll} \approx 1.2 \times 10^9$ yr, $v \approx 173\,\mathrm{km\,s^{-1}}$, $T_{\rm vir} \approx 7 \times 10^5$ K. The density ontrast of the galaxy at present will be $(\rho_{\rm coll}/\rho_0) \approx 170(1+z_{\rm coll})^3 \approx$ $.6 \times 10^3$.

These values are broadly in agreement with the parameters which one ssociates with a galactic halo. The linear evolution, studied in chapter combined with the spherical collapse model discussed above, seems to e capable of producing structures of the correct magnitude. The virial adius of the baryonic content of the galaxy will be much smaller because aryons can cool by radiative processes and contract further. This will e discussed in section 8.3.

The equation (8.33) also provides a relation we needed in chapter 5. he time of formation of a bound structure ($t_{\rm coll}$) is related to the density ontrast δ_i at an earlier time t_i by $t_{\rm coll} \propto t_i \delta_i^{-3/2}$. That is, the minimum ensity contrast needed at time t_i for a bound structure to form at $t \le$ $_{\rm coll}$ scales as $(\delta_i)_{\rm min} \propto (t_i/t_{\rm coll})^{2/3}$. This result was used in chapter 5.

The spherical top-hat model can be used to estimate nonlinear density contrast in the following way: We start with some density contrast δ_i at z_i, and compute the density contrast δ_0 at present using the linear theory to be $\delta_0 = (3/5)\delta_i(1+z_i)$. The *actual* density contrast, of course, will be higher and can be calculated as follows: (1) If $\delta_0 < 1.063$, then we can find a $\theta(\delta_0)$ in the range $0 < \theta < \pi$ by inverting the relation (8.37): $\delta_0 = (3/5)(3/4)^{2/3}(\theta - \sin\theta)^{2/3}$. The correct density contrast can now be obtained from (8.36) using this value of $\theta(\delta_0)$. (2) If $\delta_0 > 1.686$, then our analysis shows that a bound structure would have already formed at $(1+z_{\rm coll}) = (\delta_0/1.686)$ with the density $\rho_{\rm coll} \cong 170\rho_0(1+z_{\rm coll})^3 = (170/1.686^3)\rho_0\delta_0^3 \simeq 35.5\rho_0\delta_0^3$. The correct density contrast is, therefore, $\delta = (\rho_{\rm coll}/\rho_0) - 1 = 35.5\delta_0^3 - 1$. (3) For $1.063 < \delta_0 < 1.686$, the spherical collapse model is a bad approximation and cannot be used to make reliable predictions. The actual density contrast increases by two orders of magnitude during this interval.

We end this section by mentioning an extremely simple *general relativistic* solution which describes the evolution of a spherical inhomogeneity in the Friedmann universe (see exercise 8.3). It turns out[3] that such a situation can be described by a metric of the form

$$ds^2 = dt^2 - a^2(x,t)\left[\frac{dx^2}{1-k(x)x^2}\left[\frac{(ax)'}{a}\right]^2 - x^2(d\theta^2 + \sin\theta\, d\phi^2)\right] \quad (8.47)$$

where $a(x,t)$ is a *space dependent* 'expansion factor', and $k(x)$ is a space dependent curvature constant. The metric can be written in the above form as long as mass shells at different values of x do not cross ; a condition which will be satisfied by density distributions in which ρ decreases monotonically with x. The Einstein equations determining the time evolution of the expansion factor $a(x,t)$ and the matter density $\rho(x,t)$ turn out to be:

$$\frac{\dot{a}^2 + k(x)}{a^2} = \frac{C}{a^3} = \frac{8\pi G\rho(x,t)}{3}\frac{(ax)'}{a} \quad (8.48)$$

where C is a constant. Here (8.48) is actually two equations, one giving the evolution of a and the other that of ρ. If ρ, and hence a and k, are independent of x these equations reduce to the standard equations for a Friedmann universe.

This simple generalization of the homogeneous universe model offers considerable insight into the way a spherical overdense (or underdense) region behaves. Equation (8.48) shows that the behaviour of a mass shell at a comoving radius x is completely specified by the *local* value of the curvature constant k. If at some x, $k(x) \leq 0$, the corresponding mass *shell will expand for ever, while if* $k(x) > 0$, *it will turn around at some* stage and collapse.

We are now in a position to understand the evolution of different types of spherical density perturbations that may arise in the Friedmann universe. Consider the case when $k(x)$ is positive for $x < x_0$, is zero at x_0 and tends to a constant negative value, say -1, far away from the origin. One way of realizing such a situation is to embed a density hill centred around the origin in an open Friedmann universe and start off the universe expanding uniformly. From the evolution equation (8.48) we can infer that the region $x < x_0$ will eventually collapse, while the region $x \geq x_0$ will expand for ever. Here we see quite clearly that condensation in a local part of the universe does not alter the global behaviour of an open Friedmann universe. Similarly one can construct expanding voids in a closed universe. In this case, one demands that $k(x) < 0$ for $x < x_0$ (say) and positive elsewhere. This situation can be realized if there is a deep enough density valley in a closed universe. The region within x_0 will keep on expanding, whereas the region outside will initially expand at a slower rate and eventually recollapse.

8.3 Scaling laws

The analysis in the previous section dealt with the nonlinear collapse of a *single* overdense region in an otherwise smooth universe. To model the structure formation correctly, we need to find the full *power spectrum* of bound objects which are formed due to the nonlinear collapse. This is a considerably more difficult task in which only limited success has been achieved.

The mass function of the bound objects can be calculated in a fairly straightforward manner, once a choice is made for the filtering function. (This was discussed in chapter 5.) While this is adequate for some purposes, it does not provide a dynamical picture of the collapse. Somewhat more detailed modelling is possible if the form of the power spectrum at $t = t_{dec}$ is known. The further evolution depends crucially on whether the dark matter is cold or hot. Different kinds of analytic approximations are needed for the two cases. We shall first consider the case of cold dark matter. The approximate analysis of the hot dark matter will be taken up in sections 8.5 and 8.6.

We saw in chapter 5 that the density inhomogeneity can be characterized by a Gaussian distribution with some variance, σ, which is related to the power spectrum of the fluctuations. Labelling the fluctuations by the mass $M \propto \lambda^3 \propto k^{-3}$, we can relate the mean square fluctuation in the mass to the variance of the Gaussian as,

$$\sigma^2(M) = < (\delta M/M)^2 > = CM^{-(3+n)/3} \qquad (8.49)$$

where n is the index of the power spectrum (with $P(k) \propto k^n$) and C is the normalization constant which should be fixed by comparison with the

observations. Since $P(k)$ is not a strict power law, n should be thought of as an approximate local value $d(\ln P)/d(\ln k)$ in the relevant range.

The quantity $\sigma(M)$ was interpreted in chapter 5 as the typical (excess) mass contrast at some scale $R \propto M^{1/3}$. Since $(\delta M/M) \propto [(\delta\rho)R^3/\rho_b R^3] \propto (\delta\rho/\rho_b)$, we can take the quantity $\sigma(M)$ to be proportional to the average density contrast $\overline{\delta}$ inside a region of radius R which was the parameter used in the last section. More generally, we can set $\delta = \nu\sigma$ to describe a spherical region with density contrast which is ν times the standard deviation. Using this expression in (8.49) we can express all the physical quantities in terms of the mass of the overdense region. We then find the following scalings: $t_{\rm coll} \propto \nu^{-3/2} M^{(n+3)/4}$; $\rho \propto \nu^3 M^{-(n+3)/2}$; $r_{\rm vir} \propto r_m \propto \nu^{-1} M^{(n+5)/6}$; $v \propto \nu^{1/2} M^{(1-n)/12}$; $T_{\rm vir} \propto \nu M^{(1-n)/6}$. The same scalings have been obtained from simpler dimensional arguments and the linear theory in chapter 5. But only a detailed model can provide us with the constants of proportionality appearing in these relations.

For $n > -3$, the variance σ decreases with increasing M; then the scaling $t_{\rm coll} \propto M^{(n+3)/4}$ shows that, on the average, smaller masses turn around and collapse earlier than larger masses. Structures grow by the gradual separation and recollapse of progressively larger units. As each unit condenses out, it will in general be made up of a number of smaller condensations which had collapsed earlier. This leads to a heirarchical pattern of clustering.

When a larger mass collapses, its substructure is likely to be erased rapidly by the mergers and tidal disruption of its subunits, provided the specific binding energy $(GM/r_{\rm vir}) \propto v^2$ increases with M. Since $v \propto M^{(1-n)/12}$, this happens for $n < 1$. In this case, the evolution of structure will be self-similar in time with a characteristic mass $M_c(t)$ which grows with time as $M_c(t) \propto t^{4/(n+3)}$. For masses much larger than $M_c(t)$, the fluctuations will still be in the linear regime; on scales comparable to $M_c(t)$ structure will be turning around and collapsing and will show a heirarchical pattern; while on mass scales much smaller than $M_c(t)$, the structure would have been smoothed out by nonlinear relaxation effects.

It should be stressed that the processed spectrum at $t = t_{\rm dec}$ is not a pure power law. So the scaling laws derived above can only be applied piecewise, over mass intervals in which $P(k)$ can be approximated (locally) as a power law. In the cold dark matter models $n \gtrsim -3$ on small M, increases with increasing M and reaches the asymptotic value of $n = 1$ for $M \gtrsim 10^{15}\,{\rm M}_\odot$. The power spectrum on galactic scales can be approximated by $n \approx -2$. In this case, one sees from the relation $v \propto M^{(1-n)/12}$ that $M \propto v^4$. This relation connects the *total* mass of the system with the velocity dispersion in the gravitational potential produced by this mass. If we assume that the total mass is proportional

to the luminosity L of the system and that the velocity dispersion is of the same order as the rotational velocity σ of visible objects (stars, gas, ...etc.) in this potential, it follows that $L \propto \sigma^4$. This was one of the relations used in chapter 7 to estimate the distances to the galaxies.

8.4 The masses of galaxies

Galaxies have typical masses of about $10^{11}\,M_\odot$. Theories for galaxy formation, based purely on the gravitational instability of density fluctuations do not provide any natural explanation for this characteristic mass. It is, therefore, necessary to understand the extra physical considerations which lead to this characteristic mass scale.

To begin with, it should be noted that the part of a galaxy which is directly accessible to observations is the baryonic part, though the gravitationally dominant part may be the dark matter. The dynamics of the baryonic part can be properly described[5] only if the cooling mechanisms in the gas is also taken into account. Consider a gas cloud of mass M and radius R, which is supported against gravitational collapse by gas pressure. To provide this support, the gas should have a temperature T where $T = (\mu v^2/3) = (\mu/3)(6GM/10r_v) \simeq (GM\mu/5R)$ if we identify the virial radius r_v with R. Because of this rather high temperature, the gas will be radiating energy and cooling. Once the temperature changes due to cooling, the delicate balance between gravity and pressure support can be affected. The evolution of such a cloud will depend crucially on the relative values of the cooling timescale,

$$t_{\rm cool} = \frac{E}{\dot{E}} \approx \frac{3\rho kT}{2\mu\Lambda(T)} \tag{8.50}$$

and the dynamical timescale

$$t_{\rm dyn} \approx \frac{\pi}{2}\left[\frac{2GM}{R^3}\right]^{-1/2} = 5\times 10^7\ {\rm yr}\left(\frac{n}{1\,{\rm cm}^{-3}}\right)^{-1/2}. \tag{8.51}$$

Here ρ is the average baryonic density and $\Lambda(T)$ gives the cooling rate of the gas at temperature T. Note that we have taken $t_{\rm dyn}$ to be the freefall time of a uniform density sphere of radius R.

There are three possibilities which should be distinguished as regards the evolution of such a cloud. Firstly, if $t_{\rm cool}$ is greater than the Hubble time, H^{-1}, then the cloud could not have evolved much since its formation. On the other hand, if $H^{-1} > t_{\rm cool} > t_{\rm dyn}$, the gas can cool; but as it cools the cloud can retain the pressure support by adjusting the pressure distribution. In this case the collapse of the cloud will be quasi-static on a timescale of order $t_{\rm cool}$. Finally there is the possibility that $t_{\rm cool} < t_{\rm dyn}$. In this case the cloud will cool rapidly (relative to the dynamical timescale) to a minimum temperature. This will lead to

the loss of pressure support and the gas will undergo an almost freefall collapse. Fragmentation into smaller units can now occur because, as the collapse proceeds isothermally, smaller and smaller mass scales will become gravitationally unstable.

The criterion $t_{\rm cool} < t_{\rm dyn}$ can determine the masses of galaxies. Only when this condition is satisfied can a gravitating gas cloud collapse appreciably and fragment into stars. Further, in any heirarchical theory of galaxy formation, unless a gas cloud cools within a dynamical timescale and becomes appreciably bound, collapse on a larger scale will disrupt it. In these theories, galaxies are the first structures which have resisted such disruption by being able to satisfy the above criterion.

Let us first examine this model without introducing any dark matter. The cooling of primordial gas is mainly due to three processes: bremsstrahlung, recombinations in the hydrogen–helium plasma and Compton scattering of hot electrons by the colder cosmic background photons. As discussed in chapter 6, Compton cooling is important only at redshifts higher than $z \simeq 8$ or so. Since galaxy scales become nonlinear only at $z \lesssim 10$ we can ignore the Compton cooling. The cooling rate of the gas due to bremsstrahlung and recombination can then be written as[6]

$$\Lambda(T) = (A_B T^{1/2} + A_R T^{-1/2})\rho^2 \tag{8.52}$$

where the $A_B \propto (e^6 n^2 T^{1/2}/m_e^{3/2})$ term represents the cooling due to bremsstrahlung and the $A_R \simeq e^4 m A_B$ term arises from the cooling due to recombination. (The temperature dependence of both these processes was discussed in chapter 1.) This expression is valid for temperatures above 10^4 K; for lower temperatures, the cooling rate drops drastically since hydrogen can no longer be significantly ionized by collisions. Introducing the numerical values appropriate for a hydrogen–helium plasma (with a helium abundance $Y = 0.25$ and some admixture of metals) the expression for $t_{\rm cool}$ becomes

$$t_{\rm cool} = 8 \times 10^6\ {\rm yr} \left(\frac{n}{1\ {\rm cm}^{-3}}\right)^{-1} \left[\left(\frac{T}{10^6\ {\rm K}}\right)^{-1/2} + 1.5 f_m \left(\frac{T}{10^6\ {\rm K}}\right)^{-3/2}\right]^{-1} \tag{8.53}$$

Here n is the number density of gas particles and the factor f_m takes into account the possibility that the gas may be enriched with metals; $f_m \approx 1$ when there are no metals and $f_m \approx 30$ for solar abundance of metals. For gas with primordial abundance ($f_m \approx 1$), one can see from (8.53) that there is a transition temperature $T^* \approx 10^6$ K. For $T > T^*$, bremsstrahlung dominates while for $T < T^*$, the line cooling dominates.

Let us now consider the ratio $\tau = (t_{\rm cool}/t_{\rm dyn})$. The condition $\tau = 1$ defines a curve on the ρ–T space, which demarcates the region of param

eter space in which cooling occurs rapidly within a dynamical time, from the region of weak cooling (see figure 8.2).

For $T < T^*$, when line cooling is dominant, we have $t_{\rm cool} \propto (T^{3/2}/\rho)$ and $t_{\rm dyn} \propto \rho^{-1/2}$ giving $\tau \propto (T^{3/2}/\rho^{1/2}) \propto M$; hence the $\tau = 1$ curve will be parallel to the lines of constant mass in the $\rho - T$ plane. Substituting the numbers and using the expression for the cooling time from (8.53) we find that $\tau = 1$ implies $f(T/10^6\,{\rm K})^{3/2}(n/{\rm cm}^{-3})^{-1/2} = 4.28$. Expressing the mass of the cloud as $M = (5RT/G\mu) = 2.1 \times 10^{11}\,{\rm M}_\odot(T/10^6\,{\rm K})^{3/2}(n/{\rm cm}^{-3})^{-1/2}$ we can write

$$\tau = \frac{t_{\rm cool}}{t_{\rm dyn}} \approx \frac{M}{9 \times 10^{11}\,{\rm M}_\odot}, \tag{8.54}$$

if $\mu = 0.57$ and $f = 1$. Thus the criterion for efficient cooling can be satisfied for masses below a critical mass of about $10^{12}\,{\rm M}_\odot$, provided $T < 10^6$ K.

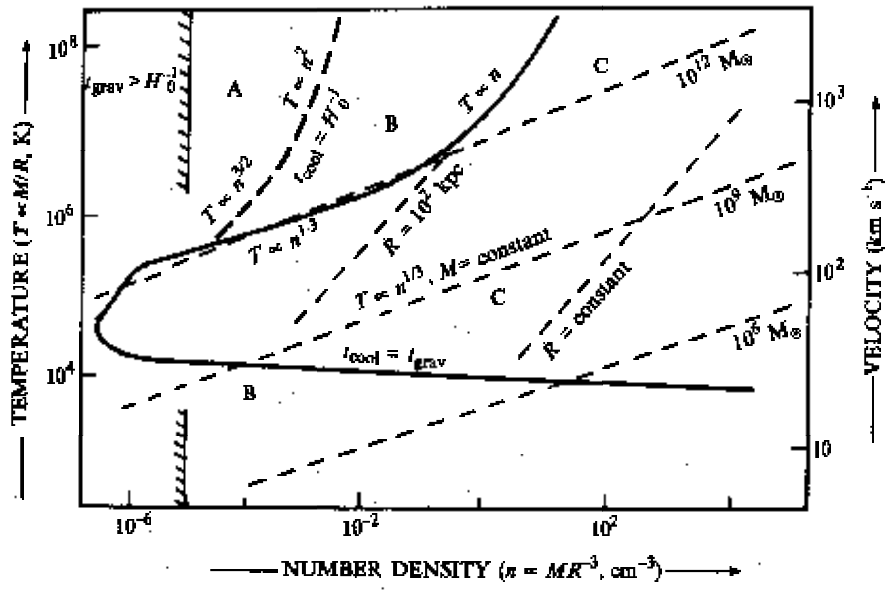

Fig. 8.2. Diagram showing the regions in which various cooling processes can be efficient. The thick line is obtained by equating the timescale for cooling with the free-fall timescale. The top part of the curve is dominated by the bremsstrahlung process while the left end of the curve is contributed by the line cooling of the ionized gas. (The exact shape of the curve depends on the composition of the gas and the amount of metals present.) The thick broken line on the top portion is obtained by equating cooling time to the Hubble time. Similarly the vertical shaded line is obtained by equating free-fall time and with Hubble time. Also marked for reference are the lines of constant mass and constant radius.

On the other hand, for $T > T^*$, when bremsstrahlung dominates the cooling process, $t_{\rm cool} \propto (T^{1/2}/\rho)$ and $t_{\rm dyn} \propto \rho^{-1/2}$. So $\tau \propto (T^{1/2}/\rho^{1/2}) \propto R$, and the curve $\tau = 1$ will be parallel to the lines of constant radius in the ρ–T space. We now find that $\tau = 1$ implies $(T/10^6\,{\rm K})^{1/2}\,(n/{\rm cm}^{-3})^{-1/2} = 6.43$. Expressing the radius of the cloud as $R = (GM\mu/5T) = 13\,{\rm kpc}\,(T/10^6\,{\rm K})^{1/2}(n/{\rm cm}^{-3})^{-1/2}$ we get

$$\tau = \frac{t_{\rm cool}}{t_{\rm dyn}} \cong \frac{R}{80\,{\rm kpc}}. \tag{8.55}$$

Therefore, clouds with high temperature ($T > T^*$) have to shrink below a critical radius of about 10^2 kpc before being able to cool efficiently to form galaxies.

These features are illustrated schematically in figure 8.2 which is usually called a 'cooling diagram'. The ρ–T space is divided into three regimes A, B and C. A gas cloud with constant mass evolves roughly along lines of constant M_J, with $T \propto \rho^{1/3}$, if it is pressure supported. Gas clouds in region A have $t_{\rm cool} > t_{\rm Hubble}$ and cool very little. Those in region B cool slowly and undergo quasi-static collapse, with the pressure balancing gravity at each instant, until they enter the region C where $\tau < 1$. Gas clouds in C can cool efficiently to form galaxies because they have masses below $10^{12}\,{\rm M}_\odot$ or radius below 10^2 kpc. These masses and radii compare well with the scales characteristic of galaxies.

Let us now consider the effects of including the dark matter component. The dynamical timescale is now determined by the total density of dark matter and baryons, whereas the cooling time still depends only on the density of the baryonic gas. In this case, the gas will not be at the virial temperature initially. It is only during collapse that the gas gets heated up by shocks produced when different bits of gas run into each other. If the cooling timescale of the shocked gas is larger than the dynamical timescale in which the cloud settles down to an equilibrium, then the gas will eventually get heated up to the virial temperature. On the other hand, if the cooling time was shorter, the gas may *never* reach such a pressure supported equilibrium. Efficient cooling will result in the gas sinking to the centre of the dark matter potential well which is being formed, until halted by rotation or fragmentation into stars.

Clearly, it is again the ratio of the cooling time to the dynamical time of the object which governs the evolution. Further, notice that smaller mass clumps are disrupted as larger masses turn around and collapse. However, if the gas component can cool efficiently enough, it may shrink sufficiently close to the centre of the dark matter potential and thus resist further disruption. This process will break the hierarchy. Galaxies could be, again, thought of as the first structures that have survived the disruption due to hierarchical clustering.

The spherical model can be used to estimate the relevant dynamical timescale. We assume $t_{\rm dyn}$ to be comparable to $(t_{\rm coll}/2)$, the time taken for a spherical top-hat fluctuation to collapse after turning around. This expression is the same as the $t_{\rm dyn}$ given in (8.51) above, provided we identify R in (8.51) with the radius of turn-around r_m. Then

$$t_{\rm dyn} \approx \frac{t_{\rm coll}}{2} \approx 1.5 \times 10^9 \left(\frac{M}{10^{12}\,{\rm M}_\odot}\right)^{-1/2} \left(\frac{r_m}{200\,{\rm kpc}}\right)^{3/2} {\rm yr}. \qquad (8.56)$$

For estimating the cooling timescale, we use (8.53) and assume that the gas makes up a fraction F of the total mass and is uniformly distributed within a radius $r_m/2$. The gas temperature is taken to be of order the virial temperature obtained in the spherical model; that is, $T_{\rm vir} \simeq (\mu v^2/3)$, where $v^2 \simeq (6GM/5r_m)$. This corresponds to the temperature achieved by heating by shocks which have a velocity of order of the virial velocity. In that case,

$$t_{\rm cool} \approx 2.4 \times 10^9 f_m^{-1} \left(\frac{F}{0.1}\right)^{-1} \left(\frac{M}{10^{12}\,{\rm M}_\odot}\right)^{1/2} \left(\frac{r_m}{200\,{\rm kpc}}\right)^{3/2} {\rm yr}. \qquad (8.57)$$

We have assumed that the line cooling dominates at the temperature $T = T_{\rm vir}$ relevant to the galaxies, and adopted a typical value of $F \simeq 0.1$. Note that the collapse, in general, is likely to be highly inhomogeneous and the above estimates are only supposed to give a rough idea of the numbers involved. From the last two equations we get

$$\tau = \left(\frac{t_{\rm cool}}{t_{\rm dyn}}\right) \approx 1.6 f_m^{-1} \left(\frac{F}{0.1}\right)^{-1} \left(\frac{M}{10^{12}\,{\rm M}_\odot}\right) \qquad (8.58)$$

so that efficient cooling (with $\tau < 1$) requires

$$M < M_{\rm crit} \approx 6.4 \times 10^{11}\,{\rm M}_\odot\, f_m \left(\frac{F}{0.1}\right). \qquad (8.59)$$

It is clear that masses of order of galactic masses are again picked out preferentially even when the dark matter is included.

The procedure outlined above can be used to analyze any particular theory of structure formation involving heirarchical clustering. The starting point will be the cooling diagram, in which the $\tau = 1$ curve is plotted. Given the power spectrum of density fluctuations, one can work out density contrast at various scales $\delta_0 = \nu\sigma(M)$. Then the various properties, like ρ and T of the collapsed objects which are formed, can be estimated using the spherical model. We saw that these properties depend only on one parameter M, once the density contrast δ_0 is fixed. Thus, for each value of ν one gets a curve on the ρ–T plane, giving the properties of collapsed objects. These curves assume that the proto-condensations have virialized, but that the gas has not cooled and condensed. Cooling moves

points on these curves to higher densities. In the same diagram one can also plot, for comparison, the observed positions of galaxies, groups and clusters of galaxies.

A simplified form of such a cooling diagram, for a particular version of cold dark matter theory[7], is given in figure 8.3. This figure suggests that, while galaxies show evidence for having cooled and condensed within their dark halos, groups and clusters of galaxies have too long a cooling time to have dissipated much of their energy. From the diagram one can also see that gas clouds with mass in the range of $10^8\,M_\odot < M < 10^{12}\,M_\odot$

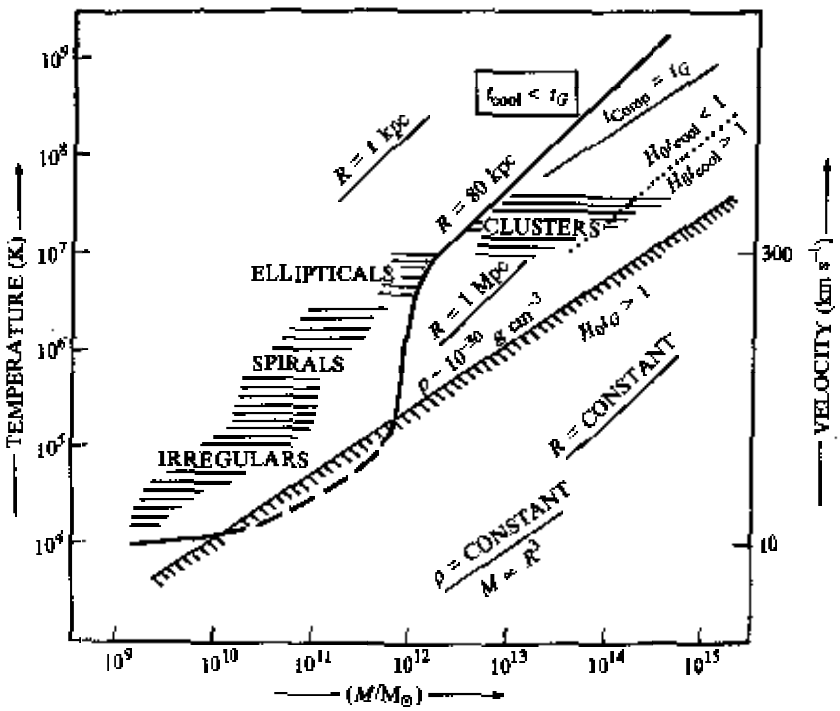

Fig. 8.3. The cooling curve of the previous figure is redrawn using different variables. The y-axis corresponds to the gravitational potential energy per particle and the x-axis, the mass. (Those were the variables used in chapter 1). The slanted, shaded, line corresponds to the line $H_0 t_G = 1$. Objects to the right of (and below) this line cannot collapse within the age of the universe. The thick line in the middle is the cooling curve obtained by $t_{cool} = t_G$. Objects to the left of (and above) this curve can cool efficiently while the objects to the right of this curve cannot. The dotted curve on the right delineates the regions which can cool efficiently within the Hubble time. Also shown is a thin straight line obtained by equating the Compton cooling time and the free-fall time. Typical locations of clusters and different kinds of galaxies are indicated. Also shown are the lines of constant radii and constant density.

can cool within the dynamical timescale. The lower limit comes from the fact that the cooling rate drops drastically below about 10^4 K, when hydrogen can no longer be significantly ionized by collisions.

Some complicating features which affect the above simple ideas deserve mention. Firstly, note that we have ignored the formation of stars and the feedback of this process on the gas. If the star formation is very efficient, the supernovae from the massive stars may provide an important heat input. It may even drive out the gas if the potential well is shallow enough[8]. In fact, such effects may be crucial in preventing all the baryons from being locked up in small objects, before typical galaxies form. Also, we see from (8.58) that if the gas were enriched with metals, much larger masses can cool within a dynamical time because of the increased cooling rate. So the chemical history of the gas could be important.

Finally, we discuss the effect of Compton cooling, which has been ignored so far. The cooling rate of a gas with electron density n_e and temperature T embedded in a blackbody radiation field of density ρ_R and temperature T_γ is given by (see chapter 6)

$$\Lambda_{\rm Comp} = \frac{4\sigma_T n_e \rho_R (T - T_\gamma)}{m_e}. \tag{8.60}$$

The cooling time for matter, due to inverse Compton scattering off the cosmic background photons, will be therefore,

$$t_{\rm Comp} = \frac{3 m_p m_e (1+z)^{-4}}{8\mu\sigma_T \Omega_R \rho_c} \approx 2.1 \times 10^{12} (1+z)^{-4}\,{\rm yr}. \tag{8.61}$$

Here we have assumed that $T \gg T_\gamma$ and used $\rho_R(z) = \Omega_R \rho_c (1+z)^4$ to take into account the expansion of the universe. Comparing $t_{\rm Comp}$ with the dynamical time in (8.56) we get

$$\tau_{\rm Comp} = \frac{t_{\rm Comp}}{t_{\rm dyn}} \approx 2 \times 10^2 (1+z_{\rm coll})^{-5/2}. \tag{8.62}$$

This ratio is less than unity for $z_{\rm coll} > 7$, independent of the mass of the collapsing object. So Compton cooling can efficiently cool an object only if it collapses at a redshift higher than $z \simeq 10$, whatever its mass. It is not clear whether galaxies can collapse that early; but if they do, then galaxy masses cannot be preferentially picked out through the cooling processes outlined above. An interesting feature emerges if one plots the the line $\tau_{\rm Comp} = 1$ in the cooling diagram (see figure 8.3). Note that this line is parallel to the lines of constant density. Galaxies and clusters are neatly separated by the Compton cooling line, suggesting that galaxy formation ceased when Compton cooling became inefficient. If galaxies could form earlier than a redshift of $z \approx 10$, then the above fact could provide a reason for their characteristic masses, which is quite different from the one discussed earlier.

8.5 Zeldovich approximation

The analysis so far was based on the assumption of spherical symmetry and used the power spectrum of the cold dark matter models in which smaller masses become nonlinear first. The evolution, however, will be different if the power spectrum is that produced in a universe dominated by hot dark matter. It was shown in chapter 4 that the power spectrum for hot dark matter is peaked at a mass scale of about $M_{FS} \approx 10^{14}\,M_\odot$. Therefore, the first structures which form due to nonlinearity will all have masses around this value. There will be very little power on small scales. To analyze this scenario, it is preferable to use a different kind of approximation.

Such an approximation, proposed by Zeldovich[9], is possible for scales which are much smaller than d_H where Newtonian analysis is possible. The starting point of the Zeldovich approximation is the result from the linear theory for the growth of small perturbations, expressed as a relation between the Eulerian and Lagrangian co-ordinates of the particles. In a smooth universe with uniform density $\rho_b(t)$, the actual position of any particle $\mathbf{r}(t)$, is related to its initial (Lagrangian) location $\mathbf{q}$ by

$$\mathbf{r}(t) = a(t)\mathbf{q}. \tag{8.63}$$

This result, of course, is altered in the presence of growing density perturbations. In the linear regime, the only modification needed is the addition of a separable function of t and $\mathbf{q}$ of the form, say, $f(t)\mathbf{p}(\mathbf{q}) = a(t)b(t)\mathbf{p}(\mathbf{q})$. That is, we can take

$$\mathbf{r}(t) \equiv a(t)\mathbf{x}(t) = a(t)[\mathbf{q} + b(t)\mathbf{p}(\mathbf{q})] \tag{8.64}$$

where $\mathbf{x}(t)$ is the comoving Eulerian coordinate. This equation gives the comoving position ($\mathbf{x}$) and proper position ($\mathbf{r}$) of a particle at time t given that at some time in the past it had the comoving position $\mathbf{q}$. To demonstrate that this equation correctly describes the linear evolution, let us calculate how the perturbed density evolves when the individual particles move according to (8.64). If the initial, unperturbed, density is $\bar{\rho}$ (which is independent of $\mathbf{q}$), then the conservation of mass implies that the perturbed density will be

$$\rho(\mathbf{r}, t)\, d^3\mathbf{r} = \bar{\rho}\, d^3\mathbf{q}. \tag{8.65}$$

Therefore

$$\rho(r, t) = \bar{\rho}\det(\partial q_i/\partial r_j) = \frac{\bar{\rho}/a^3}{\det(\partial x_j/\partial q_i)} = \frac{\rho_b(t)}{\det(\delta_{ij} + b(t)(\partial p_j/\partial q_i))} \tag{8.66}$$

where we have set $\rho_b(t) = [\bar{\rho}/a^3(t)]$. Expanding the Jacobian to the first order in the perturbation $b(t)\mathbf{p}(\mathbf{q})$, we get

$$\frac{\delta\rho}{\rho} = \frac{(\rho - \rho_b)}{\rho_b} = -b(t)\nabla_{\mathbf{q}} \cdot \mathbf{p}. \tag{8.67}$$

On the other hand, the linear theory predicts that

$$\frac{\delta\rho}{\rho}(\mathbf{x}, t) = g(t)\delta_i(\mathbf{x}) = g(t)\sum_{\mathbf{k}} A_{\mathbf{k}} \exp\{i\mathbf{k} \cdot [\mathbf{q} + b(t_i)\mathbf{p}(\mathbf{q})]\} \tag{8.68}$$

where $g(t)$ is the function describing the time evolution of the growing mode of the density contrast and $A_{\mathbf{k}}$ is the Fourier transform of the initial density contrast, δ_i. For the $\Omega = 1$ universe, for example, $g(t) = (2/5)(t/t_i)^{2/3}$. Let us choose the initial moment t_i such that the term $b(t_i)\mathbf{p}$ is negligible compared to $\mathbf{q}$. Then, if we identify $b(t)$ with $g(t)$ and set

$$\mathbf{p}(\mathbf{q}) = \sum_{\mathbf{k}} \frac{i\mathbf{k}}{k^2} A_{\mathbf{k}} \exp(i\mathbf{k} \cdot \mathbf{q}), \tag{8.69}$$

we see that the approximation given by the relation (8.64) correctly reproduces the result of the linear theory for the growth of small density perturbations. Thus, the relation (8.64) is definitely correct in the linear approximation.

From the definition of $\mathbf{p}(\mathbf{q})$ given in (8.69), one has

$$\mathbf{p}(\mathbf{q}) = \nabla_{\mathbf{q}}\Phi_0(\mathbf{q}) \tag{8.70}$$

where

$$\Phi_0(\mathbf{q}) = \sum_{\mathbf{k}} \frac{A_{\mathbf{k}} \exp(i\mathbf{k} \cdot \mathbf{q})}{k^2}. \tag{8.71}$$

This relation allows one to obtain a simple physical interpretation for $\Phi_0(\mathbf{q})$ and $\mathbf{p}(\mathbf{q})$. Note that

$$\nabla_{\mathbf{q}} \cdot \mathbf{p} = \nabla^2\Phi_0 = \sum_{\mathbf{k}} A_{\mathbf{k}} \exp(i\mathbf{k} \cdot \mathbf{q}) = -\frac{(\rho - \rho_b)}{b\rho_b}. \tag{8.72}$$

Using the Einstein equation $\ddot{a} = -(4\pi G\rho_b a)/3$, we can write this equation

$$\nabla_{\mathbf{q}}^2\Phi_0 = \frac{4\pi G a^2(\rho - \rho_b)}{(3ab\ddot{a})}. \tag{8.73}$$

The equation for the gravitational potential ϕ in the perturbed universe

$$\nabla_{\mathbf{x}}^2\phi = 4\pi G a^2(\rho - \rho_b). \tag{8.74}$$

Comparing these two expressions at an early epoch (say, at $t = t_i$) when $\mathbf{x}$ is very nearly equal to $\mathbf{q}$, we get:

$$\phi = 3ab\dot{a}\Phi_0. \tag{8.75}$$

Thus Φ_0 is proportional to the gravitational potential of the linear theory and $\mathbf{p}(\mathbf{q})$ is proportional to the peculiar velocity field of the linear theory.

Zeldovich suggested that while (8.64) is in accordance with the linear theory, it may also provide a good approximate description of the evolution of density perturbations in the *nonlinear regime* where $\delta\rho/\rho$ greatly exceeds unity. To study the consequences of this hypothesis, it is best to proceed as follows: Since $\mathbf{p}(\mathbf{q})$ is a gradient of a scalar function, the Jacobian in (8.66) is a real symmetric matrix. This matrix can be diagonalized at every point $\mathbf{q}$, to yield a set of eigenvalues and principal axes as a function of $\mathbf{q}$. If the eigenvalues of $(\partial p_j/\partial q_i)$ are $[-\lambda_1(\mathbf{q}), -\lambda_2(\mathbf{q}), -\lambda_3(\mathbf{q})]$ then the perturbed density is given by

$$\rho(\mathbf{r}, t) = \frac{\rho_b(t)}{(1 - b(t)\lambda_1(\mathbf{q}))(1 - b(t)\lambda_2(\mathbf{q}))(1 - b(t)\lambda_3(\mathbf{q}))} \tag{8.76}$$

where $\mathbf{q}$ can be expressed as a function of $\mathbf{r}$ by solving (8.64). This expression describes the effect of deformation of an infinitesimal, cubical volume (with the faces of the cube determined by the eigenvectors corresponding to λ_n) and the consequent change in the density. For a growing perturbation, $b(t)$ increases with time; therefore, a positive λ denotes collapse and negative λ signals expansion. In an overdense region the density will become infinite if one of the terms in brackets in the denominator of (8.76) becomes zero. In the generic case, these eigenvalues will be different from each other; let $\lambda_1 \geq \lambda_2 \geq \lambda_3$. At any particular value of $\mathbf{q}$ one of them, say λ_1, will be maximum. Then the density will diverge for the first time when $(1 - b(t)\lambda_1) = 0$; at this instant the material contained in a cube in the $\mathbf{q}$ space gets compressed to a sheet in the $\mathbf{r}$ space, along the principal axis corresponding to λ_1. Thus sheetlike structures, or 'pancakes', will be the first nonlinear structures to form when gravitational instability amplifies density perturbations.

Notice that the description uses trajectories which are built out of the linear theory. Such a description, of course, cannot be exact. To understand the nature of the approximation, we may proceed as follows: Given the acceleration field $\ddot{\mathbf{x}}$, we can compute the density distribution ρ as

$$\nabla_{\mathbf{x}} \cdot \ddot{\mathbf{x}} = -4\pi G\rho, \tag{8.77}$$

provided the potential producing the acceleration ($\ddot{\mathbf{x}} = -\nabla\phi_{acc}$) is the same as the potential generated by the density ($\nabla^2\phi_N = 4\pi G\rho$). For the *exact* trajectories, equation (8.77) will be an identity; for the approximate