5. Galaxy Formation

The key question for structure (galaxy) formation is how does the tiny fluctuations in the early universe (of the order of $10^{-5}$, as inferred from cosmic microwave background radiation, see Fig. 1) grow to the enormous amplitudes we see today in galaxies (e.g., $2.5 \times 10^6$ for the solar neighbourhood; see Problem set 5).

![Cosmic microwave background (CMB) radiation as observed by the Wilkinson Microwave Anisotropy Probe (WMAP). The typical temperature fluctuation on the sky is of the order of $10^{-5}$.](image)

Fig. 1.— Cosmic microwave background (CMB) radiation as observed by the Wilkinson Microwave Anisotropy Probe (WMAP). The typical temperature fluctuation on the sky is of the order of $10^{-5}$.

We currently have a structure formation model in a universe that is dominated by dark matter and dark energy. The basic picture for the formation of disk and elliptical galaxies were discussed at the end of Chapter 1.

5.1. Einstein-de Sitter universe

A rigorous treatment of galaxy formation requires the knowledge of General Relativity, a complete thermal history of the universe and linear perturbation theories that we do not possess currently (some of these topics will be covered in your third-year “cosmology” course).

Fortunately, we can still make headways using a local Newtonian approximation. This is because of the cosmological principle, which assumes that the universe is homogeneous and isotropic on large-scales (this assumption is supported by observations).

In particular, there is a so-called Birkhoff’s theorem in General Relativity, which states that for a spherical mass distribution, the motion within some radius $R$ is influenced only by matter within; external mass distribution exerts no force. There is, of course, a perfect analogue in Newtonian mechanics. We shall apply this theorem to study the motion of matter and expansion of the universe. Specifically, we will study the Einstein-de Sitter cosmology in some detail (§5.1), how an overdense region grows in such a cosmology (§5.2), the cooling of baryons and the Jeans instability (§5.3), and finish with an short epilogue.
Let us consider a small uniform spherical region in an homogeneous and isotropic universe (with no cosmological constant and pressure). The radius of the sphere is $R(t)$ and the density is $\rho(t)$. If the region is small (non-relativistic), then we can apply Newtonian mechanics. The equation of motion for a unit mass on the sphere is

$$\ddot{R} = -\frac{GM}{R^2}. \quad (1)$$

The energy conservation equation is

$$\frac{1}{2} \dot{R}^2 + \left(-\frac{GM}{R}\right) = E. \quad (2)$$

In General Relativity, the total energy $E$ is related to the curvature of the universe. For $E > 0$, the system is unbound and it expands forever; this corresponds to an open universe with a negative spatial curvature. For $E < 0$, the system is bound and it will first expand, turn around and eventually recollapse in a “big crunch”; this corresponds to a closed universe with a positive spatial curvature. The case with $E = 0$ separates the open and closed universes. It is called the Einstein-de Sitter cosmology; it has a flat geometry.

So for the Einstein-de Sitter universe, the equations of motion and energy conservation are simply given by

$$\ddot{R} = -\frac{GM}{R^2}, \quad (3)$$

$$\frac{1}{2} \dot{R}^2 - \frac{GM}{R} = 0. \quad (4)$$

These two equations are in fact exact in General Relativity and apply not only to small-scales but also to large-scales due to the cosmological principle.

From the energy conservation equation, and $M = \rho \times 4\pi R^3/3$, we find that

$$\dot{R}^2 = \frac{2GM}{R} = \frac{8\pi G \rho}{3} R^2 \rightarrow \left(\frac{\dot{R}}{R}\right)^2 = \frac{8\pi G \rho}{3} \quad (5)$$

The Hubble expansion law states that

$$\dot{R} = HR, \quad (6)$$

where $H$ is the Hubble constant\(^1\). Hence we have from eq. (5),

$$H^2 = \frac{8\pi G \rho}{3} \rightarrow \rho = \frac{3H^2}{8\pi G} \equiv \rho_{\text{cr}}, \quad (7)$$

where we have defined the critical density as $\rho_{\text{cr}} = 3H^2/(8\pi G)$.

\(^1\)It is in fact not a constant but a function of time!
It is customary to define a density parameter

$$\Omega \equiv \frac{\rho}{\rho_{cr}}. \quad (8)$$

For the Einstein-de Sitter universe, $\Omega = 1$. A closed universe corresponds to $\Omega > 1$ and an open universe corresponds to $\Omega < 1$.

Let us further restrict ourselves to a matter-dominated universe. Suppose that the spherical region has a density, $\rho_0$, a radius, $R_0$, and the Hubble constant is $H_0$ at time $t_0$, then from mass conservation we have

$$\rho \frac{4\pi}{3} R^3 = \rho_0 \frac{4\pi}{3} R_0^3 \rightarrow \rho R^3 = \rho_0 R_0^3. \quad (9)$$

The density decreases as $R^{-3}$ as the universe expands and the volume increases as $R^3$. A relation we will repeatedly use later concerns the factor $GM$

$$GM = G (\rho_0) \frac{4\pi}{3} R_0^3 = G \left( \frac{3H_0^2}{8\pi G} \right) \frac{4\pi}{3} R_0^3 = \frac{1}{2} H_0^2 R_0^3 = \frac{1}{2} H^2 R^3. \quad (10)$$

From the energy conservation equation, we have

$$\ddot{R}^2 = \frac{2GM}{R} = \frac{H_0^2 R_0^3}{R}. \quad (11)$$

The solution to the above equation is

$$R(t) = R_0 \left( \frac{3}{2} H_0 t \right)^{2/3}. \quad (12)$$

We can define a scale factor

$$a(t) \equiv \frac{R(t)}{R_0} = \left( \frac{3}{2} H_0 t \right)^{2/3}. \quad (13)$$

Whatever the initial size, the radius of a region increases as the scale-factor, $a(t)$. Notice that as a consequence of eq. (13),

$$\rho(t) \propto R^{-3} \propto t^{-2}. \quad (14)$$

The density of the universe decreases as $t^{-2}$.

We can derive the current age of the universe at $t_0$ by setting $a(t_0) = 1$, we find that

$$t_0 = \frac{2}{3H_0}. \quad (15)$$

Taking $t_0$ as the present-day, and $H_0 = 70 \text{ km s}^{-1} \text{ Mpc}^{-1}$, we find an age of 9.5Gyr, which is younger than the age of the oldest globular clusters. This is obviously absurd! This implies that the Einstein-de Sitter model cannot accurately describe our universe. Now we know that the universe is likely to have a present-day matter density parameter $\Omega = 0.3$ and a vacuum energy density parameter ($\Lambda = 0.7$), so that the universe is flat! This cosmology solves the age crisis (and other problems).
However, the Einstein-de Sitter cosmology is particularly simple to treat and offers valuable insight in a number of things so we will use it throughout this Chapter.

It is also customary to define a dimensionless deceleration parameter \( q \equiv -\frac{\ddot{R}}{R^2} \). In the Einstein-de Sitter cosmology,

\[
q = -\frac{\ddot{R}}{R^2} = -\frac{R(-GM/R^2)}{2GM/R} = \frac{1}{2}.
\]  

(16)

So the deceleration parameter is 1/2 for the Einstein-de Sitter universe. Notice that in the Einstein-de Sitter universe, the acceleration is negative, so the universe decelerates \((q > 0)\), but the magnitude of deceleration \((GM/R^2)\) becomes smaller and smaller.

We can also interpret the redshift in an expanding universe. Let us consider two very close points \( P \) and \( P' \) in an expanding universe, they have radial coordinates \( R \) and \( R + dR \). Due to the Hubble expansion, the relative velocity between these two points are

\[
dv = H dR = \frac{\dot{R}}{R} dR.
\]  

(17)

Suppose \( P' \) sends a light signal at \( t \) with frequency \( \nu \), and it is received by an observer at \( P \) at \( t + dt \). Obviously \( dt = dR/c \).

Due to the Hubble expansion and the Doppler shift, the signal received by \( P \) will have a slightly different frequency \( \nu' \), using eq. (17),

\[
\frac{\nu' - \nu}{\nu} = \frac{dv}{\nu} = -\frac{dv}{c} = -\frac{\dot{R}}{c R} (c dt) = -\frac{dR}{R}.
\]  

(18)

So this gives a relation between the frequency and expansion rate between two points separated by an infinitesimal amount \( dR \): \( dv/\nu = -dR/R \). The negative sign indicates that as the universe expands, the wavelength is redshifted (becomes larger) and hence the frequency becomes lower.

We can integrate along the radial coordinates. Suppose a light signal is emitted with frequency \( \nu \) when the radius is \( R \) and received by us at the present epoch when the radius is \( R_0 \) and frequency \( \nu_0 \). Then we have

\[
\int_\nu^{\nu_0} \frac{d\nu}{\nu} = \int_R^{R_0} \frac{dR}{R},
\]  

(19)

which gives

\[
\ln \frac{\nu_0}{\nu} = -\ln \frac{R_0}{R}.
\]  

(20)

Now the redshift is defined as \( \nu_0/\nu = 1/(1 + z) \), i.e., the frequency observed by us is smaller by a factor of \( 1/(1 + z) \); correspondingly the wavelength is redshifted by a factor of \((1 + z)\). Using the previous equation, we find that

\[
a(t) \equiv \frac{R}{R_0} = \frac{1}{1 + z}.
\]  

(21)
The redshift is therefore directly related to the scale factor. This is the correct interpretation of redshift in an expanding universe.

As \(a(t) = (3H_0 t/2)^{2/3}\), using the previous equation, we can also express the cosmic time as a function of redshift

\[
t = \frac{2}{3H_0} (1 + z)^{-3/2}
\]

in the Einstein-de Sitter universe.

### 5.2. Spherical top-hat model

Let us consider a uniform overdense spherical region embedded in an Einstein-de Sitter universe. At the initial time \(t_0\), the region has a radius \(R_0\) and density \(\rho_0\). The background universe has density \(\bar{\rho}_0\) and a Hubble constant \(H_0\). The density contrast at \(t_0\) is defined as

\[
\delta_0 \equiv \frac{\rho_0}{\bar{\rho}_0} - 1 > 0
\]

The density contrast at later times will be written as \(\delta(t)\).

The equation of motion is given by

\[
\ddot{R} = -\frac{GM}{R^2} = -(1 + \delta_0) \frac{H_0^2 R_0^3}{2 R^2},
\]

where we have used eq. (10) and the factor \((1 + \delta_0)\) arises due to the overdensity at \(t = t_0\).

The energy conservation equation reads

\[
\frac{1}{2} \dot{R}^2 + \left(-\frac{GM}{R}\right) = E.
\]

Due to the overdensity, the energy is no longer zero. If the sphere initially follows the Hubble expansion, then the energy for a particle at radius \(R_0\) at \(t = t_0\) is

\[
E = \left(\frac{\dot{R}^2}{2} - \frac{GM}{R}\right) \bigg|_{t=t_0} = \frac{(H_0 R_0)^2}{2} - \frac{(1 + \delta_0) H_0^2 R_0^3}{2 R_0} = -\frac{\delta_0 H_0^2 R_0^3}{2}
\]

And hence the energy conservation equation at later times is

\[
\frac{1}{2} \dot{R}^2 + \left(-\frac{GM}{R}\right) = -\frac{\delta_0 H_0^2 R_0^3}{2} \rightarrow \dot{R}^2 = H_0^2 R_0^2 \left[(1 + \delta_0) \frac{R_0}{R} - \delta_0\right]
\]

Eqs. (27) and (24) are two equations involving first- and second-order derivatives of \(R\) with respect to \(t\). These two equations can be solved; the parametric solution is well known

\[
R(t) = \frac{R_{\text{max}}}{2} (1 - \cos \theta), \quad R_{\text{max}} = R_0 \frac{1 + \delta_0}{\delta_0}
\]

(28)
\[ t = \frac{t_{\text{max}}}{\pi}(\theta - \sin \theta), \quad t_{\text{max}} = \frac{1}{H_0} \left( 1 + \frac{\delta_0}{2H_0^3/2} \right) \pi \] (29)

Notice that \( R_{\text{max}} \) and \( t_{\text{max}} \) are functions of \( R_0, H_0 \) and \( \delta_0 \) only; their meanings are explained below. Let us discuss the behavior of the parametric solution:

1. The maximum radius \( R_{\text{max}} \) is achieved when \( \theta = \pi \) with \( t = t_{\text{max}} \).

2. when \( \theta = 2\pi \) (i.e., when \( t = 2t_{\text{max}} \)), the radius becomes \( R(2t_{\text{max}}) = 0 \). i.e., the sphere will collapse to a singular point!

3. The overdense region first follows the universal expansion (although at a slower rate), reaches a maximum radius, turns around with radius \( R = R_{\text{max}} \) when \( t = t_{\text{max}} \), and then recollapses in a “big crunch”. The behaviour is shown in Fig. 3.

4. However, the system cannot collapse to a singular point at \( t = 2t_{\text{max}} \) due to gas pressure (which we neglected). As the radius becomes smaller, the density (and hence) the pressure becomes higher. Shocks will form when gas runs into each other and energy will be converted into random motions in the process. Small inhomogeneities will be amplified in the re-collapse phase. The associated rapid changes in the potential lead to violent relaxations to an equilibrium state, which will have some finite radius, usually called the virial radius, \( r_{\text{vir}} \). The relaxation process to the final equilibrium is usually called virialisation.

![Image](Fig. 2.— The evolution of a uniform overdense spherical region in an Einstein-de Sitter background universe.)
We can estimate the virial radius using energy conservation. At maximum expansion

\[ E = T + \phi = 0 + \left( -\frac{GM}{R_{\text{max}}} \right) M \times A, \]  

(30)

where the kinetic energy \( T = 0 \) at maximum expansion (all velocities are zero), and the term in the bracket is the potential energy per unit mass at the edge of the sphere. Multiplying by \( M \) gives the potential energy if all the mass is at the edge of the sphere. Of course, not all the mass is at the edge, so the mass distribution as a function of radius gives rises to some correction factor \( A \) in the potential energy. For a uniform sphere, one can show that \( A = \frac{3}{5} \).

After virialisation, the system is in a steady-state, so from virial theorem, \( 2T + \phi = 0 \), and the total energy is

\[ E = T + \phi = \frac{\phi}{2} = \frac{1}{2} \left( -\frac{GM}{r_{\text{vir}}} \right) M \times A', \]  

(31)

where \( A' \) is a similar factor as \( A \). Comparing the two expressions (eqs. 30 and 31) for the total energy, if \( A = A' \), then one arrives at

\[ r_{\text{vir}} = \frac{R_{\text{max}}}{2}. \]  

(32)

The virial radius is simply one half of the maximum radius at turn around, a remarkably simple result!

5. Suppose initially \( \delta_0 \ll 1 \) at time \( t = t_0 \) (see below), then when \( t \ll t_{\text{max}} \), the system will deviate only slightly from the Einstein-de Sitter solution. One can show that (see Problem set 5)

\[ R(t) = \bar{R}(t)(1 - \delta_R), \quad \bar{R}(t) = R_0(3/2H_0t)^{2/3}, \]  

(33)

where \( \bar{R}(t) \) is the radius of the system would have without the overdensity (cf. eq. 13), and the deviation \( (\delta_R) \) can be shown to be

\[ \delta_R = \frac{1}{20} \left( \frac{6\pi t}{t_{\text{max}}} \right)^{2/3}. \]  

(34)

The above equation also shows that when \( t \ll t_{\text{max}}, \delta_R \ll 1 \).

The density is therefore given by

\[ \rho(t) = \frac{M}{4\pi R^3(t)/3} = \bar{\rho}(t)(1 - \delta_R)^{-3} \approx \bar{\rho}(t)(1 + 3\delta_R) \]  

(35)

where \( \bar{\rho}(t) = M/(4\pi \bar{R}^3/3) \) and we have used the first-order approximation that \( (1 - x)^{-3} \approx 1 + 3x \) when \( x \ll 1 \). The density contrast is then

\[ \delta(t) \equiv \frac{\rho(t)}{\bar{\rho}(t)} - 1 = 3\delta_R = \frac{3}{20} \left( \frac{6\pi t}{t_{\text{max}}} \right)^{2/3} \propto \frac{R}{R_0} = \frac{1}{1 + z}. \]  

(36)

where we have used the fact that \( a(t) \equiv R/R_0 \propto t^{2/3} \) (cf. eq. 13).
6. From observations of the microwave background radiation, the temperature fluctuations on the sky is of the order of $10^{-5}$ at $z \approx 10^3$. Before this redshift, baryons are ionised, and the collisions between free electrons and photons make them tightly coupled. After $z \sim 10^3$, the temperature of the universe is low enough and the electrons and protons combine to form neutral hydrogen. As a result, the scattering of photons by free electrons (whose density has decreased dramatically) becomes rare, and they can escape freely and we see these photons as the microwave background radiation. At any given direction on the sky, the microwave background spectrum is a perfect blackbody. And the temperature of these blackbody’s only differ by $\sim 10^{-5}$ in different directions. The density contrast in baryons is of the same order, $10^{-5}$.

This immediately gives rise to a problem. As we have shown the density contrast grows as the scale factor, $a(t) \propto 1/(1+z)$, so from $z \sim 10^3$ to the present-day ($z = 0$), the density contrast is only expected to be of the order of $10^{-5} \times (1+z) \sim 10^{-2}$, many orders of magnitude lower than the density contrast observed near the Sun ($\sim 2.5 \times 10^6$).

How can we solve this discrepancy? The answer is dark matter. The dark matter component is no longer coupled to photons at $z \sim 10^3$ because they have much smaller interaction cross-sections with baryons and photons, so their density inhomogeneity can grow to much higher values than the density contrast of photons and baryons ($\sim 10^{-5}$) at $z \sim 10^3$.

At smaller $z$, the baryons also decouple from photons, and hence they can quickly fall into existing dark matter overdense regions and catch up in density inhomogeneity. The existence of dark matter is essential for structure (galaxy) formation.

At $t = t_{\text{max}}$, if our first-order expansion is correct, then $\delta = 3/20(6\pi)^{2/3} \approx 1.062$ while $t = 2t_{\text{max}}$, $\delta = 3/20(12\pi)^{2/3} \approx 1.686$. Obviously our assumption of small $\delta$ ($\ll 1$) breaks down for these two epochs, hence we have to calculate the density contrast more carefully.

7. Density contrast at $t = t_{\text{max}}$

$$\frac{\rho(t_{\text{max}})}{\bar{\rho}(t_{\text{max}})} = \frac{(1 + \delta_0)\bar{\rho}_0 [R_{\text{max}}/R_0]^{-3}}{\bar{\rho}_0 [R(t_{\text{max}})/R_0]^{-3}} = \frac{9}{16} \pi^2 = 5.552,$$  \hspace{1cm} (37)

where we have used eqs. (28), (29) and (33).

8. Density contrast at $t = 2t_{\text{max}}$

This can be obtained by comparing the densities with those at $t = t_{\text{max}}$. At $t = 2t_{\text{max}}$, the density of the background universe decreases by a factor of 4 as $\bar{\rho}(t) \propto t^{-2}$ (cf. eq. 14). The virial radius at $t = 2t_{\text{max}}$ is a factor of two smaller than $R_{\text{max}}$ at $t = t_{\text{max}}$, and hence the density has increased by a factor of $2^3 = 8$ from $t_{\text{max}}$ to $2t_{\text{max}}$. Therefore

$$\frac{\rho(2t_{\text{max}})}{\bar{\rho}(2t_{\text{max}})} = \frac{8 \times \rho(t_{\text{max}})}{1/4 \times \bar{\rho}(t_{\text{max}})} = 32 \times \frac{9}{16} \pi^2 = 18\pi^2 \approx 178$$  \hspace{1cm} (38)

So after virialisation, the density contrast is of the order of 200. This is still orders of magnitude smaller than the observed density contrast near the Sun. Part of the solution is
that we have ignored the \textit{internal} structures of virialised region. The central part will have much higher density than the average density in the region. The second factor is that baryons can further cool and sink toward the centre, and increase the density contrast along the way. We now turn to baryon cooling in the next section.

5.3. Jeans instability and Baryonic cooling

5.3.1. Jeans instability

Without pressure, an overdense region initially \textit{at rest} will collapse on a free-fall (or dynamical) time scale, $t_{ff}$. However, the pressure will resist the collapse. The pressure resistance is propagated on a sound-crossing time scale, $t_s$. Whether gravity or pressure wins depends on the relative magnitude of $t_{ff}$ and $t_s$.

\textit{Free-fall time scale:} Let us consider a uniform sphere with density $\rho$. A particle is initially at radius $R$. The equation of motion for the particle is

$$
\ddot{R} = -\frac{GM}{R^2} = -\frac{4\pi G \rho}{3} R
$$

This is a simple harmonic oscillator, with $\omega^2 = 4\pi G \rho / 3$. All the particles within the sphere will collapse from rest to the origin within a quarter of the oscillation period. We identify this time as the free-fall timescale

$$
t_{ff} = T_4 = \frac{1}{4} \frac{2\pi}{\omega} = \sqrt{\frac{3\pi}{16G\rho}} \propto (G\rho)^{-1/2}
$$

\textit{Sound-crossing time scale:}

$$
t_s = \frac{R}{v_s}, \quad v_s \sim \sqrt{v^2} \sim \left(\frac{kT}{m_p}\right)^{1/2},
$$

where $v_s$ is the sound speed\textsuperscript{2}. If $t_{ff} \gg t_s$, the pressure responds fast enough to counter the gravitational force, and the system oscillates as sound waves. On the other hand, if $t_{ff} \ll t_s$, the pressure does not respond fast enough to resist the gravitational collapse, and the system collapses on a free-fall time scale. This condition implies that

$$
\sqrt{\frac{3\pi}{16G\rho}} < \frac{R}{v_s} \rightarrow R > v_s \sqrt{\frac{3\pi}{16G\rho}}
$$

\textsuperscript{2}Rigorously, the adiabatic sound speed involves the adiabatic index, $\Gamma = 5/3$. The fact that we have a mixture of hydrogen and helium also gives rise to an additional factor (of order unity) involving the so-called mean molecular weight.
A more rigorous derivation gives
\[ R > v_s \sqrt{\frac{\pi}{G\rho}} \equiv \lambda_J, \]  \hspace{1cm} (43)
where \( \lambda_J \) is called the Jeans length. This is the same as our rough derivation except a numerical factor within the square root.

The Jeans mass is defined as the mass contained within a sphere of diameter \( \lambda_J \):
\[ M_J \equiv \frac{4\pi}{3} \rho \left( \frac{\lambda_J}{2} \right)^3 = \frac{\pi}{6} \rho \lambda_J^3. \] \hspace{1cm} (44)
\[ \text{e.g. After redshift } z \sim 10^3, \text{ baryons and photons are decoupled due to infrequent collisions. The temperature of the baryons is about } 2700 \text{ K, and the baryon density is } \rho_B \approx 4.3 \times 10^{-19} \text{ kg m}^{-3}. \text{ This gives a sound speed of } v_s \approx (kT/m_p)^{1/2} = 5 \text{ km s}^{-1}, \text{ and a Jeans length } \lambda_J = v_s \sqrt{\pi/G\rho_B} \approx 1.6 \text{ kpc}. \text{ The Jeans mass is } M_J = 3.7 \times 10^5 M_\odot. \text{ The Jeans mass at } z \sim 10^3 \text{ is similar to the mass of globular clusters and the sound speed is also roughly equal to the velocities of stars within globular clusters. This is a remarkable coincidence whose meaning is not entirely clear.}

All structures on the scale of the Jeans length above the Jeans mass can collapse and the density contrast grows as the scale factor. An overdense region in an expanding universe eventually recollapses and virialises. As we have discussed in the §5.2, the average density in the virialised region is approximately \( 18\pi^2 \) times the mean density in the universe. The density contrast is not yet as high as that in the solar neighbourhood. To achieve this, one must consider the cooling of baryons.

### 5.3.2. Cooling of Baryons

During virialisation, gas is thought to be shock-heated to the so-called virial temperature:
\[ kT_v \approx \frac{1}{2} m_p V^2 - T_v \approx \frac{m_p V^2}{2k} = 2.4 \times 10^6 \text{ K} \left( \frac{V}{200 \text{ km s}^{-1}} \right)^2. \] \hspace{1cm} (45)
So for a galaxy with \( V = 200 \text{ km s}^{-1} \) (similar to the MW), the temperature is in the (soft) X-ray range.

At such a high temperature, the gas will be fully ionised (recall that the ionisation energy for hydrogen, the most abundant element in the universe, is \( 13.6 \text{ eV} \sim 10^5 \text{ K} \)). Electrons will radiate due to Coulomb accelerations (so-called thermal bremsstrahlung radiation). The gas will cool due to this radiative loss and sink toward the centre, increasing density contrast in the process. Notice that dark matter will not undergo the same process because they are “collisionless” due to their much smaller interaction cross-sections with baryons and photons. So cooling provides a natural way of separating baryons and dark matter. Because of this, in the central part of galaxies, baryons are expected to contribute substantially to the mass budget.
The energy loss per unit time per unit volume can be approximated as, when \(10^5 \text{K} < T < 3 \times 10^6 \text{K}\),
\[
\dot{\epsilon} \approx 2.5 \times 10^{-37} \left( n_p/\text{m}^3 \right)^2 (T/10^6\text{K})^{-1/2} \text{J s}^{-1} \text{m}^{-3},
\]
where \(n_p\) is the number density of protons (which is roughly equal to the electron number density). Notice that the cooling rate is proportional to \(n_p^2\) because the radiative process involves collisions between electrons (which gives rise to a factor of \(n_e \approx n_p\)) and protons (which gives rise to another factor of \(n_p\)). As the average thermal energy per unit volume is roughly given by \(n_p \times 3kT_v/2\), and hence the cooling time scale is given by
\[
t_{\text{cool}} \approx \frac{n_p^3kT_v/2}{\dot{\epsilon}} \propto \frac{n_pT_v}{T_v^{-1/2}n_p^2} \propto T_v^{3/2}n_p^{-1},
\]
(46)
The proton number density
\[
n_p = \frac{f_B M}{4\pi R^3/3} \propto f_B MR^{-3},
\]
(48)
where \(f_B\) is the fraction of mass in baryons, \(M\) is the total mass of the Galaxy, and \(f_B M\) gives the total mass in baryons. Combined with \(T_v \propto V^2 \propto M/R\) (cf. eq. 45), we have
\[
t_{\text{cool}} \propto T_v^{3/2}(f_B MR^{-3})^{-1} \propto M^{1/2}R^{3/2}f_B^{-1} \rightarrow t_{\text{cool}} \equiv C_{\text{cool}} M^{1/2}R^{3/2}f_B^{-1},
\]
(49)
where we have written the proportionality as \(C_{\text{cool}}\).

Valuable insight can be gained by comparing the cooling time scale and the free-fall time scale. If \(t_{\text{cool}} \ll t_{\text{ff}}\), the gas cools quickly, dynamical processes are unable to adjust the pressure distribution. Pressure is lost and the system rapidly collapses on a free-fall time scale. The dynamical time scale is
\[
t_{\text{ff}} \propto (G\rho)^{-1/2} \propto M^{-1/2}R^{3/2} \rightarrow t_{\text{ff}} \equiv C_{\text{ff}} M^{-1/2}R^{3/2}
\]
(50)
where \(C_{\text{ff}}\) is the proportionality. The condition \(t_{\text{cool}} < t_{\text{ff}}\) yields
\[
C_{\text{cool}} M^{1/2}R^{3/2}f_B^{-1} < C_{\text{ff}} M^{-1/2}R^{3/2} \rightarrow M < \frac{C_{\text{ff}}}{C_{\text{cool}}} f_B \approx \text{few} \times 10^{13} f_B M_\odot.
\]
(51)

If \(f_B = 1\) (all mass is baryons), this gives an upper limit of the total baryon mass of \(\text{few} \times 10^{13} M_\odot\), which is much larger than the known baryon mass of galaxies.

However if \(f_B = 0.1\) (10% of the mass is in baryons), this gives an upper limit of the total galaxy mass of \(\text{few} \times 10^{12} M_\odot\), and a baryon mass of of \(\text{few} \times 10^{11} M_\odot\). This value is more in line with the observed baryon mass in a typical galaxy. Hence the inclusion of dark matter makes the separation of successful and failed galaxies from the baryonic cooling argument much more plausible.
5.4. Epilogue

We have covered a tiny fraction of galactic astronomy. There are many topics which one can discuss. To name a few, gravitational lensing, dark matter candidates and searches, basic theories of stellar structure and evolution and interstellar medium.

There are still many open questions in this subject. Galaxy formation and evolution is a truly fascinating subject, and it will remain so for many decades to come. I hope that some of you will be interested in exploring some of these issues in the future.
Problem set 5

1. The density around the solar neighbourhood is around $0.1 M_\odot \text{pc}^{-3}$. Assuming a Hubble constant of 70 km s$^{-1}$ Mpc$^{-1}$, and a matter density of $\Omega = 0.3$ in units of the critical density. Find the density contrast of the local density relative to the mean mass density in the universe.

2. In the lecture, we derived that after virialisation the mean density of a spherical overdense region is about $18\pi^2$ that of the average density in the Einstein de Sitter universe, $3H^2/(8\pi G)$. Let us write the mass of the overdense region as $M$, and radius as $r$. Suppose the rotation speed as a function of radius in the virialised region has a flat rotation curve with amplitude $V$. Show that $r = V/(3\pi H)$, and the mass $M = V^3/(3\pi GH) \propto V^3$.

3. The brightness of HI emission along the line of sight in the direction with longitude $l = 30^\circ$ has been measured. It is found that the brightness drops to zero at a radial velocity of 100 km s$^{-1}$. What can be inferred about the rotation curve of the Milky Way?

4. Describe how do disk and elliptical galaxies form in the current hierarchical structure formation model.

5. (optional, will not be examined) In the lecture notes, we derived that

$$R(t) = \frac{R_{\text{max}}}{2} (1 - \cos \theta), \quad t = \frac{t_{\text{max}}}{\pi} (\theta - \sin \theta),$$

where $R_{\text{max}}$ and $t_{\text{max}}$ are the maximum radius and its corresponding time. For simplicity, let us define $r = 2R/R_{\text{max}}$ and $\tau = \pi t/t_{\text{max}}$.

a) When $\theta$ is very small, using Taylor series, show that the second equation implies

$$\tau = \frac{\theta^3}{3!} - \frac{\theta^5}{5!}$$

b) Show that the lowest order solution is $\theta = (6\tau)^{1/3}$, and the next order approximation is $\theta = (6\tau)^{1/3} + \frac{1}{10} \tau$.

c) Substituting the above solution into the expression of $r$, show that the two leading orders of $r$ in terms of $\tau$ is given by

$$r = \frac{1}{2} (6\tau)^{2/3} - \frac{1}{40} (6\tau)^{4/3}$$

d) Show that the first term is simply the average expansion of the universe, and the factor $\delta_R$ we defined in the lecture notes is simply given by $(6\tau)^{2/3}/20 = (6\pi t/t_{\text{max}})^{2/3}/20$.

Galaxies: Solution 5

1. The local density is $\rho_{\text{local}} \approx 0.1 M_\odot \text{pc}^{-3} = 6.8 \times 10^{-21} \text{kg m}^{-3}$.

The matter density is $\rho_m = \Omega_m \rho_{\text{cr}} = 0.3 \times 3H_0^2/(8\pi G) = 2.8 \times 10^{-27} \text{kg m}^{-3}$. So the density contrast is about $2.5 \times 10^6$. 
2. For the virialised region with the total mass $M$ and radius $r$, we have

$$\frac{GM}{r} = V^2$$

and

$$\frac{M}{4\pi r^3/3} = 18\pi^2 \left( \frac{3H^2}{8\pi G} \right)$$

Combining these two equations, we find that

$$r = V/(3\pi H)$$

and

$$M = \frac{V^3}{(3\pi GH)} \propto V^3.$$  

3. The maximum radial velocity for $0 < l < 90^\circ$ is achieved at the tangent point, $R = R_0 \sin l = 4$ kpc for $R_0 = 8$ kpc. This implies that the circular velocity at 4 kpc from the Galactic centre is (see the Chapter on Milky Way: section 9.3.1, equation 10)

$$V_c(R = 4 \text{ kpc}) = 100 + V_0 \sin l = 100 + 220 \times \sin 30^\circ = 210 \text{ km s}^{-1}$$ (52)

One sees that the rotation curve is approximately flat from 4 kpc to 8 kpc.

4. See the end of Chapter 1.

5. a) From given equation, we have $\tau = \theta - \sin \theta$. As $\sin \theta = \theta - \theta^3/3! + \theta^5/5! + \cdots$, it follows that $\tau = \theta^3/3! - \theta^5/5! = \theta^3/6 - \theta^5/120$.

b) As $\theta \ll 1$, to the lowest order we have $\tau = \theta^3/6$, so we have $\theta \equiv \theta_0 = (6\tau)^{1/3}$. Next we write $\theta = \theta_0 + \theta_1$, where $\theta_1$ is the next-order correction. Now from $\tau = \theta^3/6 - \theta^5/120$, we have, retaining to the first order,

$$\tau = \frac{1}{6} (\theta_0 + \theta_1)^3 - \frac{1}{120} (\theta_0 + \theta_1)^5 \approx \frac{1}{6} (\theta_0^3 + 3\theta_0^2\theta_1) - \frac{1}{120} (\theta_0^5 + \cdots)$$

Substituting the expression of $\theta_0$ in the above equation, and collect all the terms, we find that

$$\theta_1 = \frac{1}{60} \theta_0^3 = \frac{\tau}{10}$$

Notice that $\theta_1 \propto \theta_0^3$ in terms of order on $\tau$.

c) As $r = 1 - \cos \theta \approx \theta^2/2 - \theta^4/24$, using $\theta = \theta_0 + \theta_1$, we find that to the lowest order

$$r = \frac{1}{2}(\theta_0 + \theta_1)^2 - \frac{1}{24}(\theta_0 + \theta_1)^4 = \frac{1}{2}(\theta_0^2 + 2\theta_0\theta_1 + \cdots) - \frac{1}{24}(\theta_1^4 + \cdots)$$

Substituting the expressions of $\theta_0$ and $\theta_1$ into the above equation, we find that

$$r = \frac{1}{2}(6\tau)^{2/3} - \frac{1}{40}(6\tau)^{4/3} = \frac{1}{2}(6\tau)^{2/3} \left[ 1 - \frac{1}{20}(6\tau)^{2/3} \right]$$

d) The zeroth order of $r$ satisfies $r = \frac{1}{2}(6\tau)^{2/3}$, which implies

$$2R/R_{\text{max}} = \frac{1}{2}(6\pi t/t_{\text{max}})^{2/3}.$$  

One can easily show that this is indeed the average expansion rate of the universe using the definitions of $R_{\text{max}}$ and $t_{\text{max}}$ in terms of $\delta_0$ and $R_0$ (see the lecture notes). Hence, the correction factor $\delta_R$ is then as given in the problem set.
Revisions questions with hints

[Jan. 2003 exam questions]

• State the Virial Theorem and explain the terms used [2 marks]
  [hint: $2T + \phi = 0$, where $T$ is the total energy and $\phi$ is the total potential energy]

What are the essential conditions that must be satisfied for this theorem to be applicable to the investigation of a cluster of galaxies? [3 marks]
  [hint: the cluster must be in statistical equilibrium and more or less spherical.]

Use the theorem to derive an approximate relationship between the total mass of a cluster, its mean radius and the mean velocities of the individual galaxies [5 marks]
  [hint: $V^2 = GM/R$]

• How can measurements of the distances from the Sun to the Galactic globular clusters give the distance of the Sun from the centre of the Galaxy? [5 marks]
  [hint: the Galactic centre can be identified as the peak of the globular density distribution. $R_0 = 8 \pm 0.8$ kpc.]

What parameters of the motions of nearby stars are measured to derive Oort’s constants? [2 marks]
  [hint: proper motions]

Calculate the mean separation of stars in the disk of a typical disk galaxy. Assume the stars are concentrated in the disk as opposed to the bulge or halo. [3 marks]
  [hint: first estimate the number surface density of stars, $\Sigma_*$ (e.g., the mass surface density is $75M_\odot$ pc$^{-2}$, which implies a number density of $75$ pc$^{-2}$ if all stars are the same as the Sun). And then $d \sim \Sigma^{-1/2} \sim 0.1$ pc. An answer within a factor of few is acceptable.]

[Jan. 2002 exam questions]

• What is the ‘winding dilemma’ when applied to spiral arms in disk galaxies? [6 marks]
  [hint: if the spiral arms are permanent material arms, then their spiral arms will be tightly wound up, the pitch angles will be much smaller than the observed values. This shows that spiral arms must be density waves.]

• If Oort’s constants $A$ and $B$ are measured as $12.5$ and $-12.5$ km s$^{-1}$ kpc$^{-1}$, respectively, use the relations
  \[
  \frac{V_0}{R_0} = A - B, \quad \left(\frac{dV}{dR}\right)_{R_0} = A + B
  \]
  to derive the galactic motions in the vicinity of the Sun, explain the meaning of the symbols. [15 marks]
[hint: For the given $A$ and $B$ values, $V_0 = R_0 \times (A - B) = 200\text{km/s}$ for $R_0 = 8 \text{kpc}$; and $(dV/dR)_{R_0} = -(A + B) = 0$, the rotation curve is perfectly flat. More recent determinations yield a slightly falling rotation curve. See Problem set 3.]

[Jan. 2001 exam questions]

• Explain how observed rotation curves of disk galaxies indicate the presence of ‘missing’ mass. [3 marks]

[hint: observed rotation curves are flat which implies that the total mass increases linearly with the radius, however, most light is in the inner part, and hence the mass in the outer part must be dominated by dark matter.]