Gravitational Collapse and Structure Formation in an Expanding Universe

J S Bagla and Pritpal Kaur Sandhu

We use Newtonian formalism to motivate the form of Friedmann equations that describe the expansion of the universe in the standard cosmological model. We use the same formalism to study the evolution of density perturbations in the universe. We show that a simple model like spherical collapse can be used to estimate the characteristics of halos of galaxies and clusters of galaxies.


(left) Jasjeet Bagla works at IISER Mohali. His research is mainly in the area of cosmology.

(right) Pritpal Kaur Sandhu works as a research scholar at IIT Indore. Her current research is on clusters of galaxies.

1. The Friedmann Equations

Observations show that, at very large scales, there is no special direction in the sky [1], nor is there any special region or location within the universe [2]. At an early stage in the development of cosmology, the notion of an isotropic and homogeneous universe was raised to the level of a guiding principle known popularly as the cosmological principle [3]. Most models of the universe take this into account and assume that the distribution of mass is homogeneous and isotropic, i.e., the density of matter is constant and not a function of position though it may be a function of time if the universe undergoes expansion or contraction. It is also clear that in such a model, there can be no bulk motions that may lead to departures from homogeneity and isotropy at a later stage. An important implication of the cosmological principle is that the choice of origin and the orientation of axes of the coordinate system used to describe the universe are completely arbitrary.

If we consider the universe to consist mainly of slowly moving particles (particles whose motions are non-relativistic), then the total energy density in the universe is

Keywords
Cosmology, gravitational collapse, galaxy formation.

dominated by the rest mass energy of these particles and the kinetic energy is a small and sub-dominant contribution. In such a case, the energy density is proportional to the mass density ρ. It is possible to construct a Newtonian model to describe the evolution of such a universe.

Let us consider an arbitrary origin in the smooth matter distribution, and then study the motion of a thin spherical shell at a distance r from the centre. The equation of motion for the shell is:

$$\ddot{r} = -\frac{GM}{r^2}. \tag{1}$$

The shell may expand or contract, but other types of motion such as rotation or shear are not allowed as these lead to the violation of homogeneity and isotropy. The motion of the shell is then conveniently described in co-moving coordinates that expand or contract with the universe as a whole. The physical coordinates $\mathbf{r}$ are related to the co-moving coordinates as $\mathbf{r}(t) = a(t)\mathbf{x}$, where the co-moving coordinates $\mathbf{x}$ are fixed for a given particle even as the universe expands or contracts. Using this, we can recast (1) as

$$\frac{\ddot{a}}{a} = -\frac{4}{3}\pi G\rho, \tag{2}$$

where ρ is the matter density and is related to the mass enclosed inside the shell at radius r as $M = 4\pi\rho r^3/3$. The expansion or contraction of the universe is described by the scale factor $a(t)$ that increases during expansion and decreases during contraction.

The mass enclosed within the shell remains constant as the matter within the shell remains inside even as the universe expands or contracts. This implies that $\rho \propto a^{-3}$ and we may write:

$$\rho(t) = \rho(t_0)\frac{a^3(t_0)}{a^3(t)} = \rho_0\frac{a_0^3}{a^3}.$$

We can also use the constancy of mass within the shell to find the first integral of motion from (1):

$$\left(\frac{\dot{a}}{a}\right)^2 + \frac{k}{a^2} = \frac{8\pi G}{3}\rho. \quad (3)$$

Here, k is a constant and may be considered equivalent to energy. It is clear that this equation admits three classes of solutions for non-relativistic matter:

- Open universe: $k < 0$. The sign of $\dot{a}$ does not change through the evolution of the universe. An expanding universe continues to expand forever and a contracting universe contracts for ever. In the case of an expanding universe, the scale factor $a(t)$ increases forever with a rate $\dot{a}$ that eventually approaches a constant value.

- Flat universe: $k = 0$. The comments relating to the unchanging sign of $\dot{a}$ for the open universe apply here as well. In the case of an expanding universe, the scale factor $a(t)$ increases forever with a rate that keeps on decelerating.

- Closed universe: $k > 0$. In this case, it is possible for the sign of $\dot{a}$ to change during the course of evolution. The expansion can slow down to a halt and start a collapsing phase. For the converse, we require $\ddot{a} > 0$ and that is not possible with normal matter (see (2)).

Even though we derived (3) and (2) using Newtonian arguments, it turns out that for non-relativistic matter these are also the equations we get in Einstein's general theory of relativity. Equation (3) holds for arbitrary constituents of the universe, even for relativistic constituents, whereas (2) requires a correction in case the constituent has non-negligible pressure. The relativistic form of the equation is:

$$\frac{\ddot{a}}{a} = -\frac{4}{3}\pi G\left(\rho + 3p\right), \qquad (4)$$

where p is pressure and we have taken $c = 1$. The set of Friedmann equations (3) and (4) contains three unknowns and, in general, we need to add another equation in order to solve for all the quantities. The relevant equation here is one that describes the relation between density and pressure for the constituents of the universe. However, here we are concerned mainly with non-relativistic matter and hence, we choose to ignore the pressure term as well as its relation with density.

Observations show that the universe is spatially flat, i.e., $k = 0$ is favoured by observations. In this case, the scale factor $a(t) \propto t^{2/3}$ if the universe contains only non-relativistic matter[1]. Such a universe is referred to as the Einstein–deSitter model. Observations made towards the end of the twentieth century show that the expansion of the universe is accelerating, which cannot happen in the framework we are using. This is attributed to a form of matter known as 'dark energy'. Thus, the Einstein-de Sittter model is not realistic in the light of present day knowledge. We will go ahead and analyse it anyway, since some of the aspects covered are unchanged. We will, indicate briefly the changes that dark energy make to the results, later on.

[1] The scale factor can be derived easily in this case by integrating the Friedmann equation (3) with $\rho = \rho_0\,(a_0/a(t))^3$.

2. The Clumpy Universe

The universe is not homogeneous and isotropic at small scales. We observe large fluctuations in density – indeed the Earth has an average density of 5 g/cc, whereas, the average density of the universe is close to 10^{-30} g/cc (see *Box* 1, (ii)). Our solar system and the Galaxy also represent very large over-densities. In the following discussion we would like to understand how large over-densities can form in a universe that was very smooth at early times.

Box 1. Density of the Universe and Hubble's Law

The relation between the co-moving and physical coordinates can be used to derive a relation between the expansion velocity and distance [A], i.e., $\mathbf{V} = H_0\mathbf{r}$, where H_0, the Hubble's constant, is the present value of $\dot{a}/a$.

Writing the relation (3) at the present time, we find:

$$H_0^2 + \frac{k}{a_0^2} = \frac{8\pi G}{3}\rho_0. \tag{i}$$

We see that there is a special value of the present day density ρ_0 for which k vanishes. This is known as the critical density: $\rho_\mathrm{c} = 3H_0^2/8\pi G$. Observations favour values of $H_0 \simeq 70$ km/s/Mpc and we get the critical density:

$$\begin{aligned} \rho_\mathrm{c} &= \frac{3H_0^2}{8\pi G} \\ &= 9.2 \times 10^{-25}\mathrm{kg/m^3}\left(\frac{H_0}{70\mathrm{km/s/Mpc}}\right)^2 \\ &\sim 10^{-30}\mathrm{g/cc}. \end{aligned} \tag{ii}$$

[A] J S Bagla, *Resonance,* Vol.14, p.216, 2009.

Our derivation of Friedmann equations using Newtonian mechanics can be generalised to the study of density perturbations if we assume that the perturbations are spherically symmetric [4]. The introduction of such a perturbation is easier to analyse as it can be shown that matter exterior to the perturbation does not affect its evolution in any way. We note here that the introduction of a spherically symmetric perturbation retains isotropy but removes homogeneity.

2.1 *Spherical Collapse Model*

Consider a spherical region with a constant over-density that is expanding along with the rest of the universe at early times. We assume that the initial over-density is small in magnitude and that the size of the over-dense region is small. Now consider a shell, initially at radius r_i, which encloses a mass M. It can be shown that mass will remain conserved as the system evolves. The equation of motion is:

$$\ddot{r} = -\frac{GM}{r^2}.$$

Using the constancy of mass enclosed we get the first integral of motion:

$$\frac{\dot{r}^2}{2} = \frac{GM}{r} + E. \tag{5}$$

The energy E has to be negative for this to be a bound perturbation and we write it explicitly as $E = -|E|$. Further, as mentioned above, we assume that the perturbation is taken to be situated in an Einstein–deSitter universe. The energy can be written in terms of initial conditions:

$$\begin{aligned} |E| &= -\frac{1}{2}\dot{r}_{\mathrm{i}}^2 + \frac{GM}{r_{\mathrm{i}}} \\ &= -\frac{1}{2}H_{\mathrm{i}}^2 r_{\mathrm{i}}^2 + \frac{4\pi G}{3}\rho_{\mathrm{i}} r_{\mathrm{i}}^2 \\ &= -\frac{1}{2}H_{\mathrm{i}}^2 r_{\mathrm{i}}^2 + \frac{4\pi G}{3}\bar{\rho}_{\mathrm{i}}\left(1+\delta_{\mathrm{i}}\right) r_{\mathrm{i}}^2 \\ &= \frac{1}{2}H_{\mathrm{i}}^2 r_{\mathrm{i}}^2 \delta_{\mathrm{i}}. \end{aligned} \tag{6}$$

Here, r_{i} and $\dot{r}_{\mathrm{i}}$ are the initial radius and velocity of the shell. We assume that these are related through the Hubble's law $\dot{r}_{\mathrm{i}} = H_{\mathrm{i}} r_{\mathrm{i}}$ with H_{i} as the value of Hubble's parameter at the initial time. $\rho_{\mathrm{i}} = \bar{\rho}_{\mathrm{i}}\left(1+\delta_{\mathrm{i}}\right)$ is the density of the perturbation and is related to the average density $\bar{\rho}_{\mathrm{i}}$ and density contrast δ_{i}. Qualitatively, we can see that as compared to the Einstein–deSitter universe, the expansion rate of the shell within the over-dense region slows down at a faster rate until it comes to a halt and then begins to recollapse. The slower expansion leads to a relative increase in the density within the shell as compared to the background universe. At the stage when its radius is maximum, we have $|E| = GM/r_{\mathrm{max}}$ and hence, $r_{\mathrm{max}} = r_{\mathrm{i}}\left(1+\delta_{\mathrm{i}}\right)/\delta_{\mathrm{i}} \simeq r_{\mathrm{i}}/\delta_{\mathrm{i}}$.

The solution to (5) can be written in a parametric form:

$$r = \frac{r_{\max}}{2}(1-\cos\theta),$$
$$t = \frac{1}{2H_{\mathrm{i}}\delta_{\mathrm{i}}^{3/2}}(\theta-\sin\theta), \qquad (7)$$

where we have assumed that $\delta_{\mathrm{i}} \ll 1$. This is obtained by integrating (5) and substituting for r. We note that the characteristic collapse time for the perturbation is $H_{\mathrm{i}}/\delta_{\mathrm{i}}^{3/2}$ and is smaller for perturbations with a larger initial density contrast. This implies that if we have a perturbation where density contrast decreases as we go to larger radii, the inner shells collapse first and outer regions collapse later[2].

[2] In such a case, the density contrast δ_{i} needs to be replaced by a corresponding volume averaged quantity in (7).

It is interesting to note that the characteristic collapse time is fairly long – indeed it is of the same order as the age of the universe. In contrast, the collapse of a perturbation of similar density in a static universe is about $\delta_{\mathrm{i}}^{3/2}$ times smaller, i.e., for a $\delta_{\mathrm{i}} \simeq 10^{-2}$, the difference is three orders of magnitude if we consider perturbations with the same density. Thus, the expansion of the universe slows down the collapse of perturbations by a significant amount and as a result, the growth of density perturbations in an expanding universe is a very slow process as compared to its counterpart in a static background.

The solution given above is cyclic in nature where shells collapse to the centre and then rebound to the same amplitude in the other direction. However, it is obvious that most perturbations do not have spherical symmetry. It can also be shown that any departures from spherical symmetry grow rapidly during collapse. Hence, our estimation is likely to deviate more and more from the collapse of real perturbations at late times. We can capture the physics of key stages during further collapse by realising that in the collapsing phase where deviations from a symmetric initial condition grow rapidly, violent relaxation [5] is likely to play an important role to bring the system close to an equilibrium state in a very short time.

In equilibrium state for self-gravitating systems, the virial theorem dictates that we must have $2K + U = 0$, where K is the kinetic energy and U is the potential energy of the system. This can be rephrased as $2K + U = 2E - U = 0$, and therefore $|E| = GM/2r_{\rm vir}$, where $r_{\rm vir}$ is the radius of the system after it has reached virial equilibrium[3]. Comparing this with the value of energy at the time when the system is at its maximum radius $r_{\rm max}$, we find that $r_{\rm vir} = r_{\rm max}/2$. Thus, the system relaxes to equilibrium at half the radius of maximum expansion and after this we do not expect its size to change, whereas, the universe will continue to expand.

[3] Virial equilibrium is a dynamical equilibrium for a system of particles interacting with each other. It can be shown that, for gravitational interaction, this implies that twice the kinetic energy is equal in magnitude to the potential energy and that these two have opposite signs.

The density contrast for the matter contained within the shell can be written in the following form at any given time:

$$1 + \delta = \frac{\rho}{\bar{\rho}} = \frac{3M/4\pi r^3}{\bar{\rho}}. \tag{8}$$

For the Einstein–deSitter universe, we have:

$$H^2 = \frac{8\pi G}{3}\bar{\rho},$$

where, $H = \dot{a}/a$ is the Hubble parameter. Using the fact that $a \propto t^{2/3}$ in this case, we have $H = 2/3t$. This gives:

$$\bar{\rho} = \frac{1}{6\pi G t^2}. \tag{9}$$

Therefore,

$$1 + \delta = \frac{9GMt^2}{2r^3} = \frac{9}{2}\frac{(\theta - \sin\theta)^2}{(1 - \cos\theta)^3}$$

The perturbation has a density contrast of $9\pi^2/16 - 1 = 4.55$ at the time of maximum expansion.

As discussed above, the perturbation is expected to take one crossing time[4] beyond the stage of maximum expansion to reach close to dynamical equilibrium. Thus, it requires time $t(\theta = \pi)$ to reach virial equilibrium from the stage of maximum expansion. We have already found that the radius at virial equilibrium is half of the radius at maximum expansion. Using these in (8) and (9), we find that the density contrast at the time it reaches equilibrium is 168. Thus, the typical over-density of a collapsed virialised halo is 168, i.e., average density inside virialised halos is 168 times the average density of the universe at the time of collapse. Note that for this second calculation we do not use the last equality, as the parametric solution for r is invalid in this regime. The parameteric solution and the expected behaviour after virialization is illustrated in *Figure* 1.

[4] Crossing time is the time taken by a typical constituent to cross the system under consideration. In the present context, this is equal to the time taken for the shell to go from maximum displacement to the origin.

The density of the halo remains constant once it reaches virial equilibrium but the density contrast continues to increase as the average density of the universe continues to decrease.

The discussion so far has focused on the evolution of the perturbation due to gravitational interaction. It is interesting to ask as to what happens to gases during

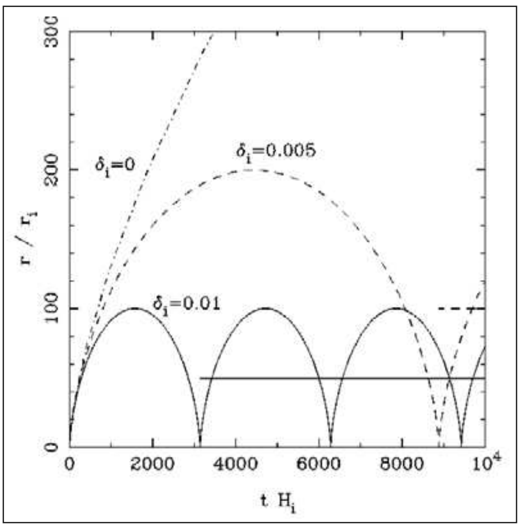

Figure 1. This figure illustrates the parameteric solution for spherical collapse in an Einstein–deSitter universe. The solid curve shows the solution for a shell with an initial density contrast δ_i = 0.01, the dashed curve is for δ_i = 0.005, and, the dot-dashed curve is for the background universe, i.e., for δ_i = 0. It can be seen that a smaller initial density contrast leads to a longer time scale for collapse. The parametric solution implies continued oscillations whereas we expect the collapsing perturbation to reach virial equilibrium. The horizontal lines indicate the expected virial radius for the corresponding curves. These lines start at the time when the halo is expected to virialize.

[5] The perturbation expands from an initial value of r_i to $r_{max} = r_i/\delta_i$. As δ_i<<1, it is clear that r_{max}>>r_i. On the other hand, the shell collapses only by a factor of 2 from r_{max} to r_{vir}. Overall, the effect during gravitational collapse of the system is an expansion. Therefore, we do not expect heating due to compression in gravitational collapse.

the evolution of the perturbation. As the perturbation expands with the universe, clearly gases must cool, and there should be some heating as the perturbation collapses after reaching maximum radius. An implicit assumption here is that the entire evolution is adiabatic and that the gases do not gain or loose energy to any external source or sink. As the expansion of the perturbation is the dominant scaling as compared to compression by a factor of two during collapse, it is clear that the temperature of gases is not very high[5].

If we assume that after collapse, gas inside the virialised halo is supported by thermal pressure, then we can estimate the temperature of gas using the expression for virial equilibrium, and by recognising that the kinetic energy of gas is related to its temperature. If the gas temperature is lower than this virial temperature $T_{\rm vir}$, then the gas must collapse towards the centre of the halo and heat up until the pressure is strong enough to provide the required support. We have:

$$K = \langle \frac{1}{2} m v^2 \rangle = \frac{3}{2} k_{\rm B} T_{\rm vir} = \frac{GMm}{2 r_{\rm vir}}$$

with $k_{\rm B}$ the Boltzmann constant and $T_{\rm vir}$ as the temperature in virial equilibrium for a halo of mass M and radius $r_{\rm vir}$. Thus,

$$T_{\rm vir} = \frac{3GMm}{r_{\rm vir} k_{\rm B}} \simeq 10^7 {\rm K} \left(\frac{M}{10^{12} M_\odot} \right) \left(\frac{r_{\rm vir}}{100 {\rm kpc}} \right)^{-1},$$

where $M_\odot = 2 \times 10^{30}$ is the mass of the Sun and 1 kpc $= 3.08 \times 10^{19}$ m. Thus, the gas in the halo of a collapsed object like our Galaxy is expected to be very hot.

The question that naturally arises is how does the gas heat up to such a high temperature when we expect the infalling gas to be cold? The answer to this puzzle lies in the formation of shocks near the virial radius. The speed of the infalling gas as it nears the virial radius is

$\dot{r} = \sqrt{GM/r_{\rm vir}}$ and for large masses this is significantly higher than the speed of sound. Thus, the infalling gas shocks as it collapses onto material that collapsed earlier and is already in equilibrium. It can be shown that shock heating takes the temperature of the gas fairly close to $T_{\rm vir}$. Using the jump conditions[6] for shock we get [6]

$$T_2 = \frac{3}{16} T_{\rm vir} + \mathcal{O}\left(T_1\right).$$

Here, T_1 is the temperature of infalling gas before it reaches the shock and T_2 is the post-shock temperature. If $T_{\rm vir} \gg T_1$ then we get $T_2 = \frac{3T_{\rm vir}}{16}$ and hence the infalling gas is heated almost to the virial temperature of the halo by shock heating. Other processes may lead to a further increase in the temperature of gas in a halo.

[6] Macroscopic quantities such as density, pressure and temperature are discontinuous across a shock. We can use conservation of mass, momentum and energy as matter crosses the shock to relate the values of density, pressure and temperature across a shock. Such relations are known as junction or jump conditions.

3. Summary

In this article, we have reviewed the spherical collapse model and shown how we may model non-linear gravitational collapse in an expanding universe. We have highlighted the fact that the expansion of the universe slows down the collapse of perturbations by a large factor. Further, we have shown that at least in the Einstein–deSitter universe, the over-density of the collapsing perturbation has characteristic values at the time when it reaches maximum radius and the time when it reaches virial equilibrium.

Using junction conditions for shock heating in the last stages of collapse, we were able to demonstrate that the gas is indeed heated to temperatures in the vicinity of the virial temperature of halos.

As an application of the spherical collapse model, we calculated quantities of interest for an isothermal halo and discussed its implications for the formation of galaxies (see *Box* 2).

Processes follow the same pattern in case of a universe with dark energy, with a well-defined epoch of maxi-

Box 2. Isothermal Sphere

An isothermal sphere is a halo with density profile $\varrho(r) \propto r^{-2}$. Halos with this density profile are of interest as this leads to flat rotation curves. Isothermal halos are also an equilibrium configuration [A, B, C]. Lastly, the comparison with universal density profile derived from cosmological N-Body simulations suggests that halos have this density profile in the vicinity of the virial radius [D], and therefore, we may use it as a simple and useful approximation for analytic estimates.

We may formally write the density of the (singular) isothermal sphere as:

$$\varrho(r) = \varrho(r_{\rm vir}) \left(\frac{r_{\rm vir}}{r}\right)^2 .$$

Here, $r_{\rm vir}$ is the virial radius and $\varrho(r_{\rm vir})$ is the density at this scale. The mass enclosed within the virial radius is:

$$M_{\rm vir} = \int_0^{r_{\rm vir}} 4\pi r^2 \varrho(r) {\rm d}r = 4\pi \varrho(r_{\rm vir}) r_{\rm vir}^3 .$$

The average density within the virial radius is then:

$$\varrho_{av} = \frac{3M_{\rm vir}}{4\pi r_{\rm vir}^3} = 3\varrho(r_{\rm vir}).$$

We have seen in the discussion on virialised halos that the average over-density at the virial radius must be equal to $\Delta = 18\pi^2$ at the time of virialisation in an Einstein–deSitter universe. Thus,

$$\varrho_{av} = \frac{3M_{\rm vir}}{4\pi r_{\rm vir}^3} = 3\varrho(r_{\rm vir}) = \Delta \varrho_c(z_c) = \Delta \varrho_c (1+z_c)^3 .$$

Here, $\varrho_c(z_c)$ is the critical density at the redshift z_c when the halo collapsed and ϱ_c is the critical density at present time. Thus, $M_{\rm vir} = 4\pi\Delta\varrho_c(1+z_c)^3 r_{\rm vir}^3/3$, and $M(r) = 4\pi\Delta\varrho_c(1+z_c)^3 r_{\rm vir}^2 r/3$.

From here, we get the circular velocity as:

$$v_c^2 = \frac{GM}{r} = 4G\pi\Delta\varrho_c(1+z_c)^3 r_{\rm vir}^2/3 = \frac{H_0^2}{2}\Delta(1+z_c)^3 r_{\rm vir}^2 .$$

Halos of a given mass that collapse at higher redshift are more compact and have a higher circular velocity.

[A] J H Jeans, RSPTA, 199, 1, 1902.
[B] S Chandrasekhar, RvMP, 15, 1, 1943.
[C] T Padmanabhan, PhR, 188, 285, 1990.
[D] J F Navarro, C S Frenk, S D M White, *ApJ*, 490, 493, 1997.

mum radius and virialisation. Shock heating is the main mechanism for heating the infalling gas to temperatures near the virial temperature. There are subtle differences however, and we will present a detailed discussion in a follow-up article.

Suggested Reading

[1] C H Lineweaver, L Tenorio, G F Smoot, *et al*, The dipole observed in the COBE DMR 4 year data, *ApJ*., 470, 38, 1996.

[2] M I Scrimgeour, T Davis, C Blake *et al*, The wiggleZ dark energy survey: the transition to large-scale cosmic homogeneity, *MNRAS*, 425, 116, 2012.

[3] G Lemaître, Un univers homogène de masse constante et de rayon croissant rendant compte de la vitesse radiale des nébuleuses extra-galactiques, *ASSB*, 47, 49, 1927.

[4] J E Gunn and J R Gott III, On the infall of matter into clusters of galaxies and some effects on their evolution, *ApJ*, 176, 1, 1972.

[5] D Lynden-Bell, Statistical mechanics of violent relaxation in stellar systems, *MNRAS*, 136, 101, 1967.

[6] L D Landau, E M Lifshitz, *Fluid Mechanics*, Course of Theoretical Physics, Pergamon Press, Oxford, 1959.

Address for Correspondence
J S Bagla
IISER Mohali Knowledge City
SAS Nagar PO Manauli
Punjab 140 306, India.
E-mail:
jasjeet@iisermohali.ac.in

Pritpal Kaur Sandhu
Indian Institute of Technology
Khandwa Road
Indore 452 017, India.
E-mail:
physicspritpal04@gmail.com