6 Large–scale structure

6.1 Introduction

The global properties of the universe – geometry, chemistry and early density fluctuations – are well described by GR, BBN and CMB respectively. The present day universe displays great complexity in terms of large–scale structure and varying galaxy properties (to name just two). Additional techniques are required to extend the standard cosmological model to describe these phenomena and to form a continuous link between the properties of the early universe and the universe we observe today. The goal is to achieve this with the simplest range of physical assumptions.

Observational measures of large–scale structure were first discussed via the galaxy number magnitude, $N(m)$, distribution. The adherence of observations to the relation predicted for a uniformly populated, flat spacetime provided the first evidence that galaxies form a standard universal component. Early, impressive work on measuring the large scale structure of the universe was performed by Shapley and Ames (1932 – 1,200 galaxies), Hubble (1934 – 44,000 galaxies) and Shane and Wirtanen (1954 – 1,000,000 galaxies) – mapping the distribution of galaxies – and by Abell (1958 – 2700 clusters) – mapping the distribution of rich galaxy clusters.

However, it was only during the 1980s that photographic galaxy surveys were combined with spectroscopic studies to create large galaxy redshift surveys from which quasi three dimensional maps of the universe could be created. The pioneer in this field was the CfA Redshift Survey (containing some 3000 galaxy redshifts out to a redshift $z = 0.04$). Already the void and filament pattern observed in this survey placed strong constraints upon HDM models – as discussed in the previous lecture, HDM could not explain the fine structures being observed. The current state–of–the–art is the SDSS which will accumulate 1,000,000 galaxy redshifts out to redshifts $z < 0.3$ and obtain photometry for up to 10^8 galaxies. These observational resources provide the raw material for quantitative models of large scale structure.

One concept that should be emphasized is the isotropy of large–scale structure. Viewed within a narrow angle the universe shows marked density variations as a function of line–of–sight. However, when one considers wider regions of the universe, e.g. corresponding to the Hubble volume, one patch of the universe looks very much like another when subject to a wide range of statistical measures. It is this observations prompts the statement that the universe is isotropic (and therefore homogeneous) on large scales.

The aim of this lecture is to describe the theoretical steps that provide the link between the current cosmological model (GR, BBN, CMB) and observed structure in the present day universe, and the observational tools employed to quantify our description of this structure.

6.2 The growth of density fluctuations in an expanding universe

In a matter dominated universe We have so far considered the expansion dynamics to be governed by the average mass density, i.e. $\bar{\rho}(t) \propto a(t)^{-3}$. We now consider the dynamics of a small[1], spherical region of the universe of local mass density

$$\rho(t) = \bar{\rho}(t)[1 + \delta(t)], \tag{1}$$

with $\delta(t) \ll 1$. The dimensionless density perturbation δ can be positive or negative and represents either a slightly over or underdense region of the universe. Consider the equation of motion of the radius of the spherical region, i.e.

$$\ddot{R} = -\frac{GM}{R(t)^2} = -\frac{G}{R(t)^2}\left(\frac{4\pi}{3}\rho(t)R(t)^3\right) = -\frac{4\pi}{3}G\bar{\rho}(t)R(t)[1 + \delta(t)]. \tag{2}$$

or

$$\frac{\ddot{R}}{R} = -\frac{4\pi}{3}G\bar{\rho}(t)[1 + \delta(t)]. \tag{3}$$

In addition to the equation of motion of the sphere we can also apply the condition that the total mass contained within the sphere is constant, i.e.

$$\begin{aligned} M &= \frac{4\pi}{3}\bar{\rho}(t)[1 + \delta(t)]\, R(t)^3 = \text{constant}. \\ \text{i.e.} \quad R(t) &\propto \bar{\rho}(t)^{-1/3}[1 + \delta(t)]^{-1/3} \\ R(t) &\propto \bar{a}(t)[1 + \delta(t)]^{-1/3}. \end{aligned} \tag{4}$$

Which provides a basic mass continuity relation. Therefore, we note that

$$\begin{aligned} \delta > 1 &\Rightarrow \text{overdense} \Rightarrow \text{reduced expansion rate} \\ \delta < 1 &\Rightarrow \text{underdense} \Rightarrow \text{increased expansion rate}. \end{aligned}$$

Taking the second derivative of the mass continuity relation and neglecting 2nd order terms we obtain

$$\begin{aligned} \dot{R} &\propto \dot{a}[1 + \delta(t)]^{-1/3} - a\frac{\dot{\delta}}{3}[1 + \delta(t)]^{-4/3} \\ \ddot{R} &\propto \ddot{a}[1 + \delta(t)]^{-1/3} - 2\dot{a}\frac{\dot{\delta}}{3}[1 + \delta(t)]^{-4/3} - a\frac{\ddot{\delta}}{3}[1 + \delta(t)]^{-4/3} + a\frac{4\dot{\delta}^2}{3}[1 + \delta(t)]^{-7/3} \\ \frac{\ddot{R}}{R} &= \frac{\ddot{a}}{a} - \frac{2}{3}\frac{\dot{a}}{a}\dot{\delta} - \frac{\ddot{\delta}}{3}, \end{aligned} \tag{5}$$

[1]How small is small? Smaller than the horizon scale but larger than the Jeans length (see later).

Recalling that $|\delta| \ll 1$ and neglecting the term $\dot{\delta}^2$. Combining Equations 3 and 5 gives

$$\frac{\ddot{a}}{a} - \frac{2}{3}\frac{\dot{a}}{a}\dot{\delta} - \frac{\ddot{\delta}}{3} = -\frac{4\pi}{3}G\bar{\rho}(t)[1+\delta(t)]. \tag{6}$$

Setting $\delta(t) = 0$ gives

$$\frac{\ddot{a}}{a} = -\frac{4\pi}{3}G\bar{\rho}(t), \tag{7}$$

which the Acceleration equation for a homogeneous universe. To obtain the time evolution of density perturbations in an expanding universe, we subtract the homogeneous component of the acceleration from Equation 6 to obtain the non-homogeneous component, i.e.

$$\begin{aligned} -\frac{2}{3}\frac{\dot{a}}{a}\dot{\delta} - \frac{\ddot{\delta}}{3} &= -\frac{4\pi}{3}G\bar{\rho}(t)\delta(t) \\ \ddot{\delta} + 2H\dot{\delta} &= 4\pi G\bar{\rho}(t)\delta(t). \end{aligned} \tag{8}$$

Setting $H = 0$ defines the static case and we obtain

$$\ddot{\delta} = 4\pi G\bar{\rho}(t)\delta(t), \tag{9}$$

i.e. the overdensity increases as the sphere collapses (for $\delta > 1$). In the case of an expanding universe the term $2H\dot{\delta}$ represents the "Hubble drag" terms that describes the fact that with time, the apparent expansion velocity of the universe dilutes the relative contribution of any peculiar motion. The right hand side of Equation 8 can be re-written noting that

$$4\pi G\bar{\rho}(t) = \frac{3}{2}\Omega_M\, H^2, \tag{10}$$

to obtain,

$$\ddot{\delta} + 2H\dot{\delta} - \frac{3}{2}\Omega_M\, H^2\, \delta = 0, \tag{11}$$

where Ω_M describes the normalised matter density of the universe. However, we must recall that the term $H(t)$ is determine the total mass-energy density of the universe. At early and late times respectively, the Ω_M term will not dominate the expansion . Consider a radiation dominated universe at early times, i.e. $\Omega_M \ll 1$ and $H = 1/2t$. We can therefore re-write Equation 11 as

$$\ddot{\delta} + \frac{1}{t}\dot{\delta} \approx 0, \tag{12}$$

which has the solution

$$\delta(t) \approx B_1 + B_2 \ln t, \tag{13}$$

i.e., the density perturbation grow according to a slow logarithmic term as a function of time. At late cosmic times$\Omega_M \ll \Omega_\Lambda$ and $H \to H_\Lambda$. Therefore

$$\ddot{\delta} + H_\Lambda \dot{\delta} \approx 0 \tag{14}$$

which has the solution

$$\delta(t) \approx C_1 + C_2 e^{-2H_\Lambda t}, \tag{15}$$

i.e., density fluctuations are effectively frozen at the amplitude when dark energy begins to dominate the expansion history. This is effectively due to the dominance of the Hubble drag term. If we now consider a matter dominated, Einstein-de Sitter universe[2], i.e. $\Omega_M = 1$ and $H = 2/3t$ we can write

$$\ddot{\delta} + \frac{4}{3t}\dot{\delta} - \frac{2}{3t^2}\delta = 0. \tag{16}$$

We guess that a power law solution of the form $\delta = Dt^n$ is required and via substitution obtain

$$\begin{aligned} n(n-1)Dt^{n-2} + \frac{4}{3}nDt^{n-1} - \frac{2}{3t^2}Dt^n &= 0 \\ n(n-1) + \frac{4}{3}n - \frac{2}{3} &= 0, \end{aligned} \tag{17}$$

which describes a quadratic equation with the solutions $n = 2/3$ and $n = -1$ – respectively referred to as the growing and decaying mode. With increasing time the growing mode dominates over the decaying mode and the time evolution of the density perturbations revert to

$$\delta(t) \propto t^{2/3} \propto a(t) \propto \frac{1}{1+z}. \tag{18}$$

We must remember that the preceding analysis is valid in the **linear regime** only, i.e. $\delta \ll 1$. Small density fluctuation therefore grow linearly with redshift until the assumption $\delta \ll 1$ no longer holds and the instability reaches the non–linear or collapse phase.

This equation provides an immediate requirement for an additional, non–coupled matter field within the universe – dark matter. Consider the typical scale of temperature fluctuations at the epoch of recombination, $\Delta T/T \sim 10^{-5}$. These temperature fluctuations are associated with matter fluctuations of similar amplitude and were fixed by the physics of the photon–baryon fluid until the epoch of recombination. These fluctuations would have grown by a factor 1000 between the epoch of recombination and the present day, reaching a level $\delta \sim 10^{-2}$ today. This is clearly not consistent with observations of large density contrasts ($\delta \sim 100$s) in the local universe. Non–linear collapse could only have occurred if a large fraction of the available matter was not coupled to the radiation field and entered the linear growth phase at some earlier epoch.

[2]This can be thought of as the "Goldilocks" model – conditions are "just right" for structure formation to proceed.

6.3 Statistics of the density field - the power spectrum

Any attempt to describe of the physics of structure formation is concerned with the statistical, rather than the individual, properties of the distribution of density perturbations. We can expand the dimensionless density enhancement at some position $\vec{r}$ as a three dimensional Fourier series, i.e.

$$\delta(\vec{r}) = \frac{V}{(2\pi)^3} \int \delta_{\vec{k}} e^{-i\vec{k}.\vec{r}} \mathrm{d}^3k, \tag{19}$$

where the individual Fourier coefficients are expressed as

$$\delta_{\vec{k}} = \frac{1}{V} \int \delta(\vec{r}) e^{i\vec{k}.\vec{r}} \mathrm{d}^3r. \tag{20}$$

Expanding the density field into its Fourier components reduces the description of large scale structure to a series of oscillating density waves of amplitude $\delta_{\vec{k}}$ and co-moving wavenumber $\vec{k}$ and wavelength $\lambda = 2\pi/k$. Strictly speaking each Fourier component is a complex number and can be written in the form

$$\delta_{\vec{k}} = |\delta_{\vec{k}}|\, e^{i\phi_{\vec{k}}}, \tag{21}$$

where $\phi_{\vec{k}}$ is a phase constant. For $|\delta_{\vec{k}}| \ll 1$ each wavemode evolves independently according to

$$\ddot{\delta}_{\vec{k}} + 2H\dot{\delta}_{\vec{k}} - \frac{3}{2}\Omega_M\, H^2\, \delta_{\vec{k}} = 0, \tag{22}$$

for wavemodes of proper wavelength $\lambda = a(t)2\pi/k$ greater than the Jeans length and less than the horizon scale at each epoch. The phase term remains constant as long as the density perturbation at each wavemode remains small. The independence of each wavemode means that even after short wavelength modes have reached $|\delta_{\vec{k}}| \sim 1$ and begun to collapse, longer wavelength modes can still evolve linearly according to Equation 22. In practical terms this means that large scale structures can still be studies using linear theory even after smaller scale structures such as galaxies and clusters have collapsed.

The **power spectrum** is defined as the mean square of the Fourier components, i.e.

$$P(k) = \langle |\delta_{\vec{k}}|^2 \rangle. \tag{23}$$

When the phases of each Fourier component are uncorrelated with each other then the physical density field $\delta(\vec{r})$ is said to be Gaussian. If the Gaussian field is also homogeneous and isotropic then the power spectrum provides a complete statistical description of the density field. Furthermore, if $\delta(\vec{r})$ is a Gaussian field then the probability of a density perturbation δ existing at some random location is given by the Gaussian distribution, i.e.

$$p(\delta) = \frac{1}{\sqrt{2\pi}\sigma} \exp\left(-\frac{\delta^2}{2\sigma^2}\right). \tag{24}$$

where the standard deviation is computed from the power spectrum

$$\sigma = \frac{V}{(2\pi)^3}\int P(k)\mathrm{d}^3k = \frac{V}{2\pi^2}\int P(k)\,k^2\,\mathrm{d}k. \tag{25}$$

The study of such Gaussian density fields is important as most inflationary models of the early universe predict a homogeneous and isotropic Gaussian density field. In addition the inflationary models of the universe predict that the power spectrum will take a simple power law form, i.e.

$$P(k) \propto k^n, \tag{26}$$

with the favoured value of $n = 1$ being proposed by Harrison and Zeld'ovich in 1970. The Harrison Zeld'ovich power spectrum is said to be *scale invariant* but what is the physical implication of this statement?

Consider a sphere of comoving radius L placed at some random location in the universe. The mass of the sphere is

$$M = \frac{4\pi}{3}L^3\bar{\rho}[1+\delta]. \tag{27}$$

Note that here δ (no subscript) refers to a real space density fluctuation. The expectation value of the mass contained in an ensemble of such spheres placed at random locations is

$$\langle M\rangle = \frac{4\pi}{3}L^3\bar{\rho} \tag{28}$$

The mean square mass fluctuation associated with the population of spheres may be written as

$$\left\langle\left(\frac{M-\langle M\rangle}{\langle M\rangle}\right)^2\right\rangle = \langle\delta^2\rangle \propto k^3P(k), \tag{29}$$

where the wavenumber associated with the sphere is $k = 2\pi/L$. Note that we will justify this statement of proportionality when we demonstrate that the real space correlation function is the Fourier transform of the volume averaged power spectrum. The root mean square deviation is defined via

$$\frac{\delta M}{M} \equiv \left\langle\left(\frac{M-\langle M\rangle}{\langle M\rangle}\right)^2\right\rangle^{1/2} \propto k^{(3+n)/2} \propto L^{-(3+n)/2} \propto M^{-(3+n)/6}, \tag{30}$$

where we have substituted $P(k) \propto k^n$ and followed sensible changes of variable. If the power spectrum is described by a power law index $n < -3$ then the amplitude of mass fluctuations diverges on the largest scales which would invalidate the observed isotropy if the universe on very large scales (the observed universe is effectively uniform on scales $\lambda > 100$ Mpc or $k < 0.01$ Mpc^{-1}). As an alternative, if we population the universe with a distribution of point masses and described fluctuations via Poisson statistics we would obtain

$$\frac{\delta M}{M} \propto N^{-1/2} \propto M^{-1/2}, \tag{31}$$

noting that the number of particles within the sphere is N and that the mass of the sphere $M \propto N$. This distribution of fluctuations corresponds to $n = 0$. Therefore, the Harrison Zeld'ovich spectrum produces greater statistical power on small scales (greater mass fluctuations at low mass) than a Poisson distribution of points. The real simplicity of the Harrison Zeld'ovich spectrum becomes clear when one considers that the potential fluctuations associated with a length scale L are

$$\delta\Phi \sim \frac{G\delta M}{L} \propto \frac{\delta M}{M^{1/3}} \propto M^{(1-n)/6}, \tag{32}$$

i.e. the case $n = 1$ ensures constant potential fluctuations on all scales.

6.4 The transfer function – the growth of structure in the linear regime

The evolution of the density power spectrum from its initial state is describe by a transfer function of the form

$$P(k, z_f) = T(k, z_f, z_i, U)\, P(k, z_i), \tag{33}$$

where $U = \Omega_b, \Omega_{DM}, \Omega_\Lambda, \Omega_R, H_0$ describes the 'background' cosmology. In the limit of small density perturbations each wavemode evolves independently.

The transfer function attempts to describe the growth of structure as a function of both scale and universal epoch. It includes many physical effects but is limited to the linear regime, i.e. small density perturbations. The transfer function is dominated by two considerations; the relative scale of a given perturbation compared to the horizon at that epoch, and whether the universe is radiation or matter dominated at that epoch. Major physical effects described within the transfer function include:

1. Meszaros effect: the rapid expansion of the universe at early times (during the radiation dominated epoch), 'freezes' the growth of structure on scales smaller than the horizon. Put another way, the Hubble Drag term dominates and density fluctuations smaller than the horizon scale are effectively fixed. Fluctuations outside the horizon grow normally and the result is the relative suppression of structure smaller than a characteristic scale – given by the horizon at matter–radiation equality.

2. Free–streaming of relativistic dark matter particles (as discussed in the previous lecture): relativistic DM particles 'free–stream' with the growing horizon and damp density density fluctuations on the horizon scale. This damping effect ceases when the DM particles become non–relativistic. Free–streaming therefore damps structure on scales smaller than the horizon at relativistic decoupling.

3. Acoustic horizon length: structure is further modified by pressure oscillations within the photon–baryon fluid. Structures displaying wavemodes that are an integer fraction of the sound horizon at the epoch of decoupling form a harmonic series of density enhancements.

These structures form a series that has had just enough time to compress, compress–rarify, compress–rarify–compress (etc.) once during the age of the universe. This effect dominates form of the temperature correlation function observed in the CMB – in effect a 'snapshot' of the transfer function.

4. Silk Damping: photons within the photo–baryon fluid travel a given 'random walk' distance before scattering. This results in a diffusion of photons from density fluctuations and the suppression of structure on scales smaller than the random walk distance. This distance is $20 - 30$ Mpc at the epoch of decoupling and is comparable to the horizon scale.

5. The Jeans scale: the Jeans relation describes baryonic structures whose self–gravity is supported by internal pressure. Structures that are Jeans stable do not collapse and structure is suppressed on scales smaller than the Jeans scale. Consider a virialised cloud of particles at the stability limit, i.e.

$$3NkT = \frac{GM_J^2}{R_J}, \tag{34}$$

where subscript J indicates the Jeans limit. Employing the relations $N = M_J/\mu m_p$ and $M_J = 4\pi\rho R_J^3/3$ one obtains the Jeans length

$$R_J = \left(\frac{3kT}{\mu m_p}.(G\rho)^{-1}.\frac{3}{4\pi}\right)^{1/2} = c_s\sqrt{\frac{\pi}{G\rho}}. \tag{35}$$

The Jeans length is therefore proportional to the sound speed of baryons at each epoch. Prior to decoupling the baryonic sound speed was determined by the properties of the relativistic photon baryon fluid, i.e. $c_s = c/\sqrt{3} = 0.58c$. The corresponding value of the Jeans length may be computed as

$$R_J = 0.58c \times \sqrt{\frac{\pi}{6.67\times10^{-11}\ \mathrm{Nm^2kg^{-2}} \times 5\times10^{-19}\ \mathrm{kmm^{-3}}}} = 5.3\times10^{22}\ \mathrm{m} \approx 1.6\ \mathrm{Mpc}, \tag{36}$$

where we have taken the physical baryon density at the epoch of decoupling. The associated Jeans mass is of order 10^{20} $\mathrm{M}_\odot$ which is several orders of magnitude larger than the most massive galaxy cluster known. Immediately after decoupling the sound speed is given by the expression

$$c_s \approx \sqrt{\frac{kT}{\mu m_p}} = \sqrt{\frac{kT}{\mu m_p c^2}}c = \sqrt{\frac{0.26\ \mathrm{eV}}{563\ \mathrm{MeV}}}c = 2\times10^{-5}c. \tag{37}$$

The corresponding Jeans length and mass are respectively $R_J = 1.84\times10^{18}$m ≈ 50 pc and $M_J \approx 4\times10^6$ $\mathrm{M}_\odot$ which lies between the mass of a typical globular cluster and a dwarf galaxy. Therefore, the epoch of decoupling marks an important transition in the development of baryonic structure. Prior to decoupling all fluctuation scales up to the horizon scale are effectively suppressed. After decoupling the Jeans scale is dramatically reduced and baryonic structures can begin to grow.

6. Fitting formulae: for a given cosmology and assumed DM properties, one may compute the transfer function as a function of k at any given epoch. The transfer function is approximated to a fitting formula in order to express the complex physical process via a single function. Such expressions are good to within 10%. The recent work of Eisenstein and Hu (1998) on the contribution of baryon oscillations to the transfer function has further improved the accuracy.

6.5 Non–linear collapse: The isolated spherical halo

We now return to the evolution of a density perturbation – recall that we had previously described the growth of such perturbations in the linear regime, i.e. $\delta \ll 1$. To follow the collapse of an isolated, spherical overdensity beyond the linear phase, we return to the exact equation of motion of a test particle at the edge of a spherical sub-region of the universe.

$$\begin{aligned} \frac{d^2r}{dt^2} &= -\frac{GM}{r^2} \\ \left(\frac{dr}{dt}\right) &= \frac{2GM}{r} + C. \end{aligned} \tag{38}$$

We consider the case $C < 0$, i.e. a bound, or closed, system. The exact solution takes the form

$$\begin{aligned} r &= A(1-\cos\theta) \\ t &= B(\theta - \sin\theta) \text{ for } \theta = 0, 2\pi. \end{aligned} \tag{39}$$

We have

$$\begin{aligned} A &= \frac{\Omega_{m0}}{2(\Omega_{m0}-1)} \\ B &= \frac{\Omega_{m0}}{2H_0(\Omega_{m0}-1)^{3/2}}. \end{aligned} \tag{40}$$

We also note that $A^3 = GMB^2$ and that $C = -A^2/B^2$. To better understand the relationship between these constants start with

$$\begin{aligned} A^3 &= GMB^2 \\ \frac{\Omega_{m0}^3}{8(\Omega_{m0}-1)^3} &= \frac{GM\Omega_{m0}^2}{4H_0^2(\Omega_{m0}-1)^3} \\ \Omega_{m0} &= \frac{2GM}{H_0^2} = \frac{8\pi G}{3H_0^2}\cdot r^3(t)\rho(t) = \frac{\rho_0}{\rho_{c,0}}\cdot r_0^3. \end{aligned} \tag{41}$$

Note that, as with a_0, we are free to choose $r_0 = 1$. The spherical overdensity expands until $\theta = \pi$ and $r_{max} = 2A$ then "turns around" and begins to collapse. Formal collapse occurs at $\theta = 2\pi, r = 0, t = 2\pi B$. At turnaround, or maximum expansion, we have

$$r_{max} = 2A = \frac{\Omega_{m0}}{\Omega_{m0} - 1} \tag{42}$$

$$t_{max} = \pi B = \frac{\pi \Omega_{m0}}{2H_0(\Omega_{m0} - 1)^{3/2}} \tag{43}$$

What is the density of the spherical region at turnaround compared to an idealised EdS background universe?

$$\frac{\rho}{\rho_c} = \frac{\Omega_{m0}\, \rho_{c0} r_{max}^{-3}}{\rho_{c0}\, a^{-3}} = \Omega_{m0} \left(\frac{a}{r_{max}} \right)^3 . \tag{44}$$

For an EdS universe we have $a = (\frac{3}{2} H_0 t)^{2/3}$.

$$a^3(t_{max}) = \left(\frac{3}{2} H_0 t_{max} \right)^2 = \frac{9\pi^2}{16} \frac{\Omega_{m0}^2}{(\Omega_{m0} - 1)^3}, \tag{45}$$

using the above expression for t_{max}. Therefore

$$\begin{aligned} \frac{\rho}{\rho_c} &= \Omega_{m0} \frac{9\pi^2}{16} \cdot \frac{\Omega_{m0}^2}{(\Omega_{m0} - 1)^3} \cdot \frac{(\Omega_{m0} - 1)^3}{\Omega_{m0}^3} \\ &= \frac{9\pi^2}{16} = 5.55, \end{aligned} \tag{46}$$

using the appropriate expression for r_{max}. So the spherical overdensity starts to collapse when its density has reached 5.55 times the background radius (linear theory gives a value of 1.06). Following turnaround, one might naively expect formal collapse to occur at $r = 0, t = 2\pi B$. But if we follow the equations to this point we expect $\rho \to \infty$ which is physically unrealistic.

6.6 Towards a virialised halo

Virialisation describes the process by which the gravitational collapse reaches an equilibrium state described by the virial equation, i.e. $P = -2K$, $E = P + K = -K$ and where the mean spatial distribution of particles is constant with time (strictly speaking $\ddot{I} = 0$).

How does this occur? Even if the dark matter (DM) were completely collisionless and collapsing on radial trajectories, each DM particle would oscillate about the halo centroid with an amplitude equal to the turnaround radius. However, infall trajectories are not purely radial – DM particles are subject to torques via gravitational interactions with nearby large-scale structure and gain angular momentum as a result. Also the form of the DM power spectrum indicates that DM structure is

present on all scales. Although DM is thought to be collisionless, such DM sub-structures can lose gravitational potential energy via dynamical friction with the parent halo and spiral to the centre.

The combination of these effects lead in simulations fo the formation of a collapsed halo largely supported by random particle motions (pressure support) with a small net rotation. Such simulations show that the halo mass is centrally concentrated according to a density function $\rho(r)$. Energy equipartition among DM particles generate an isothermal distribution where $\rho \propto r^{-2}$. This gives $M \propto r$ and $v^2 = GM/r =$ constant.

Simulations of collisionless DM particles in an expanding universe indicate a different distribution, first presented by Navarro, Frenk and White (1997) as

$$\rho \propto \frac{1}{r(r+r_S)^2}, \tag{47}$$

where r_S is a scale radius. Therefore $\rho \propto r^{-\gamma}$ where

$$\gamma \to -1 \quad \text{at} \quad r \ll r_S$$

$$\gamma \to -3 \quad \text{at} \quad r \gg r_S \tag{48}$$

with $\gamma = -2$ occurring at r somewhat greater than r_S.

Although virialisation is complex and non–linear, we can understand the properties of the virial halo which results using energy arguments. What is the halo overdensity at virialisation? At turnaround the kinetic energy is zero. Therefore $P = E$. At virial equilibrium we have $P = -2K$ and $E = P + K = -K$. Therefore $P = 2E$. As $P \propto GM/r$ the virial radius must be one-half of the turnaround (maximum) radius. As $r_{max} = 2A$, $r_{virial} = A$. At this point the overdensity is $2^3 = 8$ times more dense than at turnaround. When does this occur? We have

$$\begin{aligned} r_{virial} &= \frac{1}{2} r_{max} \\ A(1-\cos\theta) &= A \\ \cos\theta &= 0 \\ \theta &= \frac{3\pi}{2} \end{aligned} \tag{49}$$

noting that we require $\theta > \pi$ (virialisation must come after turnaround). The corresponding time is

$$t = B(\theta - \sin\theta)$$

$$\begin{aligned} t_{virial} &= B\left(\frac{3\pi}{2}+1\right) \\ &= \pi B\left(\frac{3}{2}+\frac{1}{\pi}\right) \\ &\approx 1.81 \times t_{max}. \end{aligned} \tag{50}$$

The additional factor we must consider is that in the time interval between t_{max} and t_{vir} the density of the background EdS universe has decreased by a factor

$$\left(\frac{a_{max}}{a_{virial}}\right)^3 = \left(\frac{t_{max}}{t_{virial}}\right)^2 = \frac{1}{1.81^2}. \tag{51}$$

Therefore, at virialisation the spherical region has a density relative to the background of $5.55 \times 8 \times 1.81^2 \approx 145$. Note that if – as many sources do – you take t_{virial} to occur when $r = 0$, you obtain $t_{virial} = t(\theta = 2\pi) = 2\,t_{max}$. In this case, the relative overdensity at virialisation is $5.55 \times 8 \times 2^2 \approx 178$, which is further approximated to $\delta = \frac{\rho-\bar{\rho}}{\bar{\rho}} = 200$ in textbooks and papers.

In general it is reasonable to expect a collapsing overdensity to achieve virial equilibrium when it becomes $\approx 150 - 200$ times the background density. This simple model is in agreement with the results of simulations which indicate that at $\delta > 100$ galaxies and clusters of galaxies emerge as distinct, gravitationally-bound structures.

We can also estimate the redshift at which virialisation occurs in a general matter-dominated universe, assuming that the same results as above remain approximately true even when the background density differs from critical. We denote the present-day normalised background density as Ω_{m0}^{bg}. The approximate virialisation condition is

$$\rho_{virial} \geq 100 \times \frac{3\Omega_{m0}^{bg} H_0^2}{8\pi G}(1+z_{virial})^3. \tag{52}$$

We assume that the virialised halo is characterised by a velocity dispersion σ such that

$$\frac{1}{2}M\sigma^2 = \frac{1}{2}\frac{GM^2}{r} \tag{53}$$

which gives

$$r = \frac{GM}{\sigma^2}. \tag{54}$$

The density of the spherical region at virialisation is then

$$\rho_{virial} = \frac{M}{\frac{4\pi}{3}r^3} = \frac{M}{\frac{4\pi}{3}\frac{GM}{\sigma^2}} = \frac{3\sigma^2}{4\pi G^3 M^2} \tag{55}$$

and the condition for virialisation becomes

$$\frac{3\sigma^2}{4\pi G^3 M^2} \geq 100 \times \frac{3\Omega_{m0}^{bg} H_0^2}{8\pi G}(1+z_{virial})^3. \tag{56}$$

This can be rearranged to yield

$$1 + z_{virial} \lesssim 0.275 \times \sigma^2 (GM)^{-2/3} (\Omega_{m0}^{bg} H_0^2)^{-1/3} \tag{57}$$

which scales as

$$1 + z_{virial} \lesssim 0.6 \times \left(\frac{\sigma}{100\,\mathrm{kms}^{-1}}\right)^2 \left(\frac{M}{10^{12} M_\odot}\right)^{-2/3} (\Omega_{m0}^{bg} h^2)^{-1/3}. \tag{58}$$

This is an approximate result but it gives some insight into when structures of different properties formed. Take $\Omega_{m0}^{bg} h^2 = 0.13$ as a value representative of our universe. A large galaxy like the Milky Way has $\sigma \sim 300\,\mathrm{kms}^{-1}$ and $M \sim 10^{12} M_\odot$, and we find $1 + z_{virial} \leq 11$. Thus we would be surprised to discover a fully-formed galaxy at $z \gg 10$. Massive galaxy clusters have $\sigma \sim 1,000\,\mathrm{kms}^{-1}$ and $M \sim 10^{15} M_\odot$, which gives $1 + z_{virial} \leq 1.2$ and indicates that such structures formed only recently within our universe.

6.7 Non–linear collapse: N–body simulations

Current simulations of structure formation follow the evolution of density fluctuations from the linear to non–linear (or collapse) phase via direct representation of the matter density field as an N–body system of particles. The aim of all such models is to follow the distribution of density fluctuations generated by the combination of the initial conditions and the linear growth regime described in previous sections to the regime where $\delta \gg 1$ and the local gravitational interaction of dark matter clumps becomes important.

The equation of motion of an individual density fluctuation relates the velocity to a combination of the Hubble Drag and the local gravitational field. Solution of this equation therefore reduces to solving the local Poisson equation for a system of particles, i.e.

$$\frac{\mathbf{g}}{a} = -\frac{\nabla\Phi}{a^2} = G\sum_i m_i \frac{\mathbf{x}_i - \mathbf{x}}{|\mathbf{x}_i - \mathbf{x}|^3}. \tag{59}$$

Direct computation is expensive. Therefore, to push such calculations to incorporate larger numbers of particles for a given computational effort, the local Poisson equation may be expressed via the Fourier transform of the velocity field and the equation of motion is solved in Fourier space. This is the idea underlying Particle Mesh (PM) N–body codes: the density field described by the system of particles is averaged onto a grid and a FFT is employed to compute the real space force components resulting from this density field. Such limits are obviously limited by the mesh dimension which

translates directly into spatial resolution. In addition, the larger the mesh, the more massive each "mesh particle" – such particles represent a collection of less massive particles. Such massive virtual particles suffer "close" interactions and wide angle scattering more than less massive counterparts and thus crude mesh simulations can lead to unrealistic collapse models. The solution is to refine the treatment of the mesh, i.e.

1. Particle–Particle–Particle–Mesh (P^3M) codes: direct forces are computed between all particles in adjacent cells with the FFT of the grid being employed to compute the contribution from more distant mesh elements.

2. Adaptive mesh codes: the mesh size is flexible and varies such that higher density regions are mapped by a finer mesh to preserve the number of particles per mesh within reasonable limits.

Employing such procedures it is possible to create a three dimensional map of the dark matter distribution as a function of epoch – the current state–of–the–art is to present universal "cubes" of up to $100h^{-1}$ Mpc on a side versus redshift based upon systems of 10^8 to 10^9 particles. However, additional refinements are required to translate such representations into a map of the visible large–scale properties of the present universe, i.e. the distribution of baryonic matter:

1. N–body simulations on a number of physical scales. The large scale properties of the universe can be described by large simulations mentioned above. However the mesh size represents a limiting physical scale. Additional simulations are therefore required to trace the evolution of successively smaller (i.e. less massive) structures in the universe, e.g. groups of galaxies to individual galaxies (i.e. an accreting system of halos).

2. Gas physics: visible structures require star formation. To form stars one must channel baryons to the central regions of dark matter potentials where the baryon density increases to levels where fragmentation and collapse occurs. An additional model is then required to determine at what rate the collapsing gas is transformed into stars and how this rate varies as a function of time. Finally, a model is required to describe the evolution of the stellar population as a function of time and its luminosity weighted spectral energy distribution – finally providing the observed brightness and colour of the model galaxy population. Such prescriptions are very model dependent and are often referred to a "semi–analytic" models of galaxy formation.

In summary therefore, the large scale distribution of visible structure may be related to the initial conditions of the HBB model via the following ideas:

1. Assume an initial spectrum of density fluctuations – Harrison–Zeld'ovich.

2. Describe the linear growth of density fluctuations via the Transfer function.

3. Upon reaching the limiting accuracy of the assumption of linear growth, feed the density distribution into the initial conditions of an N–body simulation of collapsing structure.

4. Apply a model for the transport of baryonic gas within collapsing structures. This can be as simple as computing the thermal energy of gas within a given potential and relating this to the cooling properties of the gas at that temperature and density (Figure 1). We can relate the thermal energy contained in the baryonic gas to their kinetic energy associated with the orbital motion in the dark matter halo, i.e.

$$\frac{3}{2}kT = \frac{1}{2}\mu m_p \langle v^2 \rangle = \frac{3}{2}\mu m_p \sigma_r^2, \tag{60}$$

i.e.,

$$T = \left(\frac{\mu m_p}{k}\right)\sigma_r^2 = \left(\frac{\mu m_p c^2}{k}\right)\left(\frac{\sigma_r}{c}\right)^2. \tag{61}$$

Inputting velocity dispersion values representative of massive galaxy clusters ($\sigma_r = 1000\,\mathrm{kms}^{=1}$) and typical galaxies ($\sigma_r = 100$ kms$^{=1}$) generates gas temperatures of order 7×10^7 K and 7×10^5 K respectively. As the cooling timescale is $t_{cool} \sim kT/\Lambda(T)$ one can see that galaxies occupy a region of the cooling curve where $\Lambda(T)$ satisfies $\mathrm{d}\Lambda/\mathrm{d}T < 0$ that the cooling time decreases with decreasing temperature – the gas contained within the potential can cool effectively and the baryons will collapse. Galaxy clusters occupy a region of the cooling curve where $\Lambda(T)$ satisfies $\mathrm{d}\Lambda/\mathrm{d}T > 0$ that the cooling time increases with decreasing temperature – the gas contained within the potential cannot cool effectively and the baryons are prevented from collapsing. This is why most of the baryons in galaxies are in the form of stars whereas the majority of baryons in galaxy clusters are in the form of diffuse, hot gas.

5. Apply a model for the rate of star formation in cool gas. Advanced models describe the distribution of cooling gas as a function of time within given halos to predict star formation (resulting in abundance and colour) gradients within individual classes of galaxy.

6. Apply a model to describe the evolution and appearance of the stellar population thus created as a function of time, e.g. Bruzual and Charlot.

This progression of ideas has come to be termed the "Hierarchical" model of galaxy and structure formation, i.e. structure forms in a hierarchical (bottom–up) fashion with baryon evolution described by semi–analytic approximations.

6.8 The limit of gravitational physics: Scaling relations of galaxy clusters

Our analytical description of the growth of structure in the universe has been naturally dominated by gravitational physics. Simulations of our universe attempt to describe the properties of visible

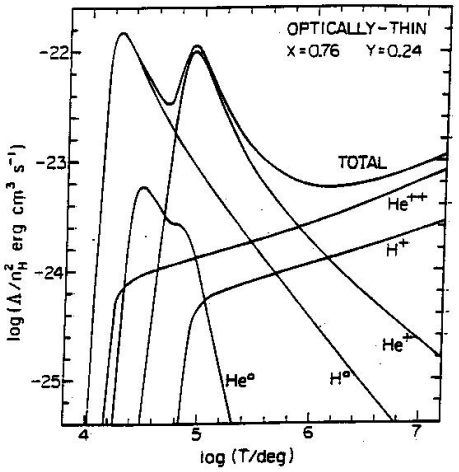

Figure 1: The cooling rate divided by the baryon density as a function of temperature for primordial gas.

structures by combining gravitational physics with baryonic processes leading either to star formation (in the case of galaxies) or to emission from hot gas (in the case of galaxy clusters). However, computer simulations are limited by the fact that they cannot simultaneously resolve the physical scales on which gravitational processes dominate and those on which baryonic processes dominate – hence the need for simple analytic rules to describe the unresolved baryons.

It is interesting to see how far we can proceed analytically from our gravitational description of virialised halos to a description of the observed properties of such structures. We consider the observed properties of galaxy clusters.

We have already seen that galaxy clusters may be described as an approximately spherical overdensity characterised by a velocity dispersion $\sigma = 300 - 1000\,\mathrm{kms}^{-1}$ and mass $M = 10^{14} - 10^{15}\,\mathrm{M}_{\odot}$. Approximately 5/6 of the mass is in the form of dark matter, 1/6 is in the form of hot, diffuse baryons and the remainder (approximately 1%) is in the form of galaxies. The observed properties of galaxy clusters are thus in principle dominated by bremsstrahlung emission from the hot baryons. If we can link the bremsstrahlung emission to the mass of the cluster we can relate observed quantities such as X-ray luminosity and temperature to the unobserved cluster mass. We can go further and predict how luminosity should scale with temperature for a population of galaxy clusters drawn from a range of masses and compare this to observations.

We begin by considering the relationship between gas temperature and mass in a virialised halo. We use $\Delta = \frac{\rho - \bar{\rho}}{\bar{\rho}} \approx \rho/\bar{\rho}$ for $\rho \gg \bar{\rho}$. We also consider that quantities such as mass and radius are measured for regions of the cluster corresponding to a given overdensity Δ w.r.t. to the background, i.e. M_{Δ} is the cluster mass corresponding to the region forming an overdensity Δ w.r.t. the background.

The kinetic energy of the gas is

$$\begin{aligned} K &= \frac{3}{2}NkT \\ &\propto NkT \propto M_{gas,\Delta}kT \\ &\propto M_\Delta kT \end{aligned} \quad (62)$$

assuming a fixed ratio between baryons and dark matter, i.e. $M_{gas,\Delta} \propto M_\Delta$.

The potential energy is

$$P \propto \frac{GM_\Delta^2}{R_\Delta} \quad (63)$$

and the virial relation gives $P = -2K$. We can therefore write

$$M_\Delta kT \propto \frac{M_\Delta^2}{R_\Delta}. \quad (64)$$

If we express R_Δ in terms of the mean cluster density, $R_\Delta \propto M_\Delta^{1/3}\rho^{-1/3}$, to obtain

$$M_\Delta = (kT)^{3/2}\rho^{-1/2}. \quad (65)$$

The mean density of the cluster is $\rho = \Delta\,\bar{\rho} = \Delta \times \frac{3H^2}{8\pi G}$ – assuming that $\bar{\rho} = \rho_c$ (and this assumption is not always made in the literature). We can further define $H(z)/H_0 = E(z)$ – which is simply a version of the Friedmann equation – to obtain $H(z) = E(z)\,H_0$, to obtain $\rho = \Delta\rho_c = \Delta E^2(z)$ and

$$M_\Delta \propto (kT)^{3/2}\,\Delta^{-1/2}\,E^{-1}(z) \text{ or} \quad (66)$$

$$kT \propto M_\Delta^{2/3}\Delta^{1/3}E(z)^{2/3}. \quad (67)$$

Note that using the $E(z)$ formalism allows us to compare the properties of galaxy clusters observed over a range of redshifts.

Having derived a relationship between gas temperature and cluster mass we next consider the emission luminosity. At temperatures $T > 10^6$K the emission from the ionised gas is dominated by collisional bremsstrahlung with a small contribution from heavy ion line emission. The X-ray luminosity of bremsstrahlung emission may be written as

$$\begin{aligned} L_X &\propto \int n_e\,n_i\,T^{1/2}\,dV \\ &\propto \rho^2(kT)^{1/2}R_\Delta^3. \end{aligned} \quad (68)$$

Strictly speaking we should use ρ_{gas} in the above expression but I am here assuming an approximate relationship $\rho_{gas} \propto \rho$. Recall that $R_\Delta \propto M_\Delta^{1/3} \rho^{-1/3}$ to obtain

$$L_X \propto \rho (kT)^{1/2} M_\Delta. \tag{69}$$

Taking $\rho = \Delta\, \rho_c \propto \Delta E^2(z)$, we can write

$$L_X \propto \Delta\, E^2(z) (kT)^{1/2} M_\Delta. \tag{70}$$

We can now substitute in our mass-temperature relationship from Equation 67 to obtain

$$\begin{aligned} L_X &\propto \Delta\, E^2(z) (kT)^{1/2} \times (kT)^{3/2}\, \Delta^{-1/2}\, E^{-1}(z) \\ &\propto \Delta^{1/2} E^{-1}(z) (kT)^2 \\ &\propto T^2 \end{aligned} \tag{71}$$

for a sample of clusters of varying mass observed at fixed overdensity and corrected to a common redshift. Is this what we observe?