

Modern Cosmology

Chapters 1-6

Ed Copeland¹ and Costas Skordis²

School of Physics and Astronomy, University of Nottingham, Nottingham NG7 2RD, UK

Abstract

These lecture notes cover the Modern Cosmology fourth year optional module (F34MCO) and Particles and Gravity M.Sc course. It begins with a review of Friedmann models including an introduction to the thermal history of the universe, freeze-out, relics, recombination, the epoch of last scattering and dark matter candidates. We then go into more details concentrating initially on the Inflationary scenario and explaining why it is required in the early Universe. It leads us into physics beyond the standard model as we are introduced to scalar fields. A number of inflation models are presented, along with their equations of motion. The slow roll conditions are obtained and we introduce the ideas behind reheating the Universe at the end of inflation. The most important aspect of inflation, the associated generation of primordial density fluctuations are derived along with the power spectra. Particular emphasis is given to the connection between the generation of the fluctuations, the slow roll parameters and the cosmological observables. Moving on from Inflation we enter the world of large scale structure formation. Initially we adopt a Newtonian approach neglecting pressure. This allows us to define and introduce perturbation modes; matter transfer functions; nonlinear effects and the spherical collapse model. The Lagrangian approach is developed leading to a description of N-body simulations, dark-matter haloes and mass functions, as well as the importance of gas cooling. This culminates in a brief overview of galaxy formation. Gravitational lensing is introduced, we describe what it is and how it can be used to detect dark matter. The mechanisms required for generating Cosmic Microwave Background anisotropies are described and linked to the inflationary perturbations. The associated Boltzmann equations, power spectrum, tensor modes and polarisation signatures are described for the case of Λ CDM models. Finally we describe the evidence for the existence of dark energy which is believed to be driving the current acceleration of the universe. Although it fits the data the best there are theoretical issues associated with using a cosmological constant and these are described along with the results obtained from adopting one. Alternative models involving an evolving scalar field, Quintessence are introduced and the associated fine tunings required with them are described. The possibility that the current acceleration is a manifestation of modified gravity is briefly reviewed

¹ed.copeland@nottingham.ac.uk

²skordis@nottingham.ac.uk

Useful resources

- S. Dodelson, *Modern Cosmology*, (Academic Press, 2003) This will be the main book we follow. Well written covering all the main topics you will need with lots of problems set and a few solutions presented.
- D.H. Lyth and A.R. Liddle, *The Primordial Density Perturbation*, (Cambridge University Press, Cambridge, 2009) This is especially strong on the inflation section and as the title says on how to generate primordial perturbations from inflation – written by two of the experts in Inflation. We make great use of it in Chapters (5) and (6).
- P.J.E. Peebles, *Principles of Physical Cosmology*, (Princeton University Press, Princeton, 1993). A classic, written by a master of the field. Tough going though, not for the light hearted.
- J.A. Peacock *Cosmological Physics*, (Cambridge University Press, Cambridge, 1999). A superb book describing the physics of the Big Bang and which is also very up-to-date.
- V. Mukhanov *Physical Foundations of Cosmology*, (Cambridge University Press, Cambridge, 2005). A wonderful book written by a pioneer in the field of cosmological perturbations.
- E. W. Kolb and M. S. Turner, *The Early Universe*, (Frontiers in Physics, Addison-Wesley Publishing Company, 1990). If there is one vintage book in cosmology then this is it. Nicely written and all material are still relevant.
- R. Durrer, *The Cosmic Microwave Background*, (Cambridge University Press, Cambridge, 2008). Only for the brave. Very mathematical, covers in great depth everything related to CMB theory and beyond. If you can master this book you can become master of the Universe.

Contents

1	Review of Friedmann Models of Cosmology	6
1.1	Observational features of our Universe	6
1.2	Metrics – a brief resume of some key results of General Relativity	6
1.3	Light propagation and redshifts	11
1.4	The Geodesic Equation	11
1.5	Einstein Equations	12
1.6	Evolution of energy	15
1.7	Cosmological solutions	17
1.7.1	Solutions in general curved space with one component of matter	18
1.7.2	Combined matter and radiation solutions – $K=0$ case	19
1.7.3	Radiation - Λ solution – $K=0$ case	19
1.7.4	Matter - Λ solution – $K=0$ case	20
1.7.5	More general solutions – $K \neq 0$ case	20
1.8	Observational Parameters	21
1.9	Horizons and distances in cosmology	22
1.10	Age of the Universe	35
2	Thermal History of the Universe – the Hot Big Bang	35
2.1	Number densities, energy densities and pressures – relativistic and non-relativistic cases	35
2.2	The density of the universe and dark matter	42
2.3	Dark Matter Candidates	45
2.4	Dark Matter Searches	46
2.5	Freezeout and Relics	47
2.6	Baryogenesis	52
2.7	Nucleosynthesis – the origin of the light elements	52
2.8	Recombination and decoupling	55
3	Large Scale Structure formation	58
3.1	Dimensional analysis	58
3.2	Elements of Newtonian fluid dynamics	59
3.2.1	Newtonian fluids	59
3.2.2	Newtonian gravity for continuous media	60
3.2.3	The Newtonian potential Φ and the Poisson equation	62
3.2.4	The continuity and Euler equations	63
3.2.5	Evolution of small fluctuations in a Newtonian fluid	67
3.3	Math recap: Fourier transforms	69
3.3.1	Fourier transforms in 3 dimensions	69
3.3.2	The three-dimensional Dirac δ -function	69
3.3.3	Solving linear partial differential equations with Fourier transforms	70
3.4	The Jeans instability	70
3.5	Newtonian gravitational collapse in an expanding universe	71
3.5.1	Setting up the system: the background equations	72
3.5.2	Equations for fluctuations	73

3.5.3	The Jeans length in an expanding Universe	75
3.5.4	Solutions during matter domination	75
3.5.5	Solutions during radiation domination: the Mészáros effect.	76
3.6	Peculiar velocities	76
3.7	Cosmological perturbation theory	78
3.7.1	Setting up the perturbations	79
3.7.2	The scalar-vector-tensor decomposition	79
3.7.3	The general form of $\delta g_{\mu\nu}$ and $\delta T_{\mu\nu}$	81
3.7.4	Einstein and fluid equations for scalar modes	82
3.7.5	Einstein and fluid equations for tensor modes	83
3.7.6	Evolution of the potential for fluids with zero shear	84
3.7.7	Simplified equations for matter and radiation	86
3.8	Probes of Large Scale Structure	87
3.8.1	The process of structure formation	87
3.8.2	Observing large scale structure	93
4	The Cosmic Microwave Background	100
4.1	Photons in the Universe	101
4.1.1	The formation of the CMB and its spectrum	101
4.1.2	Kinetic theory and the CMB	101
4.2	CMB anisotropies	105
4.2.1	Quantifying the anisotropies	106
4.2.2	The CMB sky and expansion in angular eigenmodes	107
4.2.3	Special functions: Spherical Harmonics	108
4.2.4	Special functions: Legendre Polynomials	110
4.2.5	Relations between the spherical harmonics and the Legendre polynomials	111
4.2.6	The angular power spectrum	111
4.3	Generation of CMB anisotropies	114
4.3.1	Tight-coupling, decoupling and recombination	114
4.3.2	Recombination in detail	116
4.3.3	Decoupling in detail	119
4.3.4	The optical depth, the visibility function and the Last Scattering Surface	119
4.3.5	The Boltzmann equation for photons	120
4.3.6	The photon-baryon fluid during tight-coupling	122
4.3.7	Photons after decoupling: free-streaming	124
4.3.8	The formal solution to the Boltzmann equation: the line-of-sight integral	125
4.3.9	Special functions: Spherical Bessel functions	128
4.3.10	Relating the local temperature anisotropy to the angular power spectrum	129
4.3.11	The effective temperature and the ordinary Sachs-Wolfe effect	130
4.3.12	Acoustic oscillations	132
4.3.13	The baryon drag	133
4.3.14	Photon-diffusion and Silk damping	134
4.3.15	Acoustic driving	136
4.3.16	The Integrated Sachs-Wolfe effect	136
4.3.17	Projecting the primary anisotropies today	137
4.3.18	Putting things together: the CMB spectrum today	140

4.3.19	The baryon drag and dark matter	140
4.3.20	Further effects from secondary anisotropies	141
4.4	CMB polarization	141
4.4.1	Polarization from Compton scattering	141
4.4.2	Stokes parameters	141
4.4.3	E and B modes and polarization spectra	143
5	The Inflationary Universe	146
5.1	Problems with the Hot Big Bang	146
5.2	Enter Inflation – A.Guth 1981, A. Linde 1982	148
5.3	Inflation and scalar fields	151
5.4	Inflation dynamics	152
5.5	Number of efolds	153
5.6	Some examples of Inflation: polynomial chaotic inflation	153
5.7	Inflation models	156
6	Primordial density perturbations produced during inflation	159
6.1	Harmonic oscillators	161
6.2	Quantised free scalar field	163
6.3	Generating field perturbations as the modes exit the horizon during inflation	165
6.3.1	Massless scalar field during inflation – generation of quantum fluctuations	166
6.3.2	Quantisation of the massless fields	167
6.3.3	Going from Quantum to Classical physics	168
6.3.4	Including linear corrections from the potential – going beyond slow roll	168
6.4	Calculating the curvature perturbation ζ at horizon exit	170
7	Dark energy	172
7.1	Why Dark energy and what is it?	172
7.1.1	Hints of dark energy	173
7.1.2	Acceleration observed: the Supernovae data	176
7.1.3	Distance measures	176
7.2	The Λ CDM model, its predictions and possible shortcomings	179
7.2.1	Predictions of the Λ CDM model	179
7.2.2	Shortcomings of the Λ CDM model	179
7.3	Dynamical dark energy	181
7.3.1	Phenomenological dark energy	181
7.3.2	Quintessence	182
7.3.3	K -essence	185
7.3.4	Coupled dark energy	186
7.4	Modifications of gravity	187

"The cosmos is all that is or ever was or ever will be. Our feeblest contemplations of the Cosmos stir us – there is a tingling in the spine, a catch in the voice, a faint sensation, as if a distant memory, or falling from a height. We know we are approaching the greatest of mysteries."

Carl Sagan, *Cosmos*

1 Review of Friedmann Models of Cosmology

We begin this course with a few lectures dedicated to reviewing the subject to the level we would expect a third year undergraduate to have reached. For completeness this is a fairly detailed set of notes and the lectures will not be going into as much detail. For those not feeling so familiar with the background material, we would recommend you get to grips with this section, possibly with the aid of some of the recommended background literature.

1.1 Observational features of our Universe

Over the past few decades a number of features of our Universe appear to have been established.

- The observable patch is around 3000 Mpc ($1 \text{ Mpc} \simeq 3.26 \times 10^6 \text{ years} \simeq 3.08 \times 10^{24} \text{ cm}$).
- It is homogeneous and isotropic on scales larger than 100 Mpc with well developed inhomogeneous structure (i.e. clusters of galaxies, galaxies, stars ... us) on smaller scales.
- It is expanding according to Hubble's Law, although it is accelerating at present.
- The Universe is full of thermal microwave background radiation with temperature $T \simeq 2.725\text{K}$
- It contains baryonic matter, roughly one baryon per 10^9 photons, and very little anti-matter.
- Of the baryonic matter, 75% of it is in hydrogen, 25% helium and trace amounts of heavier elements.
- Baryons contribute 4% of the total energy density, the rest is dark and appears to be composed of cold dark matter with negligible pressure ($\sim 23\%$) and dark energy with negative pressure ($\sim 73\%$).
- Observations of the fluctuations in the cosmic microwave background radiation indicate that when the universe was a thousand times smaller than it is today, the fluctuations in the energy density distribution were as small as 10^{-5} – very small indeed.

1.2 Metrics – a brief resume of some key results of General Relativity

We need to introduce the concept of a metric tensor in order to fully exploit the cosmological solutions we will be obtaining and in particular to allow us to discuss perturbations around those background solutions. Let's start with the **metric** of space-time which describes the physical distance between points. It is the metric which allows us to interpret the geometry of the Universe, ideas of luminosities and distances in cosmology.

To get an idea of the space-time metric, first consider the usual Euclidean distance between two points on a piece of flat paper with coordinate axis x_1 and x_2 . This of course is given by Pythagoras

$$\Delta s^2 = \Delta x_1^2 + \Delta x_2^2,$$

where Δx_1 and Δx_2 are the separation in the x_1 and x_2 coordinates, and it is invariant in that it does not depend on what coordinate system we use (i.e. cartesian or polars). If the paper is replaced by a rubber sheet which expands, then the coordinate grid expands with the sheet and the physical distance between the points grows as well. If the expansion is uniform (i.e. same everywhere) we write

$$\Delta s^2 = a^2(t)[\Delta x_1^2 + \Delta x_2^2],$$

with $a(t)$ giving the rate of expansion, and coordinates x_1 and x_2 are comoving coordinates. In GR, we replace the spatial coordinates by space-time coordinates, and look for the distance between points in four dimensional space-time. In other words there is nothing special about time here it is one of the coordinates. We also allow for the possibility that the spatial sections may be curved. The infinitesimal separation, ds is written in terms of the infinitesimal coordinate separation dx^μ as

$$ds^2 = \sum_{\mu, \nu=0}^3 g_{\mu\nu} dx^\mu dx^\nu,$$

where $g_{\mu\nu}$ is the **metric**, μ and ν are Greek indices taking values 0,1,2,3, x^0 is the time coordinate and x^1 , x^2 and x^3 are the three spatial coordinates. In general $g_{\mu\nu}$ is a function of the coordinates, and this allows spacetime to be curved. From now on we will drop the explicit summation sign in these expressions, but it is implied that repeated indices are to be summed over i.e.

$$a_\mu b^\mu \equiv \sum_{\mu=0}^3 a_\mu b^\mu = a_0 b^0 + a_1 b^1 + a_2 b^2 + a_3 b^3. \quad (1.1)$$

A general vector $A^\mu = (A^0, A^i)$ has A^0 being the timelike component and A^i the three spatial components. Importantly, in relativity upper and lower indices are distinct, the former are associated with vectors and the latter with 1-forms. Going back and forth between them is done via the metric tensor,

$$A_\mu = g_{\mu\nu} A^\nu ; \quad A^\mu = g^{\mu\nu} A_\nu \quad (1.2)$$

where $g^{\mu\nu}$ is the inverse of $g_{\mu\nu}$ and will be defined shortly. A vector and a 1-form can be contracted to produce an invariant, a scalar, i.e. the four-momentum squared of a massless particle must vanish

$$P^2 \equiv P_\mu P^\mu = g_{\mu\nu} P^\mu P^\nu = 0.$$

In fact the metric raises and lowers indices on tensors in general, not just vectors. For example consider raising the indices on the metric tensor itself

$$g^{\mu\nu} = g^{\mu\alpha} g^{\nu\beta} g_{\alpha\beta} \quad (1.3)$$

Note, that if the index $\alpha = \nu$, then the first term on the right is equal to the term on the left, which means that in that case we are forced to have (from the two extra factors of g on the RHS of Eqn. (1.3))

$$g^{\nu\beta} g_{\alpha\beta} = \delta_\alpha^\nu, \quad (1.4)$$

where δ_α^ν is the Kronecker delta equal to zero unless $\nu = \alpha$ in which case it is equal to unity. This is why we say $g^{\mu\nu}$ is the inverse of $g_{\mu\nu}$.

The metric tensor $g_{\mu\nu}$ is necessarily symmetric (it should not matter whether we write $dx^\mu dx^\nu$ or $dx^\nu dx^\mu$, so in principle it has four diagonal and six off-diagonal components. As stated above it provides the link between the values of the coordinates and the more physical measure of the interval ds^2 , also known as the *proper time*. Special Relativity is described by Minkowski space-time with coordinates $x^\mu = (ct, x^i)$, $i = 1, 2, 3$ and metric $g_{\mu\nu} = \eta_{\mu\nu}$ where

$$\eta_{\mu\nu} = \begin{pmatrix} -1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \end{pmatrix} \quad (1.5)$$

For the case of an expanding universe, the two grid points move apart so that their physical separation is proportional to the scale factor. If today the comoving distance is x_0 , then the physical distance between the two points at some earlier time t was $a(t)x_0$ (which is based on the normalisation that $a(t_0) = 1$). It suggests that in a spatially flat expanding universe the metric is (recall the coordinates are $x^\mu = (ct, x^i)$, $i = 1, 2, 3$)

$$g_{\mu\nu} = \begin{pmatrix} -1 & 0 & 0 & 0 \\ 0 & a^2(t) & 0 & 0 \\ 0 & 0 & a^2(t) & 0 \\ 0 & 0 & 0 & a^2(t) \end{pmatrix} \quad (1.6)$$

or

$$ds^2 = -c^2 dt^2 + a^2(t)(dx^2 + dy^2 + dz^2) \quad (1.7)$$

Let us make sure we are happy with the meaning of this metric tensor. It is representing the relationship between time intervals and comoving intervals through

$$(\text{invariant interval})^2 = -(\text{time interval})^2 + (\text{scale factor})^2(\text{comoving interval})^2$$

There exists a universal cosmological time which we would associate with a clock ticking at constant comoving coordinates on the sky. The spatial part of the metric expands with time, given by the universal scale factor $a(t)$. This implies that particles at constant coordinates recede from the origin and therefore must undergo a Doppler redshift due to the increasing scale factor. Eqn. (1.6) is the Friedmann-Robertson-Walker (FRW) metric for a spatially flat Universe.

Because the expansion is uniform we can write

$$\mathbf{r}(t) = a(t)\mathbf{x}(t)$$

where $\mathbf{r}$ is the real (physical) distance and $\mathbf{x}$ is the comoving distance. The comoving coordinates are therefore carried along with the expansion, any objects remaining at fixed coordinate values (if they are not moving relative to each other).

Differentiating $\mathbf{r}(t) = a(t)\mathbf{x}(t)$ we find

$$\mathbf{v}(\mathbf{t}) = H(t)\mathbf{r}(\mathbf{t}) + a(t)\dot{\mathbf{x}}(t), \quad (1.8)$$

where

$$H(t) = \left(\frac{\dot{a}}{a} \right). \quad (1.9)$$

defines the Hubble parameter $H(t)$. The second term on the right hand side of Eqn. (1.8) is known as the *peculiar velocity* and it accounts for the local dynamics of the objects in question being affected (for instance the local velocities of neighbouring galaxies). The first term is the key term for cosmology, it tells how the expansion rate of the universe is directly affecting the velocity of recession between the two objects, even if they are not moving relative to one another (i.e. even if $\dot{\mathbf{x}}(t) = 0$). For that case, it gives us directly Hubble's Law $\mathbf{v}(\mathbf{t}) = H(t)\mathbf{r}(\mathbf{t})$. By convention we often state that $a(t_0) = 1$ today – recall when we refer to values today we attach a subscript '0' to the quantity. In that case the comoving distance is the actual distance today. Of course it implies $a < 1$ in the past. Setting $t = t_0$ in (1.8) we obtain Hubble's law with $H_0 = H(t_0)$. A word of caution though on simply setting $a(t_0) = 1$. It is not always possible to do that, in particular as

we shall see, there are restrictions on whether that can be done in a curved space setting.

The current value , H_0 is becoming very well constrained. Our uncertainty is parameterised by a constant h through

$$H_0 = 100h \text{ km s}^{-1} \text{ Mpc}^{-1}, \quad h = 0.744 \pm 0.025. \quad (1.10)$$

where the most recent direct measurements are reported in Riess et al, *Astrophys.J.* **730** (2011) 119, Erratum-ibid. 732 (2011) 129. The uncertainty is becoming remarkably small, now around 3%. As an example if $h = 0.72$, then if $v_{exp} = 7200 \text{ km s}^{-1}$ we have a separation of 100 Mpc. Our uncertainty in h feeds into almost all of the cosmological parameters as we will see. The units of H are inverse time, and so we can use as an estimate for the cosmological expansion time H_0^{-1} , the Hubble time: $H_0 = 100h \text{ km s}^{-1} \text{ Mpc}^{-1} \rightarrow H_0^{-1} \sim 9.77h^{-1} \times 10^9 \text{ years}$.

The evolution of the scale factor depends on the density of matter in the universe. We will shortly introduce perturbations to the metric, essential if we are to understand the generation of structure in the universe. The perturbed part of the metric will be determined by the associated inhomogeneities in the matter and radiation. Equation (1.6) is the metric for a spatially flat homogeneous and isotropic universe. In the problem set, you are asked to derive the corresponding metric for the curved space generalisation. The Cosmological Principle (CP) makes life easier for us here. it means that at any time the universe should have no preferred positions. So, the spatial part of the metric must have a constant curvature (same everywhere!) which could of course be zero as in the flat metric. The most general form of the spatial metric (ds_3^2) of a three-dimensional space with constant curvature is (written in spherical polars)):

$$ds_3^2 = a^2 \left(\frac{dr^2}{1 - Kr^2} + r^2(d\theta^2 + \sin^2 \theta d\phi^2) \right), \quad (1.11)$$

where $a^2 > 0$ and the constant K measures the curvature of space. It is the same K we used in the Friedmann equation, and describes spherical ($K > 0$), flat ($K = 0$) and hyperbolic ($K < 0$) geometries respectively. It is often normalised to be $K = +1, 0, -1$ for the three geometries. Given the form of the spatial metric which satisfies the cosmological principle, we write down the full space-time metric (i.e. include the infinitesimal change in time dt). As with moving from the paper sheet to the rubber band which expands, we can allow the space to grow or shrink in time. This gives the general curved space Robertson-Walker metric

$$ds^2 = -c^2 dt^2 + a^2(t) \left[\frac{dr^2}{1 - Kr^2} + r^2(d\theta^2 + \sin^2 \theta d\phi^2) \right]. \quad (1.12)$$

Consider the spatial sections for a few minutes. Returning to Eqn. (1.11), it often proves useful to replace the radial coordinate r with χ which is defined by

$$d\chi^2 = \frac{dr^2}{1 - Kr^2} \quad (1.13)$$

By integrating this, it follows that

$$\chi = \text{arcsinh } r, \quad K = -1 \quad (1.14)$$

$$\chi = r, \quad K = 0 \quad (1.15)$$

$$\chi = \arcsin r, \quad K = +1 \quad (1.16)$$

The coordinate χ varies between $0 \leq \chi \leq \infty$ for flat and hyperbolic spaces, and $0 \leq \chi \leq \pi$ for positively curved spaces. The metric Eqn. (1.11) now becomes in terms of χ ,

$$ds_3^2 = a^2(d\chi^2 + S_K^2(\chi)d\Omega^2)$$

with

$$S_K(\chi) = \sinh \chi, \quad K = -1 \quad (1.17)$$

$$S_K(\chi) = \chi, \quad K = 0 \quad (1.18)$$

$$S_K(\chi) = \sin \chi, \quad K = +1, \quad (1.19)$$

where

$$d\Omega^2 = (d\theta^2 + \sin^2 \theta d\phi^2)$$

It is worth looking a bit closer at the case of these constant curvature spaces.

Three-dimensional sphere ($K=+1$) From ds_3^2 , the distance element on the surface of the 2-sphere of radius χ is

$$dl^2 = a^2 \sin^2 \chi (d\theta^2 + \sin^2 \theta d\phi^2).$$

You should be able to see that this is the same line element as a sphere of radius $R = a \sin \chi$ in flat three-dimensional space, which means that we can straight away write the total surface area :

$$S_{2d}(\chi) = 4\pi R^2 = 4\pi a^2 \sin^2 \chi.$$

The behaviour is at first strange. As the radius χ increases, the surface area grows to a maximum value at $\chi = \pi/2$, then decreases, vanishing at $\chi = \pi$. A lower dimensional analogy may be useful. The surface of the globe plays the role of three-dimensional space with constant curvature, and the two dimensional surfaces correspond to circles of constant latitude on the globe. Starting from the north pole ($\theta = 0$), the circle circumference grows as we go south, reaching a max at the equator ($\theta = \pi/2$), then decreases as we go further, disappearing at the south pole ($\theta = \pi$). The circles cover the whole surface of the globe as θ runs from 0 to π . The same happens here with χ running from 0 to π , it covers the whole three-dimensional space of constant curvature. The area of the globe is finite, implying the volume of the three-dimensional space should also be with constant positive curvature. To show this recall that the physical width of an infinitesimal shell is $dl = a d\chi$, hence the volume element between two spheres with radii χ and $\chi + d\chi$ is

$$dV = S_{2d}(\chi) a d\chi = 4\pi a^3 \sin^2 \chi d\chi.$$

The volume within the sphere of radius χ_0 is

$$V(\chi_0) = 4\pi a^3 \int_0^{\chi_0} \sin^2 \chi d\chi = 2\pi a^3 (\chi_0 - \frac{1}{2} \sin 2\chi_0).$$

A strange looking volume, but note that for the case where the radius is small $\chi_0 \ll 1$, we recover the familiar Euclidean result

$$V(\chi_0) = \frac{4}{3}\pi(a\chi_0)^3 + \dots$$

The total volume is for $\chi_0 = \pi$ which gives the finite result

$$V = 2\pi^2 a^3.$$

Three-dimensional pseudo-sphere ($K=-1$)- constant negative curvature. The metric on the surface of the corresponding 2-dimensional sphere of radius χ is

$$dl^2 = a^2 \sinh^2 \chi (d\theta^2 + \sin^2 \theta d\phi^2),$$

which following the argument for the positively curved space above gives the area of the sphere as

$$S_{2d}(\chi) = 4\pi R^2 = 4\pi a^2 \sinh^2 \chi$$

which increases exponentially for $\chi \gg 1$. Recall that $0 \leq \chi \leq \infty$, it follows that the total volume of the hyperbolic space is infinite, being given by $V = \int_0^\infty S_{2d} a d\chi$.

1.3 Light propagation and redshifts

We know from special relativity that light follows trajectories with zero proper time (null geodesics). Considering Eqn. (1.12) it follows that the radial equations of motion integrate to give

$$\int \frac{dr}{\sqrt{1 - Kr^2}} = c \int \frac{dt}{a(t)} \quad (1.20)$$

The comoving distance is a constant, whereas the domain of integration in time extends from t_{emit} to t_{obs} , the times of emission and detection of a photon. It therefore follows that

$$\frac{dt_{\text{emit}}}{dt_{\text{obs}}} = \frac{a_{\text{emit}}}{a_{\text{obs}}} \quad (1.21)$$

implying that events on distant galaxies time-dilate. Now this dilation also applies to frequency so

$$\frac{\nu_{\text{emit}}}{\nu_{\text{obs}}} \equiv 1 + z = \frac{a_{\text{obs}}}{a_{\text{emit}}} \quad (1.22)$$

In other words by observing shifts in spectral lines, we can determine the size of the universe at the time the light was emitted – this is the key result which enables the discipline of observational cosmology.

1.4 The Geodesic Equation

What directions do particles go in in curved space? We know that in Minkowski space they travel in straight lines unless acted on by an external force. In curved space, the straight line is generalised to a *geodesic*, the path of a particle in the absence of external forces. To obtain it we generalise Newton's Law in the absence of forces, $d^2\mathbf{x}/dt^2 = 0$, to the case of an expanding universe. This is a standard GR calculation and is done in all decent textbooks. For the case of relativity two key modifications need to be made. The first is to allow the indices to run from 0 to 3 thereby allowing time to be one of the coordinates. The second emerges because of the fact we now have time as a coordinate. It implies we can not use it as our evolution parameter. Instead we introduce a parameter λ which monotonically increases along the particles path. **[Insert figure here]**. The geodesic equation then becomes (see for example Schutz for a proof)

$$\frac{d^2 x^\mu}{d\lambda^2} + \Gamma_{\alpha\beta}^\mu \frac{dx^\alpha}{d\lambda} \frac{dx^\beta}{d\lambda} = 0 \quad (1.23)$$

where the Christoffel symbol is given by

$$\Gamma_{\alpha\beta}^{\mu} = \frac{g^{\mu\nu}}{2} \left[\frac{\partial g_{\alpha\nu}}{\partial x^{\beta}} + \frac{\partial g_{\beta\nu}}{\partial x^{\alpha}} - \frac{\partial g_{\alpha\beta}}{\partial x^{\nu}} \right], \quad (1.24)$$

where to remind you $x^{\mu} = (ct, x^i)$, $i = 1, 2, 3$. Note the use of the inverse metric $g^{\mu\nu}$ defined in Eqn. (1.4). From Eqn. (1.6) we see that in the flat ($K = 0$), FRW metric the inverse is identical to $g_{\mu\nu}$ except that its spatial elements are $1/a^2$ instead of a^2 . Using Eqns. (1.6) and (1.24) we can now derive the Christoffel symbols in a spatially flat expanding homogeneous universe. First evaluate the components with the upper index being zero, $\Gamma_{\alpha\beta}^0$. The fact that the metric is diagonal implies that $g^{\nu 0} = 0$ unless $\nu = 0$ in which case $g^{00} = -1$. We then have

$$\Gamma_{\alpha\beta}^0 = \frac{-1}{2} \left[\frac{\partial g_{\alpha 0}}{\partial x^{\beta}} + \frac{\partial g_{\beta 0}}{\partial x^{\alpha}} - \frac{\partial g_{\alpha\beta}}{\partial x^0} \right]. \quad (1.25)$$

The first two terms are just derivatives of g_{00} so vanish because $g_{00} = -1$. We are left with

$$\Gamma_{\alpha\beta}^0 = \frac{1}{2} \frac{\partial g_{\alpha\beta}}{\partial x^0}. \quad (1.26)$$

For this to be non-zero, we require α and β to be spatial indices, which we identify with the Roman letters i, j running from 1 to 3. Now since $x^0 = ct$ we have

$$\begin{aligned} \Gamma_{00}^0 &= 0 \\ \Gamma_{0i}^0 &= \Gamma_{i0}^0 = 0 \\ \Gamma_{ij}^0 &= \delta_{ij} a' a \end{aligned} \quad (1.27)$$

where $a' \equiv \frac{da}{d(ct)} = \frac{1}{c} \frac{da}{dt} = \frac{1}{c} \dot{a}$.¹ It is also straightforward to show that $\Gamma_{\alpha\beta}^i$ vanishes unless one of the lower indices is zero and one is spatial giving

$$\Gamma_{0j}^i = \Gamma_{j0}^i = \delta_{ij} \frac{a'}{a} \quad (1.28)$$

1.5 Einstein Equations

Einstein's equations relate the components of the Einstein tensor describing the geometry of space-time to the energy-momentum tensor describing the energy. The equation (including a cosmological constant Λ) is given by

$$G_{\mu\nu} \equiv R_{\mu\nu} - \frac{1}{2} g_{\mu\nu} R = \frac{8\pi G}{c^4} T_{\mu\nu} - \Lambda g_{\mu\nu} \quad (1.29)$$

where $G_{\mu\nu}$ is the Einstein tensor, $R_{\mu\nu}$ is the Ricci tensor (which depends on the metric and its derivatives), R is the Ricci scalar and is given by the contraction of the Ricci tensor ($R = g^{\mu\nu} R_{\mu\nu}$). G is Newton's constant, and $T_{\mu\nu}$ is the energy-momentum tensor. It is a symmetric tensor which describes the constituents of matter in the Universe. Note that whereas the left hand side of Eqn. (1.29) is a function of the metric, the right hand side is a function of the energy – Einstein's equation relates geometry and matter !

¹In section 1.7 we will be considering the case where primes will denote derivatives with respect to conformal time. We will remind you in that case what the new a' means.

The Ricci tensor is given by

$$R_{\mu\nu} = \Gamma_{\mu\nu,\alpha}^\alpha - \Gamma_{\mu\alpha,\nu}^\alpha + \Gamma_{\beta\alpha}^\alpha \Gamma_{\mu\nu}^\beta - \Gamma_{\beta\nu}^\alpha \Gamma_{\mu\alpha}^\beta, \quad (1.30)$$

where $\Gamma_{\mu\nu,\alpha}^\alpha \equiv \frac{\partial \Gamma_{\mu\nu}^\alpha}{\partial x^\alpha}$ etc... For the case of the FRW universe we have already obtained the Christoffel symbols Eqns. (1.27) and (1.28) with all the others being zero. Using this we see that there are only two sets of nonvanishing components of the Ricci tensor; one with $\mu = \nu = 0$ and the other with $\mu = \nu = i$. For the case of R_{00} we have

$$R_{00} = \Gamma_{00,\alpha}^\alpha - \Gamma_{0\alpha,0}^\alpha + \Gamma_{\beta\alpha}^\alpha \Gamma_{00}^\beta - \Gamma_{\beta 0}^\alpha \Gamma_{0\alpha}^\beta, \quad (1.31)$$

which then simplifies to

$$R_{00} = -\Gamma_{0i,0}^i - \Gamma_{j0}^j \Gamma_{0i}^i, \quad (1.32)$$

and finally using Eqn. (1.28) we end up with

$$\begin{aligned} R_{00} &= -\delta_{ii} \left(\frac{a'}{a} \right)' - \left(\frac{a'}{a} \right)^2 \delta_{ij} \delta_{ij} \\ &= -3 \left[\frac{a''}{a} - \left(\frac{a'}{a} \right)^2 \right] - 3 \left(\frac{a'}{a} \right)^2 \\ &= -3 \frac{a''}{a} \end{aligned} \quad (1.33)$$

The factors of three on the second line are from the Kronecker δ functions – recall δ_{ii} means summing over all three spatial indices counting one for each. The space-space component is given by

$$R_{ij} = \Gamma_{ij,\alpha}^\alpha - \Gamma_{i\alpha,j}^\alpha + \Gamma_{\beta\alpha}^\alpha \Gamma_{ij}^\beta - \Gamma_{\beta i}^\alpha \Gamma_{j\alpha}^\beta, \quad (1.34)$$

which for our FRW Universe becomes

$$R_{ij} = \Gamma_{ij,0}^0 + \Gamma_{0k}^k \Gamma_{ij}^0 - \Gamma_{ki}^0 \Gamma_{j0}^k - \Gamma_{0i}^k \Gamma_{jk}^0, \quad (1.35)$$

which in turn becomes

$$R_{ij} = \delta_{ij} [2a'^2 + aa'']. \quad (1.36)$$

The Ricci scalar follows,

$$\begin{aligned} R &= g^{\mu\nu} R_{\mu\nu} \\ &= -R_{00} + \frac{1}{a^2} R_{ii} \\ &= 6 \left[\frac{a''}{a} + \left(\frac{a'}{a} \right)^2 \right], \end{aligned} \quad (1.37)$$

where again remember the sum over i leads to a factor of three in R_{ii} . The Friedmann equation comes from the considering only the time-time coordinate of the Einstein equations:

$$R_{00} - \frac{1}{2} g_{00} R = \frac{8\pi G}{c^4} T_{00} - \Lambda g_{00} \quad (1.38)$$

For the case of a perfect isotropic fluid, the energy momentum tensor is given by

$$T_{\nu}^{\mu} = \begin{pmatrix} -\rho c^2 & 0 & 0 & 0 \\ 0 & p & 0 & 0 \\ 0 & 0 & p & 0 \\ 0 & 0 & 0 & p \end{pmatrix} \quad (1.39)$$

where $\rho(t)$ is the energy density and $p(t)$ is the pressure of the fluid. Hence using Eqn. (1.39) in Eqn. (1.38) with Eqns. (1.33) and (1.37) we obtain Einstein's equation in a spatially flat ($K = 0$) FRW universe²

$$\left(\frac{\dot{a}}{a}\right)^2 = \frac{8\pi G}{3}\rho(t) + \frac{\Lambda c^2}{3} \quad (1.40)$$

This is the **Friedmann equation** in the $K = 0$ universe. The general curvature case with a cosmological constant follows from considering the metric Eqn. (1.12) in Eqn. (1.29)

$$\left(\frac{\dot{a}}{a}\right)^2 = \frac{8\pi G}{3}\rho(t) - \frac{Kc^2}{a^2} + \frac{\Lambda c^2}{3}, \quad (1.41)$$

A second equation follows when we consider the space-space component of Einstein's equation

$$R_{ij} - \frac{1}{2}g_{ij}R = \frac{8\pi G}{c^4}T_{ij} \quad (1.42)$$

Of course the only non-trivial components are when $i = j$ and we obtain

$$2\dot{a}^2 + a\ddot{a} - \frac{1}{2}a^2 \left[\frac{\ddot{a}}{a} + \left(\frac{\dot{a}}{a}\right)^2 \right] = \frac{8\pi G}{c^2}a^2 p(t) - a^2 \Lambda c^2 \quad (1.43)$$

which simplifies to give

$$2\frac{\ddot{a}}{a} + \left(\frac{\dot{a}}{a}\right)^2 = -\frac{8\pi G}{c^2}p(t) + \Lambda c^2 \quad (1.44)$$

The general curvature case follows as before

$$2\frac{\ddot{a}}{a} + \left(\frac{\dot{a}}{a}\right)^2 = -\frac{8\pi G}{c^2}p(t) - \frac{Kc^2}{a^2} + \Lambda c^2 \quad (1.45)$$

Equations (1.41) and (1.45) are the basis of the standard big bang cosmological model including the current Λ CDM model. Note that they can be combined to give an **acceleration** equation where the curvature term drops out,

$$\frac{\ddot{a}}{a} = -\frac{4\pi G}{3} \left(\rho(t) + \frac{3p(t)}{c^2} \right) + \frac{\Lambda c^2}{3} \quad (1.46)$$

Note the cosmological constant term can be omitted if we make the following replacement in Eqn. (1.46)

$$\begin{aligned} \rho &\rightarrow \rho + \frac{\Lambda c^2}{8\pi G} \\ p &\rightarrow p - \frac{\Lambda c^2}{8\pi G} \end{aligned}$$

²Note we have gone from a' to $\dot{a}$ by multiplying through by c^2 . Recall $a' = \frac{\dot{a}}{c}$

Therefore the cosmological constant can be interpreted as arising from a form of energy which has negative pressure, equal in magnitude to its (positive) energy density:

$$p = -\rho c^2 \quad (1.47)$$

which of course is consistent with $\dot{\rho} = 0$ in Eqn. (1.56) below. Such a form of energy is a generalization of the notion of a cosmological constant and is known as dark energy. In fact, in order to get a term which causes an acceleration of the universe expansion, it is enough to have a source of unusual matter which satisfies

$$p(t) < -\frac{\rho(t)c^2}{3}. \quad (1.48)$$

Such a source is sometimes known as quintessence, and is usually associated with a fluid comprised of a scalar field. It is of course unusual to have a fluid with a negative pressure, yet this is required if we are to have a universe accelerating.

Returning to the Friedmann equation (1.41) where we have re-absorbed the explicit cosmological constant into the general energy density ρ we have

$$H^2 = \frac{8\pi G}{3}\rho(t) - \frac{Kc^2}{a^2}. \quad (1.49)$$

It allows us to define the **critical density** which is the density of matter required to yield a flat universe

$$\rho_c \equiv \frac{3H^2}{8\pi G}. \quad (1.50)$$

Another quantity that can be defined is the dimensionless **density parameter** as the ratio of the density to the critical density:

$$\Omega \equiv \frac{\rho(t)}{\rho_c} = \frac{8\pi G\rho}{3H^2} \quad (1.51)$$

Today's values of these parameters are usually given a zero subscript, i.e. H_0, ρ_0, Ω_0 . Recall Eqn. (1.10) for the Hubble parameter today, then the current density of the universe is

$$\begin{aligned} \rho_0 &= 1.878 \times 10^{-26} \Omega_0 h^2 \text{ kg m}^{-3} \\ &= 2.775 \times 10^{11} \Omega_0 h^2 M_\odot \text{Mpc}^{-3} \end{aligned} \quad (1.52)$$

Of course the critical density ρ_c corresponds to the case $\Omega_0 = 1$ or a spatially flat universe.

1.6 Evolution of energy

Let's end this section by deriving the fluid equation. The key starting point is the idea of energy momentum conservation which in the case of an expanding universe implies the *covariant* derivative of the energy momentum tensor vanishes. We are not going to derive the result here, but simply state it, although it is not difficult to derive:

$$T_{\nu;\mu}^\mu \equiv T_{\nu,\mu}^\mu + \Gamma_{\alpha\mu}^\mu T_\nu^\alpha - \Gamma_{\nu\mu}^\alpha T_\alpha^\mu = 0 \quad (1.53)$$

Eqn. (1.53) actually corresponds to four separate equations because ν can take on four values. Considering the case $\nu = 0$ then we have

$$T_{0,\mu}^\mu + \Gamma_{\alpha\mu}^\mu T_0^\alpha - \Gamma_{0\mu}^\alpha T_\alpha^\mu = 0 \quad (1.54)$$

However, from Eqn. (1.39) we see that assuming isotropy implies T_i^0 vanishes, hence the dummy indices μ in the first term and α in the second must be equal to zero leaving us with

$$-\frac{\partial(\rho(t)c^2)}{\partial(ct)} - \Gamma_{0\mu}^\mu \rho(t)c^2 - \Gamma_{0\mu}^\alpha T_\alpha^\mu = 0 \quad (1.55)$$

Now we can simplify further, since we know that $\Gamma_{0\mu}^\alpha$ vanishes unless μ and α are spatial indices and are equal to each other, as seen in Eqn. (1.28). This then leaves us with the result for the well known fluid equation in an expanding universe

$$\frac{\partial\rho}{\partial t} + 3\frac{\dot{a}}{a}\left(\rho + \frac{p}{c^2}\right) = 0 \quad (1.56)$$

Now Eqn. (1.56) is not the end of the story, we need to relate the pressure (p) and energy density (ρ) of the fluid. This is done by assuming there is a unique **Equation of state** of the form $p = p(\rho)$, for each fluid. For the types of fluid we will be considering (no torsion) it is thought to be a simple linear relation which is written generically in one of two equivalent ways:

$$p = w\rho \quad \text{or} \quad p = (\gamma - 1)\rho$$

where of course $w = \gamma - 1$, and both w and γ being known as the equation of state parameter. There are some particularly important cases which crop up in cosmology: Matter – $w = 0$, includes non-relativistic particles such as baryons as well as cold dark matter and is sometimes called dust. It is pressureless and satisfies $p = 0$, which is a good approximation for atoms which seldom interact in a cooled universe. Galaxies also obey $p = 0$, as they mainly interact gravitationally. Radiation – $w = 1/3$, describes any massless (and very light) particles which move with speed approaching c . From your electromagnetic wave courses you will recall that light exerts a radiation pressure with equation of state $p = \frac{\rho c^2}{3}$. Cosmological Constant – $w = -1$ is the energy density associated with quantum fluctuations. The corresponding equation of state satisfies $p = -\rho c^2$, in other words the pressure is negative! It is vital for the *Inflationary Universe Scenario* as we shall see later.

We can solve for more general equations of state of the form $p = w\rho c^2$. The fluid equation becomes

$$\dot{\rho} + 3\frac{\dot{a}}{a}\rho(1 + w) = 0 \longrightarrow \frac{\dot{\rho}}{\rho} + 3\frac{\dot{a}}{a}(1 + w) = 0 \longrightarrow \rho \propto a^{-3(1+w)} \quad (1.57)$$

The Friedmann equation then becomes (for $k = 0$)

$$\left(\frac{\dot{a}}{a}\right)^2 = \frac{8\pi G}{3}a^{-3(1+w)} \longrightarrow \dot{a} \propto a^{-(1+3w)/2} \longrightarrow a^{3(1+w)/2} \propto t \longrightarrow a(t) \propto t^{\frac{2}{3(1+w)}}. \quad (1.58)$$

You can check it with the known cases of matter and radiation. For example for **matter** $w = 0$ we obtain the $\Omega_m = 1$ Einstein de Sitter solution

$$\rho_m(t) = \rho_{m0} \left(\frac{a_0}{a}\right)^3 \quad \text{and} \quad a(t) = a_0 \left(\frac{t}{t_0}\right)^{2/3} \quad (1.59)$$

whereas for **radiation** $w = 1/3$ we obtain the radiation dominated or Tolman universe:

$$\rho_r(t) = \rho_{r0} \left(\frac{a_0}{a}\right)^4 \quad \text{and} \quad a(t) = a_0 \left(\frac{t}{t_0}\right)^{1/2} \quad (1.60)$$

The extra factor of a in the relative energy densities between matter and radiation is just a reflection of the fact that the number density of particles is diluted by the expansion, with photons also having their energy reduced by the redshift. For the case of a **cosmological constant** $w = -1$ we have

$$\rho_v(t) = \rho_{v0} \text{ and } a(t) = a_0 \exp(H(t - t_0)) \quad (1.61)$$

where $H^2 = \frac{8\pi G}{3}\rho_{v0}$ is a constant.

Recall that the total energy density is made up of contributions from matter ρ_m , radiation ρ_r , and something that resembles a cosmological constant type term today say ρ_v , with equation of state parameter w which we would usually set to $w = -1$. Then recall the scale factor-redshift relation given in Eqn. (1.22), (i.e. $1 + z = \frac{a_0}{a}$) and the critical density as defined in Eqn. (1.50), we can write

$$\begin{aligned} \frac{8\pi G}{3}\rho &= \frac{8\pi G}{3}(\rho_m + \rho_r + \rho_v) \\ &= \frac{8\pi G}{3H_0^2}H_0^2 \left(\rho_{m0}(1+z)^3 + \rho_{r0}(1+z)^4 + \rho_{v0}(1+z)^{3(1+w)} \right) \\ &= H_0^2 \left(\Omega_{m0}(1+z)^3 + \Omega_{r0}(1+z)^4 + \Omega_{v0}(1+z)^{3(1+w)} \right) \end{aligned} \quad (1.62)$$

Returning to the Friedmann equation (1.41) we see that it has the awkward K factor in it. It is not an observable as such and so we really need to eliminate it if we are to make progress observationally. We do it by realising that it is a constant and that it can be written in terms of the observed parameters today. In particular from (1.41) applied today we have

$$\frac{Kc^2}{a_0^2} = H_0^2 (\Omega_0 - 1) \quad (1.63)$$

where $\Omega_0 = \Omega_{m0} + \Omega_{r0} + \Omega_{v0}$. It then follows upon substitution of Eqn. (1.63) into Eqn. (1.41) that

$$H^2(z) = H_0^2 \left(\Omega_{m0}(1+z)^3 + \Omega_{r0}(1+z)^4 + \Omega_{v0}(1+z)^{3(1+w)} - (\Omega_0 - 1)(1+z)^2 \right). \quad (1.64)$$

This equation is in a very useful form as it can be integrated immediately to get $t(z)$. The integrals are straightforward, at least numerically, and for the case of a flat universe ($\Omega_0 = 1$) they can be performed analytically allowing us to obtain $a(t)$. As can be seen from Eqn. (1.64), curvature can always be neglected at sufficiently early times ($z \gg 1$), as can vacuum density (except for when we consider the theory of inflation as it postulates that the vacuum density was very much higher in the very distant past).

1.7 Cosmological solutions

We can now begin to address the question of what the cosmological solutions to the Friedmann equation for different cases. We have already obtained the case of single fluid components in a flat universe Eqns. (1.59)-(1.61). Thats fine when one fluid completely dominates the dynamics but that is not generally the case, for instance we know the universe actually contains both matter and radiation, and so we need to solve for them together. Moreover, in the very early universe we may expect it to be dominated initially by radiation before moving onto a period of Λ domination corresponding to the onset of inflation. Here we deal with these more general cases which requires

the introduction of **conformal time** as a useful aid in obtaining these more complicated solutions. Conformal time η , is defined by $cdt = a(\eta)d\eta$, or

$$\eta \equiv \int \frac{cdt}{a(t)}. \quad (1.65)$$

In the following section (1.7.1) we set the speed of light to unity for convenience (i.e. we set $c = 1$).

1.7.1 Solutions in general curved space with one component of matter

The Friedmann equation and acceleration equation become in terms of the new time variable:

$$\begin{aligned} a'^2 &= \frac{8\pi G}{3}\rho a^4 - K a^2 \\ a'' &= \frac{4\pi G}{3}(\rho - 3p)a^3 - K a \end{aligned}$$

where $a' \equiv \frac{da}{d\eta}$ etc.... (Note the new meaning of a' as opposed to that in section (1.4) – it will keep this meaning from now on.) For the case of radiation $p = \rho/3$, the acceleration equation now becomes

$$a'' + K a = 0,$$

which can be integrated and leads to

$$\begin{aligned} a(\eta) &= c_r \sinh \eta, & K = -1, & 0 \leq \eta \leq \infty \\ a(\eta) &= c_r \eta, & K = 0, & 0 \leq \eta \leq \infty \\ a(\eta) &= c_r \sin \eta, & K = +1, & 0 \leq \eta \leq \pi \end{aligned}$$

where c_r is a constant of integration and the second one has been fixed by demanding $a(\eta = 0) = 0$. The physical time then follows from,

$$t = \int a(\eta) d\eta,$$

giving

$$\begin{aligned} t &= c_r(\cosh \eta - 1), & K = -1 \\ t &= c_r \eta^2/2, & K = 0 \\ t &= c_r(1 - \cos \eta), & K = +1 \end{aligned}$$

These solutions are parametric in that although we have $a(\eta)$ and $t(\eta)$ it is not generally possible to obtain an analytic expression for $a(t)$ apart of course for the $K = 0$ case. Indeed in the $K = 0$, radiation dominated universe it immediately follows that $a(t) \propto t^{1/2}$ with $H = 1/2t$.

For the case of dust domination, $p_m = 0$ and the acceleration equation is solved to give

$$\begin{aligned} a(\eta) &= c_m(\cosh \eta - 1), & K = -1, & 0 \leq \eta \leq \infty \\ a(\eta) &= c_m \eta^2, & K = 0, & 0 \leq \eta \leq \infty \\ a(\eta) &= c_m(1 - \sin \eta), & K = +1, & 0 \leq \eta \leq 2\pi \end{aligned}$$

where c_m is a constant of integration – why not have a go to show it?

1.7.2 Combined matter and radiation solutions – K=0 case

Consider just a mixture of matter and radiation in a spatially flat universe. The energy density of matter scales as a^{-3} and radiation as a^{-4} , allowing us to write the combination as

$$\rho = \rho_m + \rho_r = \frac{\rho_{eq}}{2} \left(\left(\frac{a_{eq}}{a} \right)^3 + \left(\frac{a_{eq}}{a} \right)^4 \right),$$

where a_{eq} is the scale factor when the two components have equal energy densities – an epoch known as matter-radiation equality. Note that in the acceleration equation, the contribution from radiation vanishes because $\rho_r - 3p_r = 0$, leaving us with the pressureless matter contribution,

$$a'' = \frac{2\pi G}{3} \rho_{eq} a_{eq}^3.$$

The RHS is constant, so the integral is trivial giving,

$$a(\eta) = \frac{\pi G}{3} \rho_{eq} a_{eq}^3 \eta^2 + C\eta,$$

and one of the constants of integration has been fixed by demanding $a(\eta = 0) = 0$. We fix C by inserting the solution for $a(\eta)$ into the Friedmann equation

$$a'^2 = \frac{8\pi G}{3} \rho a^4$$

with ρ given above. This gives

$$C = (4\pi G \rho_{eq} a_{eq}^4 / 3)^{1/2}$$

hence

$$a(\eta) = a_{eq} \left(\left(\frac{\eta}{\eta_*} \right)^2 + 2 \left(\frac{\eta}{\eta_*} \right) \right)$$

with

$$\eta_* = (\pi G \rho_{eq} a_{eq}^2 / 3)^{-1/2} = \eta_{eq} / (\sqrt{2} - 1)$$

Note that we obtain the expected results in the appropriate limit. For $\eta \ll \eta_{eq}$, radiation dominates and we have $a \propto \eta$, whereas for $\eta \gg \eta_{eq}$, matter has come to dominate and we have $a \propto \eta^2$. To see that this gives the usual proper time dependence, simply insert the scale factor into $t = \int a d\eta$.

1.7.3 Radiation - Λ solution – K=0 case

We can solve for a system with just radiation and a cosmological constant ($w = -1$) in a straightforward manner. From Eqn. (1.64) we have

$$H^2(z) = H_0^2 \left(\Omega_{r0} \left(\frac{a_0}{a} \right)^4 + \Omega_{v0} \right). \quad (1.66)$$

This simplifies to

$$\frac{d}{dt} a^2 = 2H_0 \sqrt{\Omega_{r0}} a_0^2 \left(1 + \alpha^2 \left(\frac{a}{a_0} \right)^4 \right)^{\frac{1}{2}}, \quad (1.67)$$

where $\alpha^2 \equiv \frac{\Omega_{v0}}{\Omega_{r0}}$. Defining $y = \frac{a^2}{a_0^2}$ we then have

$$\dot{y} = 2H_0\sqrt{\Omega_{r0}}(1 + \alpha^2 y^2)^{\frac{1}{2}} \quad (1.68)$$

which can be integrated to give

$$a(t) = a_0 \left(\frac{\Omega_{r0}}{\Omega_{v0}} \right)^{\frac{1}{4}} \left(\sinh(2H_0\Omega_{v0}^{\frac{1}{2}}t) \right)^{\frac{1}{2}} \quad (1.69)$$

where the initial condition $a(0) = 0$ has been used. For early times or small t we obtain $a(t) \propto t^{\frac{1}{2}}$ corresponding to radiation domination, and for large t we obtain exponential expansion as found in vacuum dominated de-Sitter type evolution $a(t) \propto \exp(H_0\Omega_{v0}^{\frac{1}{2}}t)$. This would be an appropriate model for the onset of a phase of inflation following a big-bang singularity.

1.7.4 Matter - Λ solution – $K=0$ case

We follow the procedure as in the radiation case starting with

$$H^2(z) = H_0^2 \left(\Omega_{m0} \left(\frac{a_0}{a} \right)^3 + \Omega_{v0} \right). \quad (1.70)$$

This time by introducing $y = \left(\frac{a}{a_0} \right)^{\frac{3}{2}}$ we obtain

$$a(t) = a_0 \left(\frac{\Omega_{m0}}{\Omega_{v0}} \right)^{\frac{1}{3}} \left(\sinh \left(\frac{3}{2} H_0 \Omega_{v0}^{\frac{1}{2}} t \right) \right)^{\frac{2}{3}}. \quad (1.71)$$

Once again for small t we obtain the matter dominated regime $a(t) \propto t^{\frac{2}{3}}$ which then evolves into a vacuum dominated exponential expansion at late times, $a(t) \propto \exp(H_0\Omega_{v0}^{\frac{1}{2}}t)$. It could well be this type of transition which has described our recent universe.

1.7.5 More general solutions – $K \neq 0$ case

It is possible to derive some more general solutions for non-flat spacetimes and this is set as an exercise in one of the problem sets. Of course observations of the large scale features of our Universe indicate that the curvature contribution to the energy budget is likely to be no more than around 1% today, but it does not rule out the possibility that we live in an open or closed universe. From a formal point of view, knowing this difference is vital, for example a commonly recited prediction of the Landscape of string theory is that the universe is likely to be open. We have from Eqn. (1.63) that

$$\frac{Kc^2}{a_0^2} = H_0^2 (\Omega_0 - 1)$$

where $\Omega_0 = \Omega_{m0} + \Omega_{r0} + \Omega_{v0}$.

If we ignore radiation then it follows that the behaviour of $\Omega_{m0} + \Omega_{v0}$ will determine the nature of the curvature. In particular in a plot of $\Omega_{m0} - \Omega_{v0}$ the diagonal line $\Omega_{m0} + \Omega_{v0} = 1$ is the crucial one, separating open and closed models. If $\Omega_{v0} < 0$, the solution will always recollapse, whereas having $\Omega_{v0} > 0$ does not guarantee expansion to infinity, especially if the matter density is high. Also, if Ω_{v0} is large enough, then for closed models there was no big bang in the past, the universe must have emerged from a bounce at some finite minimum radius. These features can be seen in Figure. (1.1).

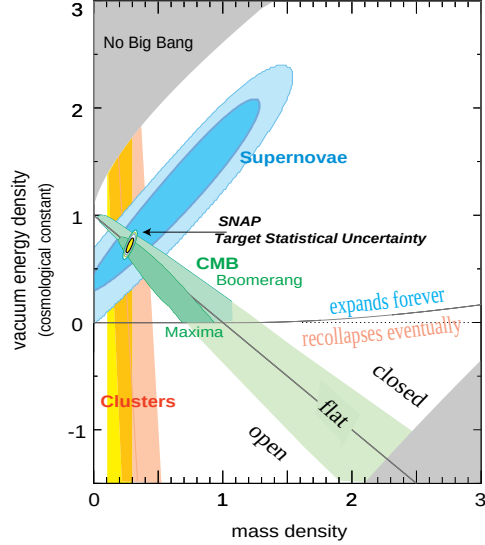


Figure 1.1: The $\Omega_0^{(0)}$ - $\Omega_\Lambda^{(0)}$ confidence regions constrained from the observations of SN Ia, CMB and galaxy clustering. We also show the expected confidence region from a SNAP (Supernova project) satellite for a flat universe with $\Omega_0^{(0)} = 0.28$.

1.8 Observational Parameters

The big bang does not fix many of the parameters we will encounter in these lectures, such as $H_0, \rho_{mat0}, \rho_{rad0}, \dots$. We can determine a few of them from observations.

Expansion rate H_0 – Hubble's constant

$$H_0 = 100h \text{ km s}^{-1} \text{ Mpc}^{-1}, \quad h = 0.744 \pm 0.025.$$

The deceleration parameter : q_0

As the name suggests q_0 provides information about the acceleration of the universe. It is obtained from the scale factor by Taylor expanding it about $a(t_0)$:

$$a(t) = a(t_0) + (t - t_0)\dot{a}(t_0) + \frac{1}{2}(t - t_0)^2\ddot{a}(t_0) + \dots$$

Dividing by $a(t_0)$ we write

$$\frac{a(t)}{a(t_0)} = 1 + (t - t_0)H_0 - \frac{q_0 H_0^2}{2}(t - t_0)^2 + \dots$$

It follows by comparing the two series that q_0 is defined by

$$q_0 \equiv -\frac{\ddot{a}(t_0)}{a(t_0)H_0^2} = -\frac{\ddot{a}(t_0)a(t_0)}{\dot{a}^2(t_0)}$$

The deceleration parameter is useful because we can measure it directly on large scales. The observations of Type Ia supernovae suggest in fact that $q_0 \sim -0.6 < 0$ implying $\ddot{a} > 0$ i.e. that

the universe is accelerating today.

1.9 Horizons and distances in cosmology

The cosmological horizon (particle horizon)

We need to think about measuring distance scales in cosmology, as its a vital skill we need to understand if we are to say anything about the make up of the universe. Lets start with a big question, how big is the observable universe, or how far has light travelled since the big bang? Recall the line element for the FRW universe given in Eqn. (1.12)

$$ds^2 = -c^2 dt^2 + a^2(t) \left(\frac{dr^2}{1 - Kr^2} + r^2(d\theta^2 + \sin^2 \theta d\phi^2) \right).$$

For a light ray travelling from $(r = 0, t = t_{\text{em}})$ to $(r = r_0, t = t)$ it travels along a radial null direction hence $ds^2 = 0$ with $(d\phi = d\theta = 0)$.

A neat way of analysing problems associated with the propagation of light is to once again introduce the conformal-time η defined in terms of proper time t by

$$\eta \equiv \int \frac{cdt}{a(t)}.$$

Then using Eqns. (1.13)-(1.19) in Eqn. (1.12) and introducing the conformal time, then the metric takes the form

$$ds^2 = a^2(\eta)(-d\eta^2 + d\chi^2 + S_K^2(\chi)(d\theta^2 + \sin^2 \theta d\phi^2)) \quad (1.72)$$

where to remind you we have

$$\begin{aligned} S_K(\chi) &= \sinh \chi, & K &= -1 \\ S_K(\chi) &= \chi, & K &= 0 \\ S_K(\chi) &= \sin \chi, & K &= +1. \end{aligned}$$

Given that the radial trajectory has $d\phi = d\theta = 0$, we see from (1.72) that the function $\chi(\eta)$ along the light geodesic is completely determined by $ds^2 = 0$ or

$$d\eta^2 - d\chi^2 = 0.$$

The beautiful result that follows is that light geodesics satisfy

$$\chi(\eta) = \pm \eta + \text{const} \quad (1.73)$$

solutions that are straight lines at angles $\pm 45^\circ$ in the $\eta - \chi$ plane. In particular it is true of all geometries, whether it be $K = 0$ or $K = \pm 1$.

Now we can begin to talk about different **horizons**. Light in a universe of finite age can only have travelled a finite distance in that time, meaning that the volume of space within which we can have received signals is finite. The boundary of this volume is the **particle horizon**, and we

would expect it to have a value of order 14 billion light years today, corresponding to the age of the universe. Given the solution Eqn (1.73), the maximum comoving distance light can propagate is

$$\chi_p(\eta) = \eta - \eta_i = \int_{t_i}^t \frac{cdt}{a} \quad (1.74)$$

where η_i (or t_i) is the beginning of the universe. So, at time η , events at $\chi > \chi_p(\eta)$ are inaccessible to us located at $\chi = 0$. It is usually fine to choose $\eta_i = t_i = 0$, especially when there is an initial singularity (big bang), but in some cases of non-singular backgrounds that isn't possible and we take a non-zero value instead – a de Sitter universe is an example. To get the **physical size** of the particle horizon, multiply χ_p by the scale factor today:

$$R(t) = a(t)\chi_p = a(t) \int_{t_i}^t \frac{cdt}{a(t)},$$

which is the **radius of the observable universe** at time t . This radius may be finite or infinite depending on how the scale factor evolves. We can easily see that for cosmological models which decelerate, $R(t)$ is always finite. To show this, consider $a(t) \propto t^\alpha$, ($0 < \alpha < 1$). This gives

$$R(t) = \frac{ct}{1 - \alpha}$$

which is finite at any given time t , but note that it grows linearly with t . For the Einstein-de Sitter Universe ($K=0$, matter dominated), we know $\alpha = \frac{2}{3}$, and in that case today we have

$$R(t_0) = 3ct_0.$$

The implication of this is that light from any galaxy that is now further away from us than $R(t_0)$ can not have reached us by today. The sphere of radius $R(t_0)$ centred on us is said to be our **cosmological horizon**, and is also known as the **particle horizon**. Note that $R(t_0) > ct_0$, the maximum distance light could travel in Minkowski space. How can that be? The reason is that the universe continues to expand as light makes its way across it. $R(t_0)$ is the distance as measured in the present universe. It was smaller earlier and easier to make progress across it.

Event horizon

This can be thought of as the complement of the particle horizon in that it encloses the set of points from which signals sent at a given moment in time t (η) will **never** be received by an observer in the future. In terms of the co-moving coordinates the points are at

$$\chi > \chi_{ev}(t) = \eta_{\max} - \eta = \int_t^{\eta_{\max}} \frac{cdt'}{a(t')}$$

where $\eta_{\max}$ refers to the final moment of conformal time. The **physical size** of the event horizon at time t is

$$R_{ev}(t) = ca(t) \int_t^{\eta_{\max}} \frac{dt'}{a(t')}.$$

Note if the universe expands forever then $\eta_{\max} \rightarrow \infty$. For the case of a $K = 0$ or $K = -1$ decelerating universe, then χ_{ev} and $R_{ev} \rightarrow \infty$. In that case there is no event horizon. However if

the universe is accelerating, then we see that R_{ev} is finite even for $K = 0$ or -1 , hence in that case there is an event horizon. To see this, consider the case of a flat de Sitter space universe given by $a(t) \propto e^{H_\Lambda t}$ where H_Λ is constant. Then we have

$$R_{\text{ev}}(t) = ce^{H_\Lambda t} \int_t^\infty e^{-H_\Lambda t'} dt' = cH_\Lambda^{-1}$$

and is finite, having a size which is the curvature scale of the universe. It means that any event that occurs at a distance larger than cH_Λ^{-1} at a time t can **never** be seen by an observer. Because the space between the event and observer is expanding so rapidly, it can not influence her future. Note that for a closed ($K = 1$) universe which is decelerating, then the time available for future observations is finite because the universe will eventually re collapse. In that case there is both an event horizon and a particle horizon.

Luminosity distance

We now turn our attention to some of the most pressing aspects of observational cosmology, determining the distances to cosmological objects. It is through this that we can talk about determining the cosmological parameters, and yet it is not easy, we cant simply use a tape measure! The Luminosity distance has a simple definition in Minkowski space (with no expansion to cause us any trouble). If we consider a source emitting light with absolute luminosity L_s , then the flux of light we receive $\mathcal{F}$ at a distance d is given by the inverse square law

$$\mathcal{F} = \frac{L_s}{4\pi d^2}$$

in other words the flux is the luminosity per unit area of the sphere of radius d . We dont know what the ‘true’ distance is in an expanding universe, so we use the Minkowski result and turn it into a definition of a new distance scale called the luminosity distance d_L :

$$d_L^2 \equiv \frac{L_s}{4\pi\mathcal{F}}. \quad (1.75)$$

Let us consider an object with absolute luminosity L_s located at a coordinate distance χ_s from an observer at $\chi = 0$. It proves convenient to adopt the metric given in Eqn. (1.72), namely

$$ds^2 = -c^2 dt^2 + a^2(t) [d\chi^2 + S_K^2(\chi)(d\theta^2 + \sin^2\theta d\phi^2)]. \quad (1.76)$$

Now the energy of light emitted from the object with time interval Δt_1 is denoted as ΔE_1 , whereas the energy which reaches us on the sphere with radius χ_s is written as ΔE_0 . We note that ΔE_1 and ΔE_0 are proportional to the frequencies of light at $\chi = \chi_s$ and $\chi = 0$, respectively, i.e., $\Delta E_1 \propto \nu_1$ and $\Delta E_0 \propto \nu_0$. The luminosities L_s and L_0 are given by

$$L_s = \frac{\Delta E_1}{\Delta t_1}, \quad L_0 = \frac{\Delta E_0}{\Delta t_0}. \quad (1.77)$$

The speed of light is given by $c = \nu_1 \lambda_1 = \nu_0 \lambda_0$, where λ_1 and λ_0 are the wavelengths at $\chi = \chi_s$ and $\chi = 0$. Then from $1 + z = \frac{\lambda_0}{\lambda_1} = \frac{a_0}{a}$ we find

$$\frac{\lambda_0}{\lambda_1} = \frac{\nu_1}{\nu_0} = \frac{\Delta t_0}{\Delta t_1} = \frac{\Delta E_1}{\Delta E_0} = 1 + z, \quad (1.78)$$

where we have also used $\nu_0 \Delta t_0 = \nu_1 \Delta t_1$. Combining Eq. (1.77) with Eq. (1.78), we obtain

$$L_s = L_0(1+z)^2. \quad (1.79)$$

The two factors of $(1+z)$ have arisen from the fact that each photon loses energy as it travels from the source to us, and that the number of photons arriving per second decreases over time as the universe expands. The light traveling along the χ direction satisfies the geodesic equation $ds^2 = -c^2 dt^2 + a^2(t) d\chi^2 = 0$. We then obtain

$$\chi_s = \int_0^{\chi_s} d\chi = \int_{t_1}^{t_0} \frac{cdt}{a(t)} = \frac{c}{a_0} \int_0^z \frac{dz'}{H(z')}. \quad (1.80)$$

Note that we have used the relation $\dot{z} = -H(1+z)$ coming from the relation $1+z = \frac{a_0}{a}$. From the metric (1.76) we find that the area of the sphere at $t = t_0$ is given by $S = 4\pi(a_0 S_K(\chi_s))^2$. Hence the observed energy flux is

$$\mathcal{F} = \frac{L_0}{4\pi(a_0 S_K(\chi_s))^2}. \quad (1.81)$$

Substituting Eqs. (1.79) and (1.81) into Eq. (1.75), we obtain the luminosity distance in an expanding universe:

$$d_L = a_0 S_K(\chi_s)(1+z). \quad (1.82)$$

For the case of a flat FRW background with $S_K(\chi) = \chi$ we then find

$$d_L = a_0 \chi_s(1+z).$$

A consequence of this relation is that distant objects appear to be further away than they really are, again because the redshift decreases their apparent luminosity L_0 . We can compare the luminosity distance to the proper or physical distance which is defined at an instant of time. A radial ray of light travels a proper distance given by $ds = a(t)d\chi$. The physical distance to the source is therefore given by integrating this at a fixed time

$$d_{\text{phy}} = a(t) \int_0^{\chi_s} d\chi = a(t_0) \chi_s \quad (1.83)$$

for today. Note that if $z \ll 1$ we have

$$d_L = a_0 \chi_s = d_{\text{phy}} \quad (1.84)$$

in all the curvature cases since from Eqns. (1.17)-(1.19) we have $S_K(\chi_s) \sim \chi_s$ for $\chi_s \ll 1$, which means that objects are really as far away as they look.

Now, the luminosity distance depends upon the cosmological model, hence we can use it to say which model best fits the data. In other words we can plot d_L v z for different cosmologies and compare it to the actual data points. This is what we turn our attention to now. Lets concentrate on the spatially flat case where

$$d_L = a_0 \chi_s(1+z).$$

Using Eqn. (1.80) we have

$$d_L = c(1+z) \int_0^z \frac{dz'}{H(z')}, \quad (1.85)$$

and the Hubble rate $H(z)$ can be expressed in terms of $d_L(z)$:

$$H(z) = \left\{ \frac{d}{dz} \left(\frac{d_L(z)}{c(1+z)} \right) \right\}^{-1}. \quad (1.86)$$

If we measure the luminosity distance observationally, we can determine the expansion rate of the universe! On the other hand substituting for $H(z)$ from Eqn. (1.64) into Eqn. (1.85) we can predict the form of d_L for any given FRW cosmology.

In Fig. 1.2 we plot the luminosity distance (1.85) for a two component flat universe (non-relativistic fluid with $w_m = 0$ and cosmological constant with $w_\Lambda = -1$) satisfying $\Omega_0^{(0)} + \Omega_\Lambda^{(0)} = 1$. Notice that $d_L \simeq z/H_0$ for small values of z . The luminosity distance becomes larger when the cosmological constant is present. We can prove that it should be like this by expanding out the integral. If we take $cH_0^{-1} = 3000h^{-1}\text{Mpc}$, then for this particular case we have

$$d_L = 3000h^{-1} \text{ Mpc} (1+z) \int_0^z \frac{dz'}{[1 - \Omega_0 + \Omega_0(1+z')^3]^{\frac{1}{2}}} \quad (1.87)$$

Solving this numerically gives Figure 1 for different values of Ω_Λ^0 (i.e. today's value). However we can obtain the $z \ll 1$ expansion. After a little bit of algebra we obtain

$$d_L \simeq 3000h^{-1} \text{ Mpc} \left[z + z^2 \left(1 - \frac{3}{4}\Omega_0 \right) + O(z^3) \right],$$

confirming the linear expansion for small z and showing the rather weak dependence on the background cosmology.

Constraints from Supernovae Type Ia – or how to win a Nobel prize in Physics

The direct evidence for the current acceleration of the universe is related to the Nobel prize winning observations of luminosity distances of high redshift supernovae. The apparent magnitude m of the source with an absolute magnitude M is related to the luminosity distance d_L via the relation

$$m - M = 5 \log_{10} \left(\frac{d_L}{\text{Mpc}} \right) + 25. \quad (1.88)$$

This comes from taking the logarithm of Eq. (1.75) by noting that m and M are related to the logarithms of $\mathcal{F}$ and L_s , respectively. The numerical factors arise because of conventional definitions of m and M in astronomy. The Type Ia supernova (SN Ia) can be observed when white dwarf stars exceed the mass of the Chandrasekhar limit and explode. The belief is that SN Ia are formed in the same way irrespective of where they are in the universe, which means that they have a common absolute magnitude M independent of the redshift z . Thus they can be treated as an ideal standard candle. We can measure the apparent magnitude m and the redshift z observationally, which of course depends upon the objects we observe. In order to get a feeling of the phenomenon let us consider two supernovae 1992P at low-redshift $z = 0.026$ with $m = 16.08$ and 1997ap at high-redshift $z = 0.83$ with $m = 24.32$. As we have already mentioned, the luminosity distance is approximately given by $d_L(z) \simeq z/H_0$ for $z \ll 1$. Using 1992P, we find that the absolute magnitude is estimated by $M = -19.09$ from Eq. (1.88). Here we adopted the value $H_0^{-1} = 2998h^{-1}\text{Mpc}$

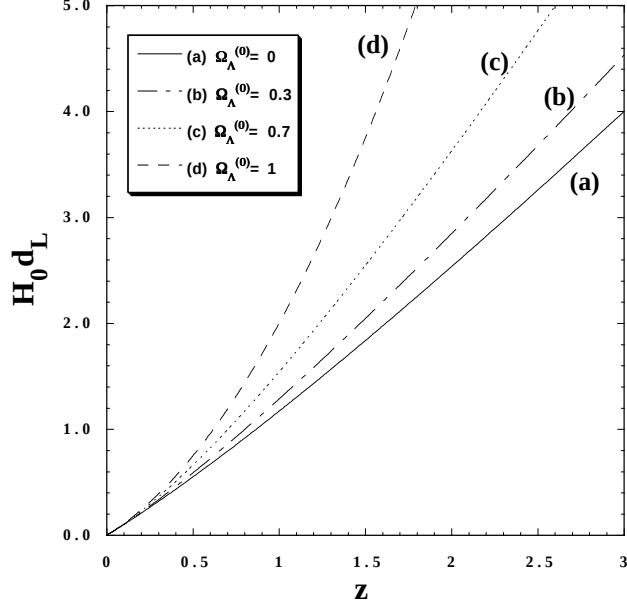


Figure 1.2: Luminosity distance d_L in the units of H_0^{-1} for a two component flat universe with a non-relativistic fluid ($w_m = 0$) and a cosmological constant ($w_\Lambda = -1$). We plot $H_0 d_L$ for various values of $\Omega_\Lambda^{(0)}$.

with $h = 0.72$. Then the luminosity distance of 1997ap is obtained by substituting $m = 24.32$ and $M = -19.09$ for Eq. (1.88):

$$H_0 d_L \simeq 1.16, \quad \text{for } z = 0.83. \quad (1.89)$$

From Eq. (1.85) the theoretical estimate for the luminosity distance in a two component flat universe is

$$H_0 d_L \simeq 0.95, \quad \Omega_0^{(0)} \simeq 1, \quad (1.90)$$

$$H_0 d_L \simeq 1.23, \quad \Omega_0^{(0)} \simeq 0.3, \quad \Omega_\Lambda^{(0)} \simeq 0.7. \quad (1.91)$$

This estimation is clearly consistent with that required for a dark energy dominated universe as can be seen also in Fig. 1.2. Of course, from a statistical point of view, one can not strongly claim that that our universe is really accelerating by just picking up a single data set. Up to 1998 Perlmutter *et al.* [supernova cosmology project (SCP)] had discovered 42 SN Ia in the redshift range $z = 0.18$ - 0.83 , whereas Riess *et al.* [high- z supernova team (HSST)] had found 14 SN Ia in the range $z = 0.16$ - 0.62 and 34 nearby SN Ia. Assuming a flat universe ($\Omega_0^{(0)} + \Omega_\Lambda^{(0)} = 1$), Perlmutter *et al.* found $\Omega_0^{(0)} = 0.28_{-0.08}^{+0.09}$ (1σ statistical) $_{-0.04}^{+0.05}$ (identified systematics), thus showing that about 70 % of the energy density of the present universe consists of dark energy.

In 2004 Riess *et al.* reported the measurement of 16 high-redshift SN Ia with redshift $z > 1.25$ with the Hubble Space Telescope (HST). By including 170 previously known SN Ia data points, they showed that the universe exhibited a transition from deceleration to acceleration at > 99 % confidence level. A best-fit value of $\Omega_0^{(0)}$ was found to be $\Omega_0^{(0)} = 0.29_{-0.03}^{+0.05}$ (the error bar is 1σ). Figure 1.3 illustrates the observational values of the luminosity distance d_L versus redshift z together with the theoretical curves derived from Eq. (1.85). This shows that a matter dominated

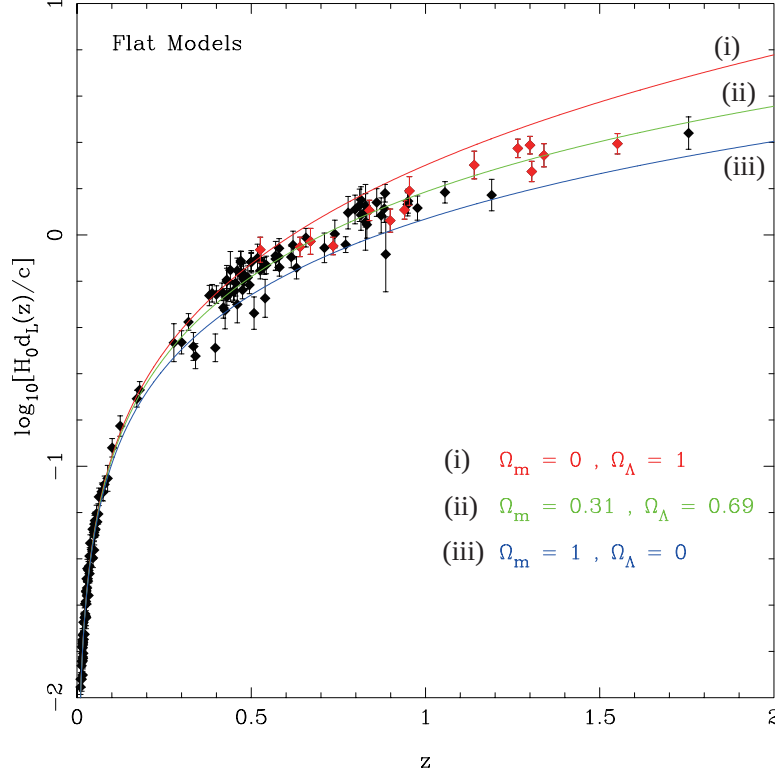


Figure 1.3: The luminosity distance $H_0 d_L$ (log plot) versus the redshift z for a flat cosmological model. The black points come from the “Gold” data sets by Riess *et al.*, whereas the red points show the recent data from HST. Three curves show the theoretical values of $H_0 d_L$ for (i) $\Omega_m^{(0)} = 0$, $\Omega_\Lambda^{(0)} = 1$, (ii) $\Omega_m^{(0)} = 0.31$, $\Omega_\Lambda^{(0)} = 0.69$ and (iii) $\Omega_m^{(0)} = 1$, $\Omega_\Lambda^{(0)} = 0$.

universe without a cosmological constant ($\Omega_0 = 1$) does not fit to the data. A best-fit value of Ω_0 obtained in a joint analysis is $\Omega_0 = 0.31^{+0.08}_{-0.08}$, which is consistent with the result by Riess *et al.*. In 2011, Saul Perlmutter, Brian Schmidt and Adam Riess deservedly shared the Nobel prize for these remarkable observations.

Angular diameter distance : d_{diam}

What follows is based on the book by Mukhanov, *Physical Foundations of Cosmology*, including the maths as well. In particular we will be deriving some of the key results in chapters 1 and 2 of the book.

Objects of a given physical size l are assumed to be perpendicular to our line of sight. If it subtends an angle $\Delta\theta$ (which is always small in astronomy because the distance scales are so large) then we define d_{diam} through

$$d_{\text{diam}} \equiv \frac{l}{\Delta\theta} \quad (1.92)$$

We begin by deriving a few useful results. Recall that in terms of conformal time η (defined through

$cdt = a(\eta)d\eta$ we can write the metric as

$$ds^2 = a^2(\eta)(-d\eta^2 + d\chi^2 + S_K^2(\chi)(d\theta^2 + \sin^2\theta d\phi^2)) \quad (1.93)$$

where χ and $S_K(\chi)$ are defined in Eqns. (1.13)-(1.19). We are going to concentrate on the case of a dust dominated universe, but we could do any cosmology, the technique applies equally well. There is a neat result for the size of the particle horizon in a dust dominated universe. It turns out that for any value of Ω_0 the following result holds

$$S_K(\chi_p) = \frac{2}{a_0 H_0 \Omega_0}. \quad (1.94)$$

Lets prove it – it will bring together many of the results of the course to date.

Case 1: $K = -1$ – dust dominated universe

We quoted the result for the scale factor earlier lecture, but lets derive it here. For dust domination we know $\rho_m \propto a^{-3}$, hence we define the constant

$$M = \frac{4\pi G}{3} \rho_m a^3. \quad (1.95)$$

The Friedmann and acceleration equations are:

$$a'^2 = \frac{8\pi G}{3} \rho a^4 - K a^2 \quad (1.96)$$

$$a'' = \frac{4\pi G}{3} (\rho - 3p) a^3 - K a \quad (1.97)$$

and with $K = -1$ the acceleration equation becomes

$$a'' = M + a.$$

The general solution is

$$a = A \sinh \eta + B \cosh \eta - M,$$

which simplifies with initial condition $a(\eta = 0) = 0$ to

$$a = A \sinh \eta + M(\cosh \eta - 1).$$

The integration constant A is determined by substitution into (1.96) yielding the final solution

$$a(\eta) = M(\cosh \eta - 1). \quad (1.98)$$

Now we want our results in terms of observables such as H_0 and Ω_0 , not in terms of M . We can easily do this, by recalling

$$\Omega_0 \equiv \frac{\rho_m}{\rho_{cr}} = \frac{3M}{4\pi G a_0^3} \times \frac{8\pi G}{3H_0^2} = \frac{2M}{a_0^3 H_0^2},$$

hence

$$\frac{a_0}{M} = \frac{2}{\Omega_0 a_0^2 H_0^2}. \quad (1.99)$$

Making use of the Friedmann equation we have

$$H^2 = \frac{8\pi G}{3}\rho_m + \frac{1}{a^2}$$

Replacing ρ_m with (1.95) and (1.99), this simplifies to give

$$1 + \Omega_0 a_0^2 H_0^2 = H_0^2 a_0^2 \quad (1.100)$$

which will prove useful shortly. Now onto $S_K(\chi_p)$. From (1.17) and (1.74) with $\eta_i = 0$ because we are looking for the particle horizon, we have

$$S_K(\chi_p) = \sinh \chi_p = \sinh \eta \quad (1.101)$$

We can rewrite this in terms of the scale factor using (1.98) to give

$$S_K^2(\chi_p) = \left(\frac{a_0}{M}\right) \left(2 + \frac{a_0}{M}\right) \quad (1.102)$$

Finally using (1.99) and (1.100) we obtain the desired result (1.94):

$$S_K(\chi_p) = \frac{2}{a_0 H_0 \Omega_0}.$$

Case 2: $K = 0$ – dust dominated universe

We adopt the same approach as before, but hopefully it will be a bit quicker. The key equations are (1.18) and (1.74). The solution to the acceleration equation (1.97) for $K = 0$, which satisfies the initial condition $a(\eta = 0) = 0$ and the Friedmann equation (1.96) is

$$a = \frac{M}{2}\eta^2. \quad (1.103)$$

Hence

$$S_K^2(\chi_p) = \frac{2a_0}{M} = \frac{4}{\Omega_0 a_0^2 H_0^2}$$

using (1.99). Recall that $K = 0$ implies by definition that $\Omega_0 = 1$, hence we can include an extra factor of Ω_0 to give the desired result

$$S_K(\chi_p) = \frac{2}{a_0 H_0 \Omega_0}.$$

Case 3: $K = +1$ – dust dominated universe

I leave this one to you my friends. Try it !

We have just obtained the particle horizon for the case of dust dominated cosmologies in the three different curved scenarios. When it comes to measuring the angular diameter distance, we are generally not looking all the way back to the beginning of the universe, rather we are looking back to a redshift z where a galaxy is emitting light that we are detecting today. We need to evaluate $S_K(\chi_{em}(z))$ and so turn our attention to this. Of course, the limit $z \rightarrow \infty$ should recover our result $S_K(\chi_p)$. We will do it again for a matter dominated scenario, and consider the case of

an open $K = -1$ universe. The same result actually applies to the closed universe, and you are encouraged to try and show it. The technique is the same.

We are wanting to evaluate (1.17) using (1.74) but with $\eta_i = \eta_{\text{em}}$ corresponding to the finite conformal time when the light ray was emitted. Therefore we have

$$S_K(\chi_{\text{em}}(z)) = S_K(\eta_0 - \eta_{\text{em}}) = \sinh(\eta_0 - \eta_{\text{em}}).$$

Expanding, we have

$$S_K(\chi_{\text{em}}(z)) = \sinh \eta_0 \cosh \eta_{\text{em}} - \cosh \eta_0 \sinh \eta_{\text{em}}.$$

We can use the solution (1.98) to rewrite this as

$$S_K(\chi_{\text{em}}(z)) = \left(\frac{a_{\text{em}}}{M} + 1 \right) \sqrt{\left(\frac{a_0}{M} + 1 \right)^2 - 1} - \left(\frac{a_0}{M} + 1 \right) \sqrt{\left(\frac{a_{\text{em}}}{M} + 1 \right)^2 - 1}.$$

Now recalling that $(1+z) = \frac{a_0}{a_{\text{em}}}$ for the redshift of the emitting galaxy, and rewriting $\frac{a_{\text{em}}}{M}$ as $\frac{a_{\text{em}}}{a_0} \frac{a_0}{M}$, we can use (1.99) and (1.100) to obtain, after some manipulating:

$$S_K(\chi_{\text{em}}(z)) = \frac{2}{\Omega_0^2 a_0 H_0 (1+z)} \left(1 + \Omega_0 z - \sqrt{1 + \Omega_0 z} + \frac{1 - \sqrt{1 + \Omega_0 z}}{a_0^2 H_0^2} \right)$$

which using (1.100) to write $\frac{1}{a_0^2 H_0^2} = 1 - \Omega_0$ eventually leads to the desired result:

$$S_K(\chi_{\text{em}}(z)) = \frac{2}{\Omega_0^2 a_0 H_0 (1+z)} \left(\Omega_0 z + (\Omega_0 - 2)(\sqrt{1 + \Omega_0 z} - 1) \right). \quad (1.104)$$

Although we have obtained it for the open universe, the result also holds true for the closed $K = 1$ universe, as well as a flat universe. Note, that as $\Omega_0 z \gg 1$ we recover the result for $S_K(\chi_p)$ of (1.94), corresponding to the case where we are going further back in time.

We now turn to the actual calculation of the Angular-diameter-redshift relation, as applied to general curved spacetime, and not just flat spatial sections. Usually we don't worry about an expanding universe, and think Euclidean with a static space in which case an object with a given fixed transverse size l on the sky is subtends an angle which is inversely proportional to the distance to the object, i.e. $d = \frac{l}{\Delta\theta}$. In an expanding universe we have to be more careful. Start with an extended object of given transverse size l situated at a comoving distance χ_{em} from us the observer. We are free to align the object as we want to, and so set $\phi = \text{const}$. Photons emitted from the object at time t_{em} propagate along radial geodesics arriving today with an apparent angular separation $\Delta\theta$. The proper size of the object, l is given by the interval between the emission events at the endpoints. In this case $dt = d\chi = d\phi = 0$ and so in the full metric

$$ds^2 = -c^2 dt^2 + a^2(t) [d\chi^2 + S_K^2(\chi)(d\theta^2 + \sin^2 \theta d\phi^2)], \quad (1.105)$$

we have

$$l = \sqrt{|ds^2|} = a(t_{\text{em}}) S_K(\chi_{\text{em}}) \Delta\theta. \quad (1.106)$$

Comparing this with Eqn. (1.92) and Eqn. (1.82) we see that

$$d_{\text{diam}} = a(t_{\text{em}}) S_K(\chi_{\text{em}}) = a(t_0)(1+z)^{-1} S_K(\chi_{\text{em}}) = d_L(1+z)^{-2} \quad (1.107)$$

or

$$d_L = d_{\text{diam}}(1+z)^2 \quad (1.108)$$

for all curved spaces.

Returning to d_{diam} we see that the angle subtended by the object is

$$\Delta\theta = \frac{l}{a(t_{\text{em}})S_K(\chi_{\text{em}})} = \frac{l}{a(\eta_0 - \chi_{\text{em}})S_K(\chi_{\text{em}})}, \quad (1.109)$$

the final term arising because the physical time t_{em} corresponds to the conformal $\eta_{\text{em}} = \eta_0 - \chi_{\text{em}}$. A **nearby object** satisfies $\chi_{\text{em}} \ll \eta_0$, hence

$$a(\eta_0 - \chi_{\text{em}}) \simeq a(\eta_0), \quad S_K(\chi_{\text{em}}) \simeq \chi_{\text{em}},$$

where the final equality follows from (1.17)-(1.19) using the approximation $\chi_{\text{em}} \ll 1$. It follows that

$$\Delta\theta \simeq \frac{l}{a(\eta_0)\chi_{\text{em}}} = \frac{l}{d_{\text{phy}}}, \quad (1.110)$$

where the physical distance is given by Eqn. (1.83), hence we see that for nearby objects, $\Delta\theta$ is inversely proportional to the physical distance as we would expect, which of course from Eqn. (1.84) is the same as the luminosity distance. What if the object is **far away**, say near the particle horizon? In that limit $\eta_0 - \chi_{\text{em}} \ll \eta_0$, so

$$a(\eta_0 - \chi_{\text{em}}) \ll a(\eta_0), \quad S_K(\chi_{\text{em}}) \rightarrow S_K(\chi_p) = \frac{2}{a_0 H_0 \Omega_0},$$

where $S_K(\chi_p)$ was derived in (1.94) and of importance here is the fact it is constant. The angular size of the object now becomes

$$\Delta\theta \propto \frac{l}{a(\eta_0 - \chi_{\text{em}})}, \quad (1.111)$$

which increases with distance away from us. In fact as it approaches the horizon its image covers the whole sky ! So, the angular size of objects peak nearby as expected but also far away. Why then isn't the sky full of these images of distant objects? Its because the luminosity drops off rapidly with increasing distance, so the remote object do not outshine the nearby ones.

How can we imagine this behaviour? For an incomplete but useful analogy, lets go down a dimension and live at the north pole on the surface of a 2-sphere again. We are looking at the way the size of a given object (say a hug iceberg) varies as we change its distance from us. The object lies across lines of latitude, meaning that the light from it travels to us on lines of longitude (or meridians), because these are the geodesics on the earths surface. We find that if we are north of the equator, the angular size of the iceberg decrease as it goes further away towards the equator. However, once south of the equator, its angular size increases as we go further south, eventually covering the whole sky at the south pole. In actual fact, the angular size of a very remote object grows in a flat universe as well, because the scale factor is changing with time.

Angular diameter size is usually given as a function of redshift z . Using

$$1+z = \frac{a_0}{a(t_{\text{em}})}$$

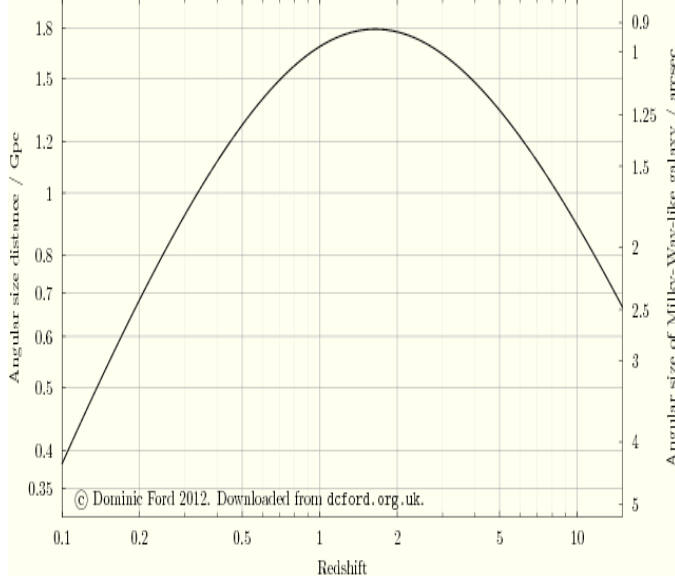


Figure 1.4: The angular distance versus redshift for a flat matter dominated universe – credit Dominic Ford, dcford.org.uk

equation (1.109) becomes

$$\Delta\theta = (1+z) \frac{l}{a_0 S_K(\chi_{\text{em}}(z))}, \quad (1.112)$$

where $\chi_{\text{em}}(z)$ is obtained from the usual definition in the following way: the comoving distance to an object that emitted a photon at t_{em} which arrives today is

$$\chi = \eta_0 - \eta_{\text{em}} = \int_{t_{\text{em}}}^{t_0} \frac{cdt}{a(t)}.$$

Changing variable to z using $(1+z) = \frac{a_0}{a(t)}$ we have

$$dz = -\frac{a_0}{a^2(t)} \dot{a}(t) dt = -(1+z)H(t)dt,$$

from which two neat things follow, the age of the universe at a redshift z is given by

$$t = \int_z^\infty \frac{dz}{(1+z)H(z)} \quad (1.113)$$

and we can write

$$\chi_{\text{em}}(z) = \frac{1}{a_0} \int_0^z \frac{cdz}{H(z)}. \quad (1.114)$$

Ok, now we have to cut to the chase and start putting in some solutions.

Lets start with a flat universe filled with dust, ($K = 0, p = 0$). From (1.18) we know that $S_K(\chi_{\text{em}}) = \chi_{\text{em}}$, hence we need $\chi_{\text{em}}(z)$. For that we require $H(z)$ in (1.114). This is given by Eqn. (1.64) with $\Omega_{m0} = \Omega_0 = 1, \Omega_{v0} = \Omega_{r0} = 0$.

We can easily solve the integrals in (1.113) and (1.114) to give

$$t(z) = \frac{2}{3H_0} \frac{1}{(1+z)^{3/2}} \quad (1.115)$$

$$\chi(z) = \frac{2c}{a_0 H_0} \left(1 - \frac{1}{\sqrt{1+z}} \right) \quad (1.116)$$

Substituting (1.116) into (1.112) we obtain

$$\Delta\theta = \frac{lH_0}{2c} \frac{(1+z)^{3/2}}{(1+z)^{1/2} - 1}. \quad (1.117)$$

Notice the small and large z limits (recalling $H_0 = \frac{2}{3t}$):

$$\Delta\theta = \frac{l}{3ct_0} \frac{(1 + \frac{3}{2}z + O(z^2))}{(1 + \frac{1}{2}z - 1 + O(z^2))} \simeq \frac{2l}{3ct_0 z} \quad z \ll 1,$$

$$\Delta\theta = \frac{z}{3ct_0} \quad z \gg 1.$$

In both limits therefore the object appears large. In fact objects appear at their smallest when $\frac{d\Delta\theta}{dz} = 0$. It then follows by differentiating (1.117) that after a little bit of straightforward algebra, the corresponding redshift is given by

$$\frac{d\Delta\theta}{dz} = 0 \longrightarrow z = \frac{5}{4},$$

as can be seen in Figure. (1.4). The angular diameter distance can easily be obtained for more general cosmologies given the general result (1.112) and (1.104) for the case of dust dominated non-flat universe:

$$\Delta\theta = \frac{lH_0}{2} \frac{\Omega_0^2(1+z)^2}{\Omega_0 z + (\Omega_0 - 2)((1 + \Omega_0 z)^{1/2} - 1)}. \quad (1.118)$$

We can look at the small and large z behaviour, which is as in the flat case. That means there is a minimum somewhere? Where is it as a function of Ω_0 ?

The use of angular diameter versus redshift to test cosmological models has met with limited success to date, mainly because of the lack of standard rulers. One exception though is the single standard ruler obtained from measurements of the CMB. It has been possible to measure temperatures in two random directions in the sky, and the temperature difference depends on the angular separation. Measuring the power spectrum associated with this temperature difference shows a series of peaks and troughs as the angular separation is varied from large to small scales. The ‘first acoustic peak’ is determined by the sound horizon at recombination, which corresponds to the maximum distance a sound wave in the baryon-radiation fluid can have propagated by recombination. This sound horizon acts as a standard ruler of length $l_s \sim H^{-1}(z_r)$. Recombination occurs at $z_r \simeq 1100$. Now since we are at such a large redshift, it implies $\Omega_0 z_r \gg 1$, so we can set $\chi_{\text{em}}(z_r) = \chi_p$. This then means we can use $S_K(\chi_p)$ which we have evaluated for a dust dominated universe in (1.94). Substituting into (1.112) we obtain

$$\Delta\theta_r \simeq \frac{z_r H_0 \Omega_0}{2H(z_r)} \simeq \frac{1}{2} z_r^{-1/2} \Omega_0^{1/2} \simeq 0.87^\circ \Omega_0^{1/2}, \quad (1.119)$$

having used $H_0/H(z_r) \simeq (\Omega_0 z_r^3)^{-1/2}$ from (1.64). The beauty of this result is that it only depends on Ω_0 , so the first doppler peak determines the spatial curvature of the universe ! The results to date suggest everything is consistent with a flat $\Omega_0 = 1$ universe.

1.10 Age of the Universe

Determining the age of the Universe is one of the big challenges in cosmology and vital if we are to understand its evolution. Eqn. (1.113) provides the age at a redshift z in any cosmology given by $H(z)$ in Eqn. (1.64). To get the present age we take $z \rightarrow 0$ in the integral. In general this can not be solved analytically but a few easy cases can be seen quickly.

At $10 < z < 1000$, where matter dominates, we have $H \simeq \frac{2}{3t}$ hence from Eqn. (1.64) this corresponds to

$$t(z) \simeq \frac{2}{3} H^{-1}(z) \simeq \frac{2}{3} H_0^{-1} \Omega_{m0}^{-\frac{1}{2}} (1+z)^{-\frac{3}{2}} \quad (1.120)$$

For a flat universe, the current age is $H_0 t_0 \simeq (2/3) \Omega_{m0}^{-\frac{1}{2}}$. One of the early pieces of evidence for the need of a cosmological constant type term was when independent tests indicated the product $H_0 t_0 \sim 1$. This required a very low Ω_{m0} to be consistent.

2 Thermal History of the Universe – the Hot Big Bang

One of the more useful relations we can derive is that for temperature v time or energy v time. It allows us to quickly estimate when some of the major events occurred in the history of the Universe. If we want to include all forms of relativistic particles, and not just photons, we have to also consider in the sum for today the contribution of neutrinos. Neutrinos are fermions, obeying Fermi-Dirac statistics, and so they have a different number density than photons (there is a famous factor of 7/8). Moreover when they decoupled from the matter, they had a different temperature than the photons (another famous factor of $(\frac{4}{11})^{\frac{4}{3}}$). Let's spend a few minutes discussing the origin of these numbers and the fact that the number of light degrees of freedom is temperature dependent (i.e. there were more than just photons and neutrinos in the early universe). Of particular importance in the analysis is the notion of statistical mechanics both in and out of equilibrium. The key objects we would like to determine are the number densities, energy densities and pressure.

2.1 Number densities, energy densities and pressures – relativistic and non-relativistic cases

For a particle species A (with mass m) in statistical equilibrium, the number density n , energy density ρ and pressure p are given as integrals over the distribution function $f_A(\mathbf{p}, t)$ where $\mathbf{p}$ is the 3-momentum of the particle. Different species of particles interact, exchange energy and momentum. Now the rate of interaction is $\Gamma(t) = n \langle \sigma v \rangle$ where σ is the interaction cross section and v is the velocity of the particles. As long as $\Gamma(t) > H(t)$ the Hubble expansion parameter, then these interactions lead to and maintain thermodynamic equilibrium among the interacting particles with some temperature T . In general the interactions have a short range, we may assume that the role of these interactions is just to provide a mechanism for thermalisation, and they do not determine the form of the distribution function. Particles may be treated as an ideal (Bose or

Fermi) gas, with an equilibrium distribution function:

$$f_A(\mathbf{p}, t) d^3\mathbf{p} = \frac{g_A}{(2\pi)^3} (\exp[(E_A - \mu_A)/k_B T_A] \pm 1)^{-1} d^3\mathbf{p} \quad (2.1)$$

where k_B is the Boltzmann constant, g_A is the spin degeneracy factor determining the number of relativistic particles present at any given temperature, μ_A is the chemical potential, T_A is the temperature of this species and $E(\mathbf{p}) = \sqrt{\mathbf{p}^2 c^2 + m^2 c^4}$. The “+” sign corresponds to fermions and the “-” sign to bosons. For a gas in thermal equilibrium the chemical potential is always zero. That is because there are no overall changes in the particle number and if you recall your first law of thermodynamics μ_A is associated with such a change through $dE = TdS - PdV + \mu_A dN_A$.

Given the distribution function, we can obtain the background number density, energy density and pressure of the particles n , ρ and p .

$$\begin{aligned} n &= \frac{1}{\hbar^3} \int f(\mathbf{p}) d^3\mathbf{p} = \frac{g}{2\pi^2 \hbar^3} \int (\exp[E(p)/k_B T] \pm 1)^{-1} p^2 dp \\ &= \frac{g}{2\pi^2 c^3 \hbar^3} \int_{mc^2}^{\infty} \frac{(E^2 - m^2 c^4)^{1/2}}{(\exp[E/k_B T] \pm 1)} E dE \end{aligned} \quad (2.2)$$

$$\rho c^2 = \frac{1}{\hbar^3} \int E(p) f(\mathbf{p}) d^3\mathbf{p} = \frac{g}{2\pi^2 c^3 \hbar^3} \int_{mc^2}^{\infty} \frac{(E^2 - m^2 c^4)^{1/2}}{(\exp[E/k_B T] \pm 1)} E^2 dE \quad (2.3)$$

$$p = \frac{1}{\hbar^3} \int \frac{|pc|^2}{3E(p)} f(\mathbf{p}) d^3\mathbf{p} = \frac{g}{6\pi^2 c^3 \hbar^3} \int_{mc^2}^{\infty} \frac{(E^2 - m^2 c^4)^{3/2}}{(\exp[E/k_B T] \pm 1)} dE \quad (2.4)$$

The factors of $\hbar$ are present because we are dealing with identical particles and in quantising them for the energy levels we are going from a discrete to a continuous representation of the momentum ($\sum_{\mathbf{p}} \rightarrow \frac{V}{\hbar^3} \int d^3p$). We can now begin to consider the different behaviour of these functions for relativistic and non-relativistic species. It is useful to introduce $x \equiv E/k_B T$

1. Relativistic species : $mc^2 \ll k_B T$

$$n = \left(\frac{k_B T}{c} \right)^3 \frac{g}{2\pi^2 \hbar^3} \int_0^{\infty} \frac{x^2 dx}{(\exp[x] \pm 1)} \propto T^3 \quad (2.5)$$

Now since the Riemann zeta function $\zeta(n) \equiv \sum_{m=1}^{\infty} m^{-n} = (1/\Gamma(n)) \int_0^{\infty} u^{n-1} du / (\exp(u) - 1)$, (note $\zeta(3) \simeq 1.202$; $\zeta(4) = \frac{\pi^2}{90} \simeq 1.0823$) we have for bosons:

$$n_B = \left(\frac{k_B T}{c\hbar} \right)^3 \frac{g\zeta(3)}{\pi^2} \quad (2.6)$$

where as for fermions, by using the intriguing identity (which by the way implies that the distribution of fermions looks like a mixture of bosons at two different temperatures, one half the other)

$$\frac{1}{\exp(x) + 1} = \frac{1}{\exp(x) - 1} - \frac{2}{\exp(2x) - 1} \quad (2.7)$$

we then obtain

$$n_F = \left(\frac{k_B T}{c\hbar} \right)^3 \frac{g\zeta(3)}{\pi^2} \left(1 - \frac{1}{4} \right) = \frac{3}{4} n_B \propto T^3 \quad (2.8)$$

Putting in some numbers, the cosmic microwave background photons today have a temperature of $T = 2.725$ K, hence from Eqn. (2.6) (with $g = 2$) that gives a number density $n_\gamma = 4.1 \times 10^8 \text{ m}^{-3}$.

For the energy density and pressure we obtain

$$\rho_B c^2 = (k_B T)^4 \frac{g}{2\pi^2 c^3 \hbar^3} \int_0^\infty \frac{x^3 dx}{(\exp[x] - 1)} = \frac{\pi^2}{30 c^3 \hbar^3} g (k_B T)^4 \propto T^4 \quad (2.9)$$

$$\rho_F c^2 = \left(1 - \frac{1}{8}\right) \frac{\pi^2}{30 c^3 \hbar^3} g (k_B T)^4 = \frac{7}{8} \rho_B c^2 \quad (2.10)$$

with

$$p = \frac{\rho c^2}{3} \quad (2.11)$$

for both cases.

A few points worth mentioning: Note the factor of $7/8$ which appears in the fermion energy density. It arises simply from the fact Fermions satisfy Fermi-Dirac statistics as opposed to the Bose-Einstein statistics satisfied by Bosons. Eqn. (2.11) allows us to obtain the equation of state for radiation but derived from statistical mechanics, and as expected we find that $w = p/(\rho c^2) = 1/3$. Eqn. (2.9) is the famous Stefan-Boltzmann law $\rho_\gamma = \sigma_{SB} T^4$. Since we know from earlier lectures that the energy density in radiation scales as $\rho_\gamma \propto a^{-4}$, then combining the two results leads to the result presented in class that the temperature of the radiation (and of any relativistic species) scales like

$$T_\gamma \propto \frac{1}{a} \quad (2.12)$$

This of course means the universe was much hotter when it was smaller. We are thinking of the temperature of the radiation as the ‘temperature of the universe’, because it is well defined as the radiation has a thermal spectrum, so a well defined temperature and because at early times the other particle species interact with the radiation and so share its temperature. This eventually breaks down as the universe cools down and particles drop out of thermal equilibrium.

We can also think about the **entropy of the background**. In thermodynamics we think of entropy and energy as extensive quantities (i.e. they are additive for subsystems, being proportional to the amount of material in the system). This means $\partial S/\partial V = S/V$ and $\partial E/\partial V = E/V$ where we have $E(T, V)$ and $S(T, V)$. Starting from

$$dE = TdS - PdV \quad (2.13)$$

we then substitute for dE and dT giving

$$\frac{\partial E}{\partial T} dT + \frac{\partial E}{\partial V} dV = T \frac{\partial S}{\partial T} dT + T \frac{\partial S}{\partial V} dV - PdV \quad (2.14)$$

This is true for arbitrary changes dT and dV so collecting terms we end up with

$$\frac{\partial E}{\partial T} = T \frac{\partial S}{\partial T} \quad \text{from dT term} \quad (2.15)$$

$$S = \frac{E + PV}{T} \quad \text{from dV term} \quad (2.16)$$

In the the ultrarelativistic limit we have been considering ($k_B T \gg mc^2$) we have using Eqn. (2.9) and (2.11) in Eqn. (2.16) the **entropy density** $s = S/V$

$$s = \frac{4}{3} \frac{\rho_B c^2}{T} = \frac{2\pi^2 k_B}{45c^3 \hbar^3} g(k_B T)^3 \quad (2.17)$$

with a factor of (7/8) of this for fermions. Recalling the number density of relativistic particles as given in Eqn. (2.5) also scales as T^3 then we see that the entropy density also counts the number of particles. This is why we can say that the ratio of the number density of photons in the universe to the number density of baryons is called the entropy per baryon. It has a value of order 10^9 today.

2. Non-relativistic (massive) species : $mc^2 \gg k_B T$

$$\begin{aligned} n &= \frac{g(k_B T)^3}{2\pi^2 c^3 \hbar^3} \int_{mc^2/k_B T}^{\infty} \exp(-x) (x^2 - (mc^2/k_B T)^2)^{1/2} dx \\ &\simeq \frac{g}{\hbar^3} \left(\frac{mk_B T}{2\pi} \right)^{3/2} \exp(-mc^2/k_B T) \end{aligned} \quad (2.18)$$

$$\rho = mn \quad (2.19)$$

$$p \simeq n(k_B T) \ll \rho \quad (p \simeq 0) \quad (2.20)$$

For the maths geeks amongst us, perhaps a few words are in order here about how we can get these results. Consider the number density Eqn. (2.18). Introduce $x = \mu y$ where $\mu = mc^2/k_B T$ we see that we have

$$n = \frac{g(k_B T)^3}{2\pi^2 c^3 \hbar^3} \left(\frac{mc^2}{k_B T} \right)^3 \int_1^{\infty} \exp(-\mu y) (y^2 - 1)^{1/2} y dy \quad (2.21)$$

Now using

$$I_\nu(\mu) \equiv \int_1^{\infty} \exp(-\mu y) (y^2 - 1)^{(\nu-1)/2} y dy = \frac{2^{\nu-1/2}}{\sqrt{\pi}} \mu^{1/2-\nu} \Gamma(\nu) K_{(\nu+1/2)}(\mu) \quad (2.22)$$

where $\Gamma(\mu)$ is a Gamma function and $K_\nu(\mu)$ is a Bessel function of the second kind, we see

$$n = \frac{g(mc)^3}{2\pi^2 \mu \hbar^3} K_2(\mu). \quad (2.23)$$

We are in the regime $\mu \gg 1$, hence we look for asymptotic expansions. In this regime

$$\text{Lim}_{\mu \rightarrow \infty} K_\nu(\mu) \simeq \sqrt{\frac{\pi}{2\mu}} \exp(-\mu) \quad (2.24)$$

then we see

$$n \simeq \frac{g(mc)^3}{2\pi^2 \mu \hbar^3} \sqrt{\frac{\pi}{2\mu}} \exp(-\mu) = g \left(\frac{mk_B T}{2\pi} \right)^{3/2} \exp(-mc^2/k_B T) \quad (2.25)$$

as in Eqn. (2.18). Lets look to derive Eqn. (2.19). Starting with Eqn. (2.3), under the substitution $x = E/k_B T$, it becomes

$$\rho c^2 = \frac{g}{2\pi^2 c^3 \hbar^3} (k_B T)^4 \int_{mc^2/k_B T}^{\infty} \exp(-x) (x^2 - (mc^2/k_B T)^2)^{1/2} x^2 dx$$

which as before under $x = \mu y$ becomes

$$\rho c^2 = \frac{g}{2\pi^2 c^3 \hbar^3} (mc^2)^4 \int_1^\infty \exp(-\mu y) (y^2 - 1)^{1/2} y^2 dy \quad (2.26)$$

Hence

$$\rho = \frac{gm}{2\pi^2 \hbar^3} (mc)^3 \int_1^\infty \exp(-\mu y) (y^2 - 1)^{1/2} y^2 dy = -\frac{gm}{2\pi^2} (mc)^3 \frac{d}{d\mu} \left(\frac{K_2(\mu)}{\mu} \right) \quad (2.27)$$

from Eqn. (2.22). Now from the integral definition of the Bessel function

$$K_\nu(z) = \int_0^\infty e^{-z \cosh t} \cosh(\nu t) dt \quad (2.28)$$

it follows by direct differentiation that $\frac{dK_2(z)}{dz} = -\frac{1}{2}(K_3(z) + K_1(z))$. Therefore in Eqn. (2.27) we have

$$\rho = \frac{gm}{2\pi^2 \hbar^3} (mc)^3 \left(-\frac{1}{2\mu} (K_3(\mu) + K_1(\mu)) - \frac{1}{\mu^2} K_2(\mu) \right) \quad (2.29)$$

Once again taking the limit $\mu \gg 1$ we obtain

$$\rho \simeq -\frac{gm}{2\pi^2 \hbar^3} (mc)^3 \frac{1}{\mu} \sqrt{\frac{\pi}{2\mu}} \exp(-\mu) = mn \quad (2.30)$$

from Eqn. (2.25) as advertised in Eqn. (2.19). To determine the pressure in Eqn. (2.20) from Eqn. (2.4) we need the additional piece of information

$$\int_1^\infty \exp(-\mu y) (y^2 - 1)^{(\nu-1)} dy = \frac{2^{\nu-1/2}}{\sqrt{\pi}} \mu^{1/2-\nu} \Gamma(\nu) K_{(\nu-1/2)}(\mu). \quad (2.31)$$

It follows from Eqn. (2.4) that $\nu = 5/2$ hence

$$p = \frac{g}{6\pi^2 c^3 \hbar^3} (mc^2)^4 \frac{K_2(\mu)}{\mu^2} \frac{4}{\sqrt{\pi}} \Gamma(5/2) \quad (2.32)$$

This simplifies recalling Eqn. (2.23) to give

$$p = n(k_B T) \quad (2.33)$$

as given in Eqn. (2.20).

A useful quantity is the **baryon to photon ratio** today $\eta \equiv \frac{n_B}{n_\gamma}$. We have seen that today $n_\gamma \sim 411 \text{ cm}^{-3}$, and we have $n_B = \frac{\rho_b}{m_b} = \frac{\Omega_b \rho_c}{m_b}$ hence substituting for m_b as the mass of the proton ($m_p \sim 1.67 \times 10^{-24} \text{ gm}$) and using Eqn. (1.52) we obtain

$$\eta \equiv \frac{n_B}{n_\gamma} = 5.5 \times 10^{-10} \left(\frac{\Omega_b h^2}{0.02} \right). \quad (2.34)$$

There are of order one billion photons for every baryon in the universe, and the typical density of baryons is $n_B = \frac{\rho_b}{m_b} \simeq 0.22 \text{ m}^{-3}$.

We can summarise. In general for relativistic species the number densities go as T^3 and the energy density behaves as T^4 , while for massive species they are suppressed by the Boltzmann factor $\exp(-mc^2/k_B T)$. In fact it is Eqn. (2.18) that plays an important role in **nucleosynthesis**. The exponential suppression of the number density means that non-relativistic particles soon drop below the limit where they interact sufficiently often to stay in equilibrium.

Dealing with several relativistic species : number of degrees of freedom

How do we interpret the g factors in the equations we have just presented and do they change with temperature? When we have a collection of relativistic species, each of them in equilibrium at different temperatures T_i , we can write the total energy density ρ_R , summing over all the contributions:

$$\rho_R c^2 = \frac{k_B^4 T_\gamma^4}{2\pi^2 \hbar^3 c^3} \sum_i g_i \left(\frac{T_i}{T_\gamma} \right)^4 \int_{x_i}^{\infty} \frac{u^3 du}{(\exp[u] \pm 1)} \quad (2.35)$$

where T_γ is the temperature of the photons. This can be rewritten in a more compact form as

$$\rho_R c^2 = \frac{(k_B T_\gamma)^4}{30 \hbar^3 c^3} \pi^2 g_* \quad (2.36)$$

where g_* is the ‘effective’ number of degrees of freedom, given by

$$g_* = \sum_{\text{bosons}} g_i \left(\frac{T_i}{T_\gamma} \right)^4 + \frac{7}{8} \sum_{\text{fermions}} g_i \left(\frac{T_i}{T_\gamma} \right)^4 \quad (2.37)$$

As the temperature of the photons decreases, the effective number of degrees of freedom in radiation will decrease, as massive particles become non-relativistic when their mass becomes larger than T_i . To be more precise:

$T \ll 1 \text{ MeV}$: the only relativistic particles would be the 3 neutrino species (fermions with 2 degrees of freedom each) and the photon (boson, 2 polarisation states). Neutrinos at this temperature are decoupled from the thermal bath (as we shall see below) and they are slightly colder than the photons, with a temperature $T_\nu = (4/11)^{1/3} T_\gamma$. We then have

$$g_* = 2 + \frac{7}{8} \times 3 \times 2 \times \left(\frac{4}{11} \right)^{4/3} \simeq 3.36$$

$1 \text{ MeV} \leq T \leq 100 \text{ MeV}$: Electrons and positrons have a mass of about 0.5 MeV, and so they are now also relativistic. As the difference between neutrino and photon temperature is due to the electron-positron annihilation, we have $T_\nu = T_\gamma$ and so

$$g_* = 2 + \frac{7}{8} \times (3 \times 2 + 2 \times 2) = 10.75$$

$T \leq 300 \text{ GeV}$: this is above the electroweak unification scale, and for particles in the standard model we have $g_* \simeq 106.75$.

Temperature v Time

The temperature was hotter in the past, a result that follows from the evidence that the universe has been expanding adiabatically. Given that the temperature of the present day cosmic microwave background (CMB) has been measured to be

$$T_0 = 2.725 \pm .001\text{K} \quad (2.38)$$

then we usually assume the temperature it evolves as

$$T(z) = 2.725(1+z). \quad (2.39)$$

We have argued for this based on the fact the energy density of photons scales as $\rho_{\text{rad}} \propto (1+z)^4$ and also from the statistical discussion earlier we see from Eqn. (2.9) that $\rho_{\text{rad}} \propto T^4$. Actually a more accurate argument for this relationship is based on the adiabatic expansion assumption, in other words that entropy is conserved. We have seen earlier that the entropy density scales as $s \propto a^{-3}$ and $s \propto T^3$, hence this requires $T \propto (1+z)$.

Relativistic particles

Lets assume we only have contributions from photons and massless neutrinos. We have from Eqns. (1.52) and (2.9)

$$\Omega_{\text{rad}} = \frac{\rho_{\text{rad}}}{\rho_c} = \frac{\pi^2}{30c^5\hbar^3\rho_c} g(k_B T)^4 \quad (2.40)$$

hence we find that today

$$\Omega_{\text{rad}} = 2.47 \times 10^{-5} h^{-2} \quad (2.41)$$

Considering neutrinos and assuming for the sake of argument that they are massless then we have (see below for an explanation of the factor of (4/11))

$$\Omega_{\nu} = 3 \times \frac{7}{8} \times \left(\frac{4}{11}\right)^{\frac{4}{3}} \Omega_{\text{rad}} \simeq 0.68 \Omega_{\text{rad}} = 1.68 \times 10^{-5} h^{-2}.$$

The total contribution today from relativistic particles then follows

$$\Omega_{\text{rel}} = \Omega_{\nu} + \Omega_{\text{rad}} = 4.15 \times 10^{-5} h^{-2}$$

which of course is negligibly small compared to $\Omega_0 \sim 0.3$, the density of non-relativistic material today. We are now in a position to look at the time evolution of these quantities. Recall $\rho_{\text{rel}} \propto a^{-4}$ and $\rho_{\text{mat}} \propto a^{-3}$, we then obtain at any time t

$$\frac{\Omega_{\text{rel}}}{\Omega_{\text{mat}}} = \frac{4.15 \times 10^{-5}}{\Omega_0 h^2} \times \frac{a_0}{a} = \frac{4.15 \times 10^{-5}}{\Omega_0 h^2} \times (1+z). \quad (2.42)$$

A very important epoch is **equality** when the matter and radiation densities were equal. This occurs when

$$(1+z_{\text{eq}}) = 24074 \Omega_0 h^2 \left(\frac{T_0}{2.725 K} \right)^4$$

where we have reinserted the T dependence from Eqn. (2.9) so that it is clear how the redshift of equality depends on the temperature today. For $T > T_{\text{eq}}$, we know $a < a_{\text{eq}}$, hence $\frac{\Omega_{\text{rel}}}{\Omega_{\text{mat}}} > 1$ in Eqn. (2.42) indicating that we are then in a Universe dominated by relativistic particles, or a

radiation dominated Universe. Similarly for $T < T_{\text{eq}}$ we are in a matter dominated regime. For example at **decoupling** occurs where $(1 + z_{\text{dec}}) = 10^3$ we see

$$\frac{\Omega_{\text{rel}}}{\Omega_{\text{mat}}} = \frac{0.04}{\Omega_0 h^2}.$$

It follows that unless $\Omega_0 h^2 \ll 1$, the universe was matter dominated at decoupling.

Let's just summarise the temporal history as we have seen it so far, working backwards. We will assume a flat model with $\Omega = 0.3$ and $h = 0.7$:

1. Today – $t_0 = 13.7$ Gyr, $T_0 = 2.725\text{K}$, $z = 0$.
2. Distant galaxies – $t = 1\text{Gyr}$, $T = 16\text{K}$, $z = 5$,
3. Decoupling – $t_{\text{dec}} \sim 350,000\text{years}$, $T_{\text{dec}} \sim 3000\text{K}$, $z \sim 1100$ – formation of microwave background – the last time photons had enough energy to ionise atoms. Note this was in matter dominated era.
4. Equality – $t_{\text{eq}} \sim 3400 \Omega_0^{-\frac{3}{2}} h^{-3}$ years, with $T_{\text{eq}} = 66000 \Omega_0 h^2$ K, $z \sim 24,000$ – the moment of equal matter and radiation energy densities.
5. Nucleosynthesis – $t \sim 1$ sec, $T \sim 10^{10}\text{K}$, $z = 10^{10}$ – photons typically have enough energy to overcome nuclear binding energy of atoms $\sim \text{O}(\text{MeV})$.
6. Nucleon pair threshold – $t \sim 10^{-6.6}$ sec, $T \sim 10^{13}\text{K}$, $z = 10^{13}$ – Photons destroy nuclei, split neutrons and protons away from each other, leaving the Universe as a sea of separate protons neutrons and electrons. The corresponding epoch known as Nucleosynthesis marks the transition to atomic nuclei.
7. Electroweak unification – $t \sim 10^{-12}$ sec, $T \sim 10^{15}\text{K}$, $E \sim 250\text{GeV}$, $z = 10^{15}$ – Probing the Electroweak era where the weak and electromagnetic force is unified into one force.
8. Grand Unification – $t \sim 10^{-36}$ sec, $T \sim 10^{28}\text{K}$, $E \sim 10^{15}\text{GeV}$, $z = 10^{28}$ – Unification of string force with ewk force. Epoch of inflation, topological defects ...
9. Quantum Gravity – $t \sim -43$ sec, $T \gg 10^{32}\text{K}$, $E \gg 10^{19}\text{GeV}$, $z = 10^{32}$ – area of speculation including unifying gravity with the other forces

2.2 The density of the universe and dark matter

There are a number of independent estimates of the density of matter in the universe, and on the forms it can take. The density appears to increase as we look on larger and larger length scales, eventually reaching a constant value. We are using the definition where the total density in matter is given by Ω_0 . Recall first of all, the characteristic scale for the density of the universe is the critical density

$$\rho_c(t_0) = 1.88 h^2 \times 10^{-26} \text{kg m}^{-3} = 2.78 h^{-1} \times 10^{11} M_{\odot} (h^{-1} \text{Mpc})^{-3}.$$

By counting stars we have $\Omega_{\text{stars}} \equiv \frac{\rho_{\text{stars}}}{\rho_c} \simeq 0.005 \rightarrow 0.01$, which is very small compared to unity. Nucleosynthesis provides a very tight constraint on the baryonic matter contribution giving

$$0.016 \leq \Omega_B h^2 \leq 0.024,$$

which is important because it implies for $h \sim 0.7$ that $\Omega_B \geq 0.03 \gg \Omega_{\text{stars}}$, hence there is more baryonic material in the universe than is visible in stars. Over all baryons appear to accounts for between 3 and 5% of the critical density. As we move to larger scales the need for some non-baryonic component of matter becomes clear. First of looking at Virial dynamics of galaxy clusters and and Galaxy rotation curves we infer the existence of a dark matter spherical halo surrounding the luminous region of our galaxy, and the estimates are

$$\Omega_{\text{halo}} \sim 0.1$$

clearly incompatible with just baryonic matter.

Galaxy clusters are the largest gravitationally collapsed objects in the Universe, which implies that they contain almost all forms of matter, because all matter is affected by gravity. Bounds arising from studying these objects suggest $\Omega_0 \simeq 0.3h^{-\frac{1}{2}} \simeq 0.35$. Note that Ω_0 is dominated by dark matter, however it remains less than the critical density case of $\Omega_0 = 1$, hence $\rho_{DM} < \rho_c$. Observations of the bulk motions of galaxies relative to one another (i.e. deviations from the cosmological principle). These bulk flows reflect the gravitational interactions which depends on the **total** mass of galaxies and leads to a similar bound, as do observations of large scale structures arising as a result of gravitational instabilities. Numerical models of structure formation indicate that we again require $\Omega_0 > 0.2$ for the models to work.

Determining the precise geometry provides a wonderful route to determine Ω_0 . The geometry is accessed primarily through precision cosmic microwave background experiments, in particular through the location of the acoustic peaks associated with the anisotropies in the CMB. Structure formation scenarios tend to predict a characteristic angular scale of around one degree for these CMB features, the precise scale depends on the geometry of the universe. Recent observations from WMAP, and older ones from Boomerang and Maxima have measured these features and the general conclusion is that they are consistent with a spatially flat ($K = 0$) universe, with $\Omega_0 + \Omega_\Lambda = 1 \pm 0.1$. The observations related to the CMB and large-scale structure (LSS) independently support the ideas of a dark energy dominated universe. The position of the first acoustic peak around $l = 200$ constrains the curvature of the universe to be $|1 - \Omega_{\text{total}}| = 0.030^{+0.026}_{-0.025} \ll 1$ which as we will see is predicted by the inflationary paradigm. Using the most recent WMAP data, then combining WMAP and the Supernova legacy Survey implies $\Omega_K = -0.015^{+0.02}_{-0.016}$, consistent with a flat universe. Combining with the Hubble Space Telescope key project constraint on H_0 provides a tighter constraint, $\Omega_K = -0.010^{+0.016}_{-0.009}$ and $\Omega_\Lambda = 0.72 \pm 0.04$ (to be compared with earlier pre WMAP3 results $\Omega_\Lambda^{(0)} = 0.69^{+0.03}_{-0.06}$, which assumed a flat universe with a prior for the Hubble constant $h = 0.71 \pm 0.076$).

In figure (1.1) we plot the confidence regions coming from SN Ia, CMB(WMAP1) and large-scale galaxy clustering. Clearly the flat universe without a cosmological constant is ruled out. The compilation of three different cosmological data sets strongly reinforces the need for a dark energy dominated universe with $\Omega_\Lambda^{(0)} \simeq 0.7$ and $\Omega_0^{(0)} \simeq 0.3$. Amongst the matter content of the universe, baryonic matter amounts to only 4 %. The rest of the matter (27 %) is believed to be in the form of a non-luminous component of non-baryonic nature with a dust like equation of state ($w = 0$) known as Cold Dark Matter (CDM). Dark energy is distinguished from dark matter in the sense that its equation of state is different ($w < -1/3$), allowing it to give rise to an accelerated expansion. The observation of the Bullet Cluster in 2006 has been seen by many as a smoking gun for dark



Figure 2.1: Bullet cluster – taken by CHANDRA. The collision has taken place and the inferred dark matter distributions are in blue and the measured hot gas distribution in red.

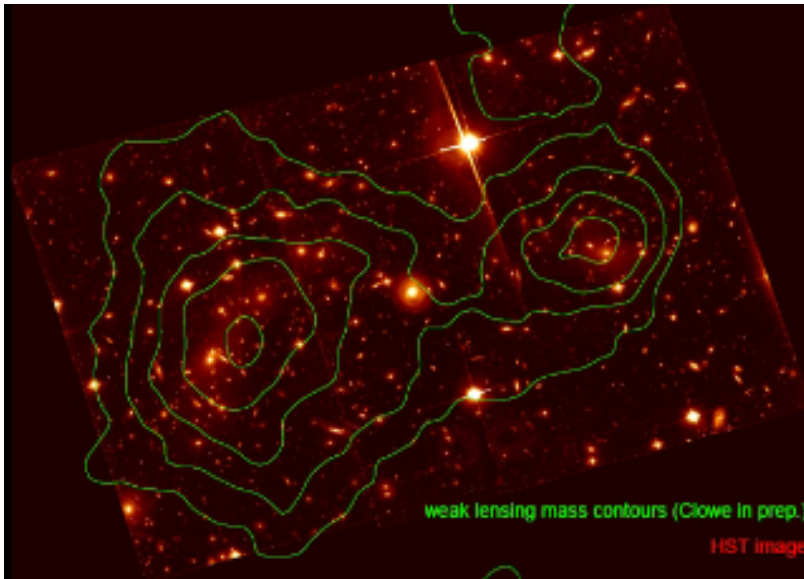


Figure 2.2: Bullet cluster – the mass density contours superimposed over the photograph of the same region taken by HST. Note the two concentrated regions showing how the dark matter from the two clusters have passed through each other.

matter (see Clowe et al in *Astrophys.J.648:L109-L113,2006*). Two colliding clusters of galaxies passed through each other around 150 million years ago. By studying it we can investigate the distribution of stars, hot X-ray gas and indirectly, dark matter in the carnage of the collision. The stars which are observable in the visible light were not really affected by the collision, and passed right through being slowed by gravity. However, the hot gas from the two clusters, when seen in X-rays comprise most of the ordinary (baryonic) mass in the cluster pair. It interacts strongly electromagnetically, meaning they lose energy and slow right down compared to the stars, showing up in the central region as very hot X-rays. The dark matter is collisionless and passes right through, again only affected by gravity. It is detected indirectly by gravitational lensing of background objects and seems impressive confirmation of the dark matter paradigm, although there are many questioning the conclusion, trying to reproduce the features with modified theories of gravity and searching for more colliding clusters. In the two figures from NASA, figure (2.1) taken by CHANDRA shows the inferred dark matter distribution as blue and the measured hot gas distribution in red. Figure (2.2) taken with the HST shows the mass density contours superimposed over the photograph of the same region.

2.3 Dark Matter Candidates

We have seen that the majority of the dark matter in the Universe is non-baryonic, so now we have to ask the question what is it then if it isn't made of baryons? There are a number of candidates around the majority of them from the world of particle physics. The favourite types are known as **WIMPS**, Weakly Interacting Massive Particles. This is a very short resume of the favourite possibilities. First of all the three types can be summarised as:

1. **Hot Dark Matter (HDM)** Relic Particles that decouple when relativistic, hence have a number density approximately equal to that of photons. The favourite example are eV-mass neutrinos, and their relic density is proportional to the particle mass.
2. **Warm Dark Matter (WDM)** Particle that are able to decouple early enough so that the relative abundance of photons can then be boosted by annihilations other than just through electron-positron annihilation. Given that there are of order 100 distinct particle species, the critical particle mass of these WDM particles required to make $\Omega = 1$ is around 1-10 keV. These were out of favour, but issues in numerical simulations of LSS suggest they may yet be required to fit the data !
3. **Cold Dark Matter (CDM)** Relic particles that decouple while they are nonrelativistic, implying from Eqn. (2.18) that their number density can be exponentially suppressed. For interactions that mimic those of neutrinos, the freezeout temperature is about 1 MeV, and the corresponding relic mass density then falls with increasing mass. For weak interactions, since the cross-sections scale with energy as E^2 , it means that the relic density falls with particle mass as m^{-2} . As a result, cosmologically interesting masses are of order 10 GeV range, implying that they cannot be any of the known neutrinos (we would have seen them already in accelerators). However, as we go beyond about 90 GeV (the mass of the Z boson), the strength of the weak interaction is reduced, with cross-section going as E^{-2} . Hence the relic density now rises as m^2 , leading to the observed dark matter density being reached at $m \simeq 1$ TeV – within the reach of the LHC. Favourite candidates of this sort of CDM particle are from the world of supersymmetry, for example the neutralino.

2.4 Dark Matter Searches

There are many experiments searching for dark matter particles using either direct detection methods or indirect detection methods. The former is possible because these particles which pervade the universe (remember they have to account for the rotation curves we see) are so abundant that even though they are weakly interacting they will occasionally interact with protons and neutrons and the collision can lead to a signal. However the numbers are both huge and small, huge numbers of WIMP particles, incredibly small interaction rates! The Boulby mine in North Yorkshire is on a WIMP hunt. They expect an interaction rate of order one per day per kg of detecting material, and they need to be underground (1100 metres down) in order to try and eliminate the dangerous confusion background events. To check on the validity of a signal is incredibly demanding. One possible route is through the fact we expect there to be some annular modulation of the signal. There should be a prevailing flow of dark matter in the solar neighbourhood and at some parts of its orbit, the earth goes generally in the direction of the flow, there by decreasing the flux we receive. Six months or so later, it goes against the flow and so we should see a larger signal! Time will tell! Tantalising claims have been made for the discovery of these particles with an energy scale of below 10 GeV or so, but the statistical significance of the data is not yet reliable enough to believe the results. (For a nice fairly up to date review of dark matter see: “Dark Matter Candidates”, Lars Bergstrom, New J.Phys.**11**:105006,2009 – SPIRES arXiv:0903.4849v4)

Let us summarise the state of play with regards candidates thanks to my colleague Dr Anne Green:

Axions: $10^{-6} \text{eV} < m < 10^{-3} \text{eV}$

These are no WIMPS Roughly speaking : masses larger than 10^{-3}eV are ruled out by some combination of (depending on the axion coupling constants) energy loss from SN 1987A, cooling of stars in globular clusters and accelerator searches. The lower limit actually comes from the cosmological axion density and is something like 10^{-6}eV , but there’s probably at least an order of magnitude uncertainty in that (from our lack of knowledge of how the physics of the QCD phase transition works, and the difficulties in calculating the abundance of axions produced by cosmic strings).

WIMPs (or any thermal relics):

There aren’t any really firm pure experimental/observational limits. If the WIMP is too light (roughly keV) it will be warm and will be able to free-stream far enough to erase structure on dwarf galaxy scales. There’s a (model independent) theoretical upper limit of a couple of hundred TeV from partial-wave unitarity of the S matrix. From a theoretical point of view in SUSY models the lightest neutralino is typically in the range 10s of GeV to a few TeV.

PBHs: $10^{15} \text{ grams} < M \leq 10^{26} \text{ grams}$

Lower limit: they’re evaporating today (and producing too many gamma-rays), upper limit: microlensing

Finally we should always take on board the possibility that dark matter does not exist and what we are seeing in these large scale features are evidence of modifications of Einstein’s theory of General Relativity. We are not going to go into this route here, but it is an area that has gained considerable popularity not least because the dark matter is proving very illusive, we have a problem explaining the nature of the cosmological constant, and some of the models used to describe

the early universe, such as models arising out of string theory are by default modified gravity theories. Missions are now being proposed to directly test for such modifications, for example EUCLID.

2.5 Freezeout and Relics

In determining the nature of the dark matter candidates we need to discuss when they come to dominate. So far in our earlier calculations, we have assumed thermodynamic equilibrium always occurs in the early universe, but that is not always the case for all particles. Consider for example the annihilation of electron positron pairs ($e^+ e^- \leftrightarrow 2\gamma$). Equilibrium is maintained when the process occurs equally fast in both directions. However, as the temperature of the photon pairs drops, then a typical photon energy will eventually be too low to create particle pairs, in which case there will be nothing to balance the annihilation of the particles. The annihilations occur at a finite rate governed by the Boltzmann equation for the number density n , of say electrons

$$\dot{n} + 3Hn + \langle \sigma v \rangle n^2 = S \quad (2.43)$$

where σ is the reaction cross section, v is the particle velocity and S is a source term representing thermal particle production. Neglecting S initially, clearly there are two key timescales in this equation which govern the rate of change of n . The first is the *expansion* timescale given by $H(z)^{-1}$ and the second is the *interaction* timescale given by $(\langle \sigma v \rangle n)^{-1}$. Both of these timescales increase as the universe expands but the interaction time changes fastest (at least as fast as a^3) where as H changes no faster than (a^2) in the radiation era, so there is always a crossover. So we have thermal equilibrium at early times where the interaction time is much shorter than the expansion time, but as the universe expands the situation reverses and we evolve to a state of *freeze out* or *decoupling* at late times. This corresponds to a situation where the particle has effectively stopped interacting and provides vital information of how the universe was at the time the particle was last in thermal eqm. Freeze out implies there are whole range of relics existing from different stages of the hot big bang, an example being photons of the CMB generated at $z \sim 1100$; as well as neutrinos; axions; neutralinos ... For a given particle therefore, once we are at the Freezeout temperature for that particle defined by

$$\Gamma(T_f) \equiv \langle \sigma v \rangle_{T_f} n^{eq}(T_f) \simeq H(T_f), \quad (2.44)$$

then annihilations cease and the WIMP comoving number density becomes constant. Different WIMP candidates experience this at different times and so we have a series of particle species freezing out at different epochs as the universe expands and cools down.

The source term in Eqn. (2.43) can be accounted for by realising that in thermal equilibrium in a non-expanding universe ($H = 0$), the number density remains fixed at the equilibrium value it has corresponding to the particular temperature T , say n_T . Therefore $S = \langle \sigma v \rangle n_T^2$. Returning to the Boltzmann equation, and substituting for S we see that it can be rewritten as

$$\frac{d}{dt}(a^3 n) = - \langle \sigma v \rangle n(a^3 n) + \langle \sigma v \rangle n \frac{(a^3 n_T)^2}{(a^3 n)} \quad (2.45)$$

Introducing the comoving number density $N \equiv a^3 n$ (recall $n \propto a^{-3}$, hence N is comoving), and using $\Gamma = \langle \sigma v \rangle n$ for the particle interaction rate, we see that

$$\frac{d}{dt}N = -\Gamma N \left(1 - \left(\frac{N_T}{N} \right)^2 \right) \quad (2.46)$$

or using $\frac{d}{dt} = H \frac{d}{d \ln a}$,

$$\frac{d \ln N}{d \ln a} = -\frac{\Gamma}{H} \left(1 - \left(\frac{N_T}{N} \right)^2 \right) \quad (2.47)$$

This needs to be solved numerically but a few features can be seen that are useful for us to understand the nature of the freeze-out. Consider the case where the universe is expanding rapidly enough to sustain a population in almost thermal equilibrium, i.e. $N \simeq N_T$, because $\dot{N} \simeq 0$. We have seen earlier for the case of relativistic particles that $n_T \propto T^3$ (see Eqn. (2.5), and since $T \propto a^{-1}$, it follows that N_T is constant. In particular it means that it is possible to keep $N = N_T$ exactly, independent of Γ/H . However, this does not mean the population remains in thermal equilibrium. For $\Gamma/H \ll 1$ a particle experiences effectively no interactions and remember the universe is constantly growing in size so lowering its temperature. For the other extreme of being a non-relativistic particle, recall Eqn. (2.18) where the thermal distribution of such particles of mass m are exponential suppressed $n_T \propto (mk_B T)^{3/2} \exp(-mc^2/k_B T)$. Therefore the comoving number of the particles would evolve as $(mk_B)^{3/2} T^{-3/2} \exp(-mc^2/k_B T)$. Now this means that the term $\frac{d \ln N}{d \ln a}$ will be large in magnitude and negative (basically from the $e^{-m/T}$ term in the number density). Consider what happens as T decreases, the number density drops rapidly. For this to be maintained on the rhs of the rate equation we can have $\frac{\Gamma}{H} \gg 1$ whilst $N \simeq N_T$. What happens though once $\frac{\Gamma}{H} \ll 1$? Now, as N_T begins to drop rapidly with a , the term $(N_T/N)^2$ rapidly becomes negligible, leaving us with $\frac{d \ln N}{d \ln a} \simeq -\frac{\Gamma}{H} \ll 1$. A point is reached where the reaction rate has dropped so much that the particles are basically conserved as the universe expands, the population is **frozen-out**. It provides a more rigorous definition of freeze-out or decoupling and matches the approximate regime which is Eqn. (2.44) defined as $N(a \rightarrow \infty) = N_T(\Gamma/H = 1)$. Figure (2.3) shows how freeze-out occurs as a function of temperature and how the final density depends on the interaction cross-section.

We can obtain the associated present day density of a non-relativistic relic. Associating freeze-out with the condition when $\Gamma/H = 1$ then from Eqn. (2.44) we have

$$n_f = \frac{H_f}{\langle \sigma v \rangle} \quad (2.48)$$

where n_f is the number density of the relics at freeze-out. The present relic density of this particular particle mass m is given by

$$\Omega_{\text{relic},0} = \frac{\rho_{\text{relic},0}}{\rho_{c,0}} = \frac{m n_{\text{relic},0}}{\rho_{c,0}} = \frac{8\pi G m n_{\text{relic},0}}{3H_0^2} \quad (2.49)$$

Assuming the background dynamics is dominated by relativistic radiation at freeze-out then we have the Friedmann equation

$$H_f^2 = \frac{8\pi G \rho_{\text{rad},f}}{3} \quad (2.50)$$

where the radiation energy density is given by Eqn. (2.36). We now need to link the relic density at freeze-out with temp T_f Eqn. (2.48) to the present day value $n_{\text{relic},0}$. We do this by recalling that the definition of freeze-out is effectively when the particle comoving number N freezes at a given value, it is conserved as the universe expands. and so its number entropy density must have fallen at the same rate as the entropy density which is given by Eqn. (2.17), i.e.

$$\frac{n_f}{n_{\text{relic},0}} = \frac{g_{*f} T_f^3}{g_{*0} T_0^3} \quad (2.51)$$

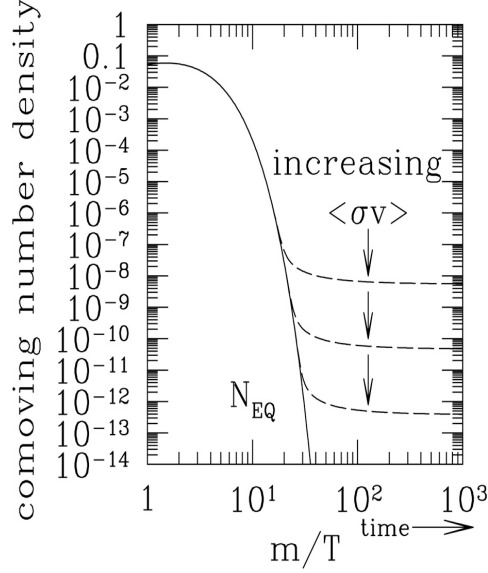


Figure 2.3: The comoving number density N of a typical relic particle as a function of m/T and of interaction cross section. Note that as the cross-section increases the final relic density decreases. Also note that freezeout occurs when $\frac{mc^2}{k_B T_f} \sim 10$ – figure c/o Paolo Gondolo – <http://ned.ipac.caltech.edu/level5/Sept05/Gondolo/Gondolo2.html>

Substituting for $n_{\text{relic},0}$ in Eqn. (2.49) and using Eqn. (2.48) with H_f given by Eqn. (2.50) we obtain, after a bit of careful rearranging

$$\Omega_{\text{relic},0} = \left(\frac{8\pi G}{3} \right)^{3/2} \left(\frac{\pi^2}{30\hbar^3 c^9} \right)^{1/2} \left(\frac{g_{*0}(k_B T_0)^3}{H_0^2 <\sigma v>} \right) \left(\frac{mc^2}{k_B T_f} \right) g_{*f}^{-1/2} \quad (2.52)$$

Using $H_0 = 3.26 \times 10^{-18} h \text{ sec}$, $g_{*0} \simeq 3.36$ for the low temperature effective number of degrees of freedom, and $T_0 = 2.725 K$ it follows that

$$\Omega_{\text{relic},0} h^2 = \frac{10^{-33} [\text{m}]^2}{<\sigma v>} \left(\frac{mc^2}{k_B T_f} \right) g_{*f}^{-1/2} \quad (2.53)$$

where $[\text{m}]^2$ indicates the units are in metres². Hence we have a prediction for the present relic density as a function of the mass, cross-section and temperature of freezeout. Now the actual simulations indicate the typical value of $\frac{mc^2}{k_B T_f} \sim 10$ (as seen in Figure (2.3)) and for the high energy regimes where freeze out occurs, $g_{*f} \sim 100$, hence the final two factors in Eqn. (2.53) effectively cancel. We can replace the velocity v by the speed of light c because at freeze-out the particles are nearly relativistic. Doing that we reach the final result

$$\Omega_{\text{relic},0} h^2 \simeq 0.03 (\sigma/\text{pb})^{-1} \quad (2.54)$$

where the picobarn is a very small area, $1\text{pb} = 10^{-40} \text{m}^2$. It shows that it is only a relatively small range of annihilation cross-sections that will be of interest from an observational point of view. As can be seen from the figure, Eqn. (2.54) makes sense. The higher the cross-section, the lower the relic density. That is because as σ increases, there are more interactions, the longer the particle

stays in equilibrium and the more annihilation events to decrease the number density.

The thermal history of neutrino decoupling

Here we explain the origin of the mysterious factor of $(11/4)^{1/3}$ associated with the temperature of neutrinos which had decoupled from the photons. At around $T \simeq 10^{12} K \simeq O(100) \text{ MeV}$, the energy density of the universe is almost all in relativistic particles $e^\pm$, ν , $\bar{\nu}$ and photons. They are in equilibrium with the same temperature, hence the effective number of degrees of freedom is $g_* = 10.75$ as we have just seen. The corresponding rate of expansion in this radiation dominated regime is

$$H^2(T) = \frac{8\pi G}{3} \rho_R \quad (2.55)$$

where ρ_R is given by Eqn. (2.36). Neutrinos are kept in equilibrium via weak interaction processes ($\nu\bar{\nu} \leftrightarrow e^+e^-, \dots$) with a cross section given by

$$\sigma_F \simeq G_F^2 E^2 \simeq G_F^2 T^2 \quad (2.56)$$

where $G_F = 1.1664 \times 10^{-5} \text{ GeV}^{-2}$ is the Fermi constant. The interaction rate per (massless) neutrino is:

$$\Gamma_F = n \langle \sigma_F v \rangle \simeq 1.3 G_F^2 T^5. \quad (2.57)$$

The factor of T^5 comes from the number density (T^3) and the cross-section (T^2). From the expressions for Γ_F and $H(T)$ we obtain (after substituting in the correct numbers)

$$\frac{\Gamma_F}{H(T)} = \frac{0.24 T^3 G_F^2}{\sqrt{8\pi G}} \simeq \left(\frac{T}{1 \text{ MeV}} \right)^3 \quad (2.58)$$

Therefore neutrinos decouple from the rest of the matter when $T_D \simeq 1 \text{ MeV}$. Once we get below this temperature the neutrino temperature scales as a^{-1} . Now the key thing that happens, occurs just below neutrino decoupling because the temperature drops below the mass of the electron ($T < 0.5 \text{ MeV}$). At the same time, the entropy in the electron-positron pairs is transferred to the photons, but not to the neutrinos. We then have

$$g_*(T_D > T > m_e) = 2 + \frac{7}{8} \times 4 = \frac{11}{2}, \quad g_*(T < m_e) = 2 \quad (2.59)$$

Here comes the key thing. We know that for the particles which are in equilibrium with radiation, **entropy is conserved** where $S = g_*(aT)^3$. So we can equate the entropies before and after $T < m_e$:

$$(g_*)_{\text{after}} (aT_\gamma)_{\text{after}}^3 = (g_*)_{\text{before}} (aT_\gamma)_{\text{before}}^3 \quad (2.60)$$

or

$$\frac{(aT_\gamma)_{\text{after}}^3}{(aT_\gamma)_{\text{before}}^3} = \frac{(g_*)_{\text{before}}}{(g_*)_{\text{after}}} = \frac{11}{4}. \quad (2.61)$$

Now neutrinos do not participate in this process and their entropy is separately conserved. But before e^+e^- annihilation began, photons and neutrinos had the same temperature. Therefore we have

$$(aT_\gamma)_{\text{after}} = \left(\frac{11}{4} \right)^{1/3} (aT_\gamma)_{\text{before}} = \left(\frac{11}{4} \right)^{1/3} (aT_\nu)_{\text{before}} = \left(\frac{11}{4} \right)^{1/3} (aT_\nu)_{\text{after}} \quad (2.62)$$

The temperature of the photons is larger than that of the neutrinos today by a factor $(11/4)^{1/3}$. Therefore given that today $T_\gamma = 2.725\text{K}$, the corresponding distribution of background relativistic relic neutrinos has an effective temperature

$$T_\nu = \left(\frac{4}{11}\right)^{1/3} T_\gamma = 1.945 \text{ K}. \quad (2.63)$$

Massive neutrinos

Recent high precision experiments have confirmed that the neutrinos are not massless, or at least not all of them, they have a small mass. Given that relic neutrinos are abundant, this could be important for cosmology. In fact we know from Eqn. (2.8) that at a given temperature T the number density of relativistic fermions is related to that of bosons by $n(\nu + \bar{\nu}) = (3/4)n(\gamma, T = 1.945\text{K})$, which gives a relic number density of around 112 relic neutrinos in every cm^3 for each species. If these neutrinos were ultra relativistic at decoupling, then as the universe expands to $k_B T < m_\nu c^2$, the total number of neutrinos is preserved, meaning that the present-day mass density in neutrinos is the number density of massless neutrinos times m_ν . For light neutrinos this implies that today their cosmological density is given by

$$\Omega_\nu = \frac{\sum n_i m_i c^2}{\rho_c} = \frac{112 \sum m_i c^2}{1.88 h^2 \times 10^{-29} \text{ gm}^{-1}} \quad (2.64)$$

But we have the conversion between grams and eV, $1\text{gm} \equiv 5.6 \times 10^{32} \text{ eV}/c^2$ hence the present density in neutrinos is given by

$$\Omega_\nu h^2 = \frac{\sum m_i c^2}{94.1 \text{ eV}}. \quad (2.65)$$

Direct laboratory limits on the masses are

$$\nu_e \leq 2.2 \text{ eV} \quad \nu_\mu \leq 0.17 \text{ MeV} \quad \nu_\tau \leq 15 \text{ MeV}. \quad (2.66)$$

However cosmology provides tighter constraints. For example, large scale structure constraints suggest $\sum m_i < 0.68 \text{ eV}$. Now we also know from neutrino mixing experiments in which each neutrino type is a mixture of energy eigenstates that the energy difference can be measured. These give a direct measurement of the difference in the square of the masses. To see this consider the relativistic energy equation $E^2 = p^2 c^2 + m^2 c^4$ and expand to get $E = pc + m^2 c^3 / 2p$. These mixings are known now from experiment detecting neutrinos generated either in the sun or in the Earth's atmosphere. They give for the mass differences

$$\begin{aligned} \Delta(m_{21})^2 &= 8 \times 10^{-5} \text{ eV}^2 \\ \Delta(m_{32})^2 &= 2.5 \times 10^{-3} \text{ eV}^2 \end{aligned} \quad (2.67)$$

where m_1, m_2 and m_3 are the three mass eigenstates. From this we do not have the absolute mass scales, rather differences. There are two possible regimes: the **normal hierarchy** with $m_3 \gg m_2 \gg m_1$ or the **inverted hierarchy** with $m_1 \simeq m_2 \gg m_3$. Cosmology may well provide the solution as it will be possible directly measure the total density in neutrinos. The easiest case is the normal hierarchy with m_1 negligible and the mass dominated by m_3 which is around 0.05 eV. Time will tell if this turns out to be correct.

2.6 Baryogenesis

The freeze-out calculations assume baryons and anti-baryons freeze out under the same conditions and at the same rate. there should be no difference between their relic number densities. Yet there is, as far as we know the number density of anti-particles is negligible compared to that of particles. In fact for every billion anti-particles there will have been one extra particle (one billion and one) in the high energy early universe. This would then account for the observed abundance of baryonic particles compared to photons, where $n_B/n_\gamma = 10^{-9}$. A big unanswered question is what caused this initial asymmetry, and is known as baryogenesis. It is thought to be an early universe process but the standard model can not generate a large enough initial asymmetry. Some new physics is required. In particular If baryon number is conserved, this imbalance cannot be altered once it is set in the initial conditions; but what generates it?

2.7 Nucleosynthesis – the origin of the light elements

Nucleosynthesis is regarded by many as one of the pillars on which the Hot Big Bang stands proud. It explains the formation and abundance of the lightest nuclei, Hydrogen, Deuterium, Helium 3, Helium 4 and Lithium 7. Generally covered in basic cosmology courses, we just briefly review the salient points here. It occurs around 1 sec into the universe when the energy scales suitable for nuclei formation are reached:

$$\left(\frac{1 \text{ sec}}{t}\right)^{\frac{1}{2}} = \left(\frac{T}{2 \times 10^{10} \Omega_0^{\frac{1}{4}} h^{\frac{1}{2}} \text{ K}}\right) = \left(\frac{k_B T}{2 \Omega_0^{\frac{1}{4}} h^{\frac{1}{2}} \text{ MeV}}\right)$$

To simplify matters consider only the formation of helium-4 nuclei, with the left over material being in hydrogen nuclei (i.e. individual protons). We also assume $m_{\text{proton}} c^2 = 938.3 \text{ MeV} < m_{\text{neutron}} c^2 = 939.6 \text{ MeV}$, free neutrons decay to protons with half-life given by $t_{\frac{1}{2}} \simeq 614 \text{ sec}$, stable isotopes of the light elements exist, and the neutrons bound into them do not decay. In other words, once a neutron has become part of a stable isotope it no longer decays.

The protons and neutrons are in thermal equilibrium at high energies and as the universe cools down they can bind into nuclei. When $k_B T > O(\text{MeV})$ which is the nuclear binding energy, but when the particles are non-relativistic i.e $k_B T < m_p c^2$, we have $O(\text{MeV}) \leq k_B T < m_p c^2$. In that energy regime, the particles are in thermal equilibrium with a Maxwell-Boltzmann number density $\mathcal{N} \propto (mT)^{\frac{3}{2}} e^{-\frac{mc^2}{k_B T}}$. Hence $\frac{\mathcal{N}_n}{\mathcal{N}_p} = \left(\frac{m_n}{m_p}\right)^{\frac{3}{2}} \exp\left[-\frac{(m_n - m_p)c^2}{k_B T}\right]$. Now since $m_n \sim m_p$, the prefactor is $O(1)$, hence in the regime $k_B T \gg (m_n - m_p)c^2$ then $\mathcal{N}_n \sim \mathcal{N}_p$ implying that early on there were identical numbers of protons and neutrons in the Early Universe. At these energies and temperatures the equilibrium conversion reactions were primarily $n + \nu_e \leftrightarrow p + e^-$; $n + e^+ \leftrightarrow p + \bar{\nu}_e$ where ν_e and $\bar{\nu}_e$ are the electron neutrino and its anti-particle respectively. The neutrons and protons remain in thermal equilibrium with the ratio $\frac{\mathcal{N}_n}{\mathcal{N}_p}$ given above if the reactions proceed rapidly enough. This happens until the universe has cooled so there is no longer enough energy available for the interactions to proceed in both directions. This corresponds to the interaction rate becoming longer than the age of the universe at that time. It occurs when $k_B T \simeq 0.8 \text{ MeV}$, and it marks the moment when the relative abundances of protons and neutrons become fixed $\frac{\mathcal{N}_n}{\mathcal{N}_p} = \exp\left[-\frac{1.3 \text{ MeV}}{0.8 \text{ MeV}}\right] \simeq \frac{1}{5}$.

For $k_B T < 0.8 \text{ MeV}$, only the decay of free neutrons can change the abundance further. Now the formation of the light elements arises from a complex reaction chain with nuclear fusion leading to the formation of the nuclei. Remember though the effect of the high energy photon tail of the

distribution which tends to break up the newly formed nuclei, and so as with estimating the temperature of decoupling, nucleosynthesis occurs at a lower temperature than you might originally have guessed. As an example of the type of reactions involved, if we consider the formation of Deuterium and Helium-3 and Helium-4 $p + n \leftrightarrow D$; $D + p \leftrightarrow {}^3\text{He}$; $D + D \leftrightarrow {}^4\text{He}$. the destruction processes which happen in the opposite direction occur less and less frequently as the universe cools, so eventually the nuclei can build up. Applying the same high energy tail argument, but this time to the Deuterium binding energy of 2.2 MeV, the nuclei begin to be stable when the energy available is around 0.1 MeV. After this moment the nuclei can begin to build up.

For $0.1 \text{ MeV} < k_B T < 0.8 \text{ MeV}$, a small fraction of the free neutrons decay into protons. How many decay? From the temperature-time relationship we see that an energy of $k_B T = 0.1 \text{ MeV}$ corresponds to a time of around $t_{\text{nuc}} \sim 400$ seconds, a number that is about 2/3 that of the neutron half life of $t_{\text{half}} \sim 614 \text{ sec}$. As a result neutron decays reduce $\mathcal{N}_n$ by $\exp\left[-\ln 2 \frac{t_{\text{nuc}}}{t_{\text{half}}}\right]$. With this suppression we see that by the time the nuclei become stable the ratio has reduced to $\frac{\mathcal{N}_n}{\mathcal{N}_p} = \frac{1}{5} \times \exp\left[-\frac{400 \text{ sec} \times \ln 2}{614 \text{ sec}}\right] \simeq \frac{1}{8}$. Only Hydrogen and Helium form in any significant amount because ${}^4\text{He}$ is the most stable of the light nuclei and Hydrogen forms because there are not enough neutrons around for all the protons to bind with implying some protons are left over. We can estimate the relative abundance of $H : {}^4\text{He}$ quoted as a mass fraction (not a number density) of the universe in ${}^4\text{He}$. Because a ${}^4\text{He}$ nucleus contains two neutrons and a hydrogen nucleus contains no neutrons, then by assuming *all* the neutrons end up in ${}^4\text{He}$, we can then obtain the number density of ${}^4\text{He}$, $\mathcal{N}_{4\text{He}} = \frac{\mathcal{N}_n}{2}$. The Helium nucleus contains two protons and two neutrons hence $m_{4\text{He}} = 4m_p$. If Y_4 is the fraction of the total mass of particles in ${}^4\text{He}$ then

$$\begin{aligned} Y_4 &= \frac{m_{4\text{He}} \times \mathcal{N}_{4\text{He}}}{m_p \times \mathcal{N}_p + m_n \times \mathcal{N}_n} = \frac{4m_p \times \frac{\mathcal{N}_n}{2}}{m_p(\mathcal{N}_p + \mathcal{N}_n)} \\ &= \frac{2\mathcal{N}_n}{\mathcal{N}_p + \mathcal{N}_n} = \frac{2}{1 + \frac{\mathcal{N}_p}{\mathcal{N}_n}} \\ Y_4 &\simeq 0.22 \end{aligned}$$

In other words 22% of matter in the universe is in the form of Helium-4, with 78% in Hydrogen. More detailed calculations involve solving for the whole network of nuclear reactions and a careful analysis of the balance between the reaction rate and expansion rate of the universe. For an up to date review check the excellent review article written by B.D. Fields and Subir Sarkar in the particle physics data book at <http://pdg.lbl.gov/2011/reviews/rpp2011-rev-bbang-nucleosynthesis.pdf>. The best predictions for all the light elements as a fraction of the Hydrogen abundance to date based on that reference is :

$$Y_4 = 0.249 \pm 0.009 : \quad D/H = (2.78 \pm 0.29) \times 10^{-5} : \quad Li/H = (1.7 \pm 0.02_{-0}^{+1.1}) \times 10^{-10}.$$

The prediction of such a low abundance for Deuterium and Lithium is made all the more remarkable by the fact that it is confirmed by observation. Remember they span nine order of magnitude! This can be seen in Figure (2.4).

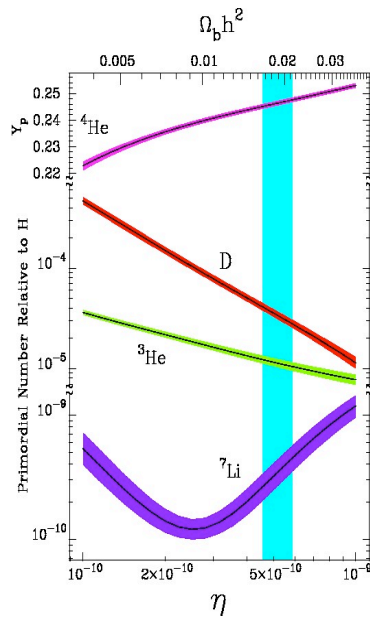


Figure 2.4: The predicted values on the relative abundances of Helium-4, Deuterium, Helium-3 and Lithium-7 as a function of the baryonic density $\Omega_b h^2$. Note how all four observed elemental abundances fit in with a narrow range of predictions for $\Omega_b h^2$ – a great success story of the HBB through its prediction of nucleosynthesis.

2.8 Recombination and decoupling

The universe passes through nucleosynthesis, past matter-radiation equality at $t \sim 3400$ years, $T_{\text{eq}} \sim 66,000K$ and $z \sim 24,000$. The next major cosmological event is reached when the universe has cooled to around $T \sim 1000K$ when it becomes possible for the ionised plasma to form neutral atoms. This is **recombination**. As the temperature drops to of order 1eV , photons remain tightly coupled to electrons via Compton scattering and electrons to protons through Coulomb scattering. There is very little Hydrogen even though the binding energy for neutral hydrogen is $\epsilon_0 = 13.6\text{eV}$. This is simply because there are so many high energy photons flying around, ionising any Hydrogen that may try and form. Now as long as the reaction $e^- + p \leftrightarrow H + \gamma$ is in equilibrium then we have

$$\frac{n_e n_p}{n_H} = \frac{n_e^{(0)} n_p^{(0)}}{n_H^{(0)}} \quad (2.68)$$

where $n_i^{(0)}$ is defined to be the species-dependent equilibrium number density given by Eqn. (2.8) or Eqn. (2.18) depending whether we are in the relativistic or non-relativistic regimes. This condition comes from the Boltzmann equation which tells us how we move out of equilibrium and which we state here without derivation (see Dodelson's book for a derivation in chapter 3)

$$a^{-3} \frac{d(n_e a^3)}{dt} = n_e^{(0)} n_p^{(0)} < \sigma v > \left(\frac{n_H}{n_H^{(0)}} - \frac{n_e n_p}{n_e^{(0)} n_p^{(0)}} \right) \quad (2.69)$$

where $< \sigma v >$ is the thermally average cross-section. Note this is similar to the Boltzmann equation (2.45) for the case of freeze-out discussed in section 2.5. It is important to realise that all of these processes involving the freeze out of particles, the fixed ratio of neutrons to protons or recombination and decoupling, **all involve the same basic physics**, namely solving Boltzmann equations for out of equilibrium phenomenon.

It follows that equilibrium is maintained when the terms inside the brackets vanish. Now neutrality of the universe ensures $n_e = n_p$, Defining the **free electron fraction**

$$X_e \equiv \frac{n_e}{n_e + n_H} = \frac{n_p}{n_p + n_H} \quad (2.70)$$

we see the denominator is the total number of hydrogen nuclei. Now rearranging Eqn. 2.70) we have

$$\frac{X_e}{1 - X_e} = \frac{n_p}{n_H} \simeq \exp(-(m_p - m_H)c^2/k_B T) \quad (2.71)$$

where we have used Eqn. (2.18) to deal with the equilibrium terms $n_e^{(0)}$, $n_p^{(0)}$ etc..., and we have ignored the small mass difference between m_p and m_H in the prefactor. Finally using

$$X_e = \frac{1}{n_e + n_H} \left(\frac{m_e k_B T}{2\pi} \right)^{3/2} \exp(-m_e/T) \quad (2.72)$$

we obtain the Saha equation

$$\frac{X_e^2}{1 - X_e} = \frac{1}{n_e + n_H} \left[\left(\frac{m_e k_B T}{2\pi} \right)^{3/2} \exp(-(m_e + m_p - m_H)c^2/k_B T) \right] \quad (2.73)$$

Now the argument of the exponential is just the binding energy for Hydrogen, $-\epsilon/k_B T$. Neglecting the small numbers of helium atoms, then the denominator $n_e + n_H = n_p + n_H$ is just the baryon density which is given by $n_b = \eta n_\gamma \sim 10^{-9} T^3$. Hence when the temperature is of order ϵ_0 i.e. 13eV the RHS of Eqn. (2.73) is of order $\frac{1}{n_b} \left(\frac{m_e T}{2\pi}\right)^{3/2} = 10^9 (m_e/T)^{3/2} \sim 10^{15}$. In other words it is huge which can only be accommodated if the denominator of the LHS nearly vanishes or $X_e \simeq 1$, implying all the hydrogen is ionised. It is only when the temperature has dropped well below ϵ_0 that significant recombination can take place. In fact as X_e falls it becomes more difficult to maintain equilibrium as the rate for recombination also falls. In order to solve for the free electron fraction accurately the Boltzmann equation (2.69) needs to be solved (remember $n_e = n_p$):

$$\begin{aligned} a^{-3} \frac{d(n_e a^3)}{dt} &= n_e^{(0)} n_p^{(0)} < \sigma v > \left(\frac{n_H}{n_H^{(0)}} - \frac{n_e^2}{n_e^{(0)} n_p^{(0)}} \right) \\ &= n_b < \sigma v > \left((1 - X_e) \left(\frac{m_e k_B T}{2\pi} \right)^{3/2} \exp(-\epsilon_0/k_B T) - X_e^2 n_b \right) \end{aligned} \quad (2.74)$$

where we have used the fact that the ratio $n_e^{(0)} n_p^{(0)} / n_H^{(0)}$ is equal to the term in square brackets in Eqn. (2.73). Now since $n_b a^3$ is a constant and using $n_e = X_e n_b$, we can replace n_e on the LHS of Eqn. (2.74), pull the n_b through the derivative to get

$$\frac{dX_e}{dt} = \left[(1 - X_e) \beta - X_e^2 n_b \alpha^{(2)} \right] \quad (2.75)$$

where the ionisation rate is denoted by

$$\beta \equiv < \sigma v > \left(\frac{m_e k_B T}{2\pi} \right)^{3/2} \exp(-\epsilon_0/k_B T) \quad (2.76)$$

and the recombination rate is

$$\alpha^{(2)} \equiv < \sigma v > . \quad (2.77)$$

The superscript (2) in $\alpha^{(2)}$ is because recombination to the ground (n=1) state is not relevant, they lead to the production of an ionising photon which immediately ionises the neutral atom, so the net effect of such a recombination is zero as no new neutral atoms are formed this way. In fact to proceed capture must be to an excited state of hydrogen, and the temperature dependent part of the rate for this is (without proof)

$$\alpha^{(2)} \propto \left(\frac{\epsilon_0}{k_B T} \right)^{1/2} \ln \left(\frac{\epsilon_0}{k_B T} \right) \quad (2.78)$$

In general the Saha approximation Eqn. (2.73) does a good job in predicting the the redshift of recombination but fails as the electron fraction drops and the system moves out of equilibrium. The full solution is obtained by solving Eqn. (2.75), and a typical solution is shown in figure (2.5). We see that recombination occurs suddenly at $z \sim 1000$ corresponding to $T \sim 1/4\text{eV}$. The Saha approximation holds in equilibrium and correctly identifies the redshift of recombination but not the detailed evolution of X_e . **Recombination** is directly related to the **decoupling** of photons from matter. Decoupling is in turn important as it affects the CMB anisotropies we observe. Decoupling occurs roughly when the rate for photons to Compton scatter off electrons becomes smaller than the expansion rate, i.e when $\frac{n_e \sigma_T}{H} < 1$. Let's work out when that occurs and show

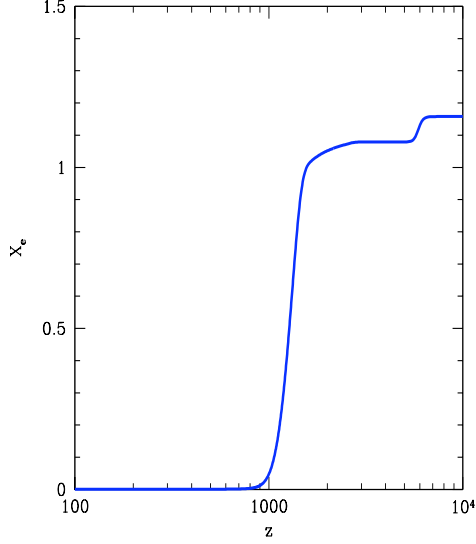


Figure 2.5: The evolution of the electron fraction X_e as a function of redshift z . Note how it drops abruptly around $z \sim 1000$ as the system moves out of equilibrium. Decoupling occurs during that period before recombination comes to an end.

that it occurs during recombination. The scattering rate $n_e \sigma_T$ can be written as $X_e n_b \sigma_T$, where $\sigma_T = 0.665 \times 10^{-24} \text{ cm}^2$ is the Thomson cross-section. Now since $\rho_b = \rho_{b0}(1+z)^3 = \Omega_{b0}\rho_c(1+z)^3$ and also $\rho_b = m_b n_b$ it follows that

$$n_b = \frac{\rho_c}{m_b} \Omega_{b0} (1+z)^3. \quad (2.79)$$

Hence inserting for ρ_c and $m_b = m_p$ we obtain

$$n_e \sigma_T = 7.477 \times 10^{-30} \text{ cm}^{-1} X_e \Omega_{b0} h^2 (1+z)^3 \quad (2.80)$$

It follows that

$$\frac{n_e \sigma_T}{H} = 0.0692 (1+z)^3 X_e \Omega_{b0} h \frac{H_0}{H} \quad (2.81)$$

where we have divided and multiplied by $H_0 = 3.26 h \times 10^{-11} \text{ sec}^{-1}$ and remembered to convert H_0^{-1} to a distance by multiplying through by c . The RHS depends on the Hubble rate which we get from the Friedmann equation. During this epoch we expect both radiation and matter to be important so in Eqn. (1.64) we have

$$H^2(z) = H_0^2 (\Omega_{m0}(1+z)^3 + \Omega_{r0}(1+z)^4). \quad (2.82)$$

or

$$\frac{H(z)}{H_0} = \Omega_{m0}^{1/2} (1+z)^{3/2} \left(1 + \frac{\Omega_{r0}(1+z)}{\Omega_{m0}} \right)^{1/2} \quad (2.83)$$

Now at equality we have $\rho_r = \rho_m$, hence $\Omega_{r0}(1+z_{\text{eq}}) = \Omega_{m0}$. It follows that

$$\frac{H(z)}{H_0} = \Omega_{m0}^{1/2} (1+z)^{3/2} \left(1 + \frac{(1+z)}{(1+z_{\text{eq}})} \right)^{1/2} \quad (2.84)$$

Inserting into Eqn. (2.81) and using ‘best-fit’ values for the baryon and matter densities, as well as $1 + z_{\text{eq}} = 24096\Omega_{m0}h^2$ we obtain

$$\frac{n_e\sigma_T}{H} = 113X_e \left(\frac{\Omega_{b0}h^2}{0.02}\right) \left(\frac{0.15}{\Omega_{m0}h^2}\right)^{1/2} \left(\frac{1+z}{1000}\right)^{3/2} \left[1 + \frac{(1+z)}{3600} \frac{0.15}{\Omega_{m0}h^2}\right]^{-1/2}. \quad (2.85)$$

Given we have used the best fit values, what that means is we can consider the brackets (..) to be of order unity. Therefore when the free electron fraction X_e drops below $\sim 10^{-2}$, photons decouple. From the figure it is clear that X_e drops very quickly from unity to 10^{-3} , therefore decoupling takes place during recombination. That is the formation of the CMB background !

3 Large Scale Structure formation

So far we have assumed that on the average the Universe is homogeneous and isotropic. This assumption, although very powerful in its simplicity, is also of limited validity. To be able to study things like the formation of structure such as galaxies, or the fluctuations of the Cosmic Microwave Background radiation, or the theory of initial conditions such as inflation, or practically anything that can be used to tie our cosmological theory with observations we need to go beyond the homogeneous and isotropic Universe. To study all of these things, in principle, one should be using the full non-linear Einstein equations. However, this is impossible in practice. To make things easier, we resort to a very simplifying (but not too much) assumption. That the Universe on large scales is ALMOST homogeneous and isotropic, that is, all relevant variables such as the metric, the energy density, etc, depart only slightly from their FRW counterparts. The tool we need to use in this case is called ”Cosmological Perturbation Theory”.

Cosmological perturbation theory is relativistic and so before introducing it we shall consider a simpler case: Newtonian perturbation theory. Although it sounds like we are cheating (and we are) for no reason, this is not completely true. One of the main problems that Modern Cosmology tries to describe is the formation of structure under gravitational instability. In effect we are trying to answer the following questions. Can structure in an expanding universe form from gravitational collapse of small fluctuations? If the answer is yes, then does the observed amount of structure agree with our theory? For the most part, these questions only deal with non-relativistic matter. Furthermore, most observable structure, for instance galaxies, is on scales much smaller than the cosmological horizon. Because of these two reasons, to describe structure formation in cosmology it is sufficient to use Newtonian perturbation theory. We shall intruduce the fully relativistic treatment afterwards.

3.1 Dimensional analysis

Quite often choosing the right units for a problem can reveal much about the underlying physics and the approach we need to take in order to solve it. In particle physics and cosmology we often use ”natural units”. In this system of units one sets the fundamental constants $c = \hbar = k_B = 1$. In a way, this will make everyone’s life easier as we won’t have to carry around factors of c , k_B or $\hbar$. This choice of units has a very important consequence. It leaves only one fundamental dimension free: that of energy. In particular, the other dimensions are either equal or inversely proportional to energy so that

$$[Energy] = [Mass] = [Temperature] = [Length]^{-1} = [Time]^{-1}. \quad (3.1)$$

A quite conventional unit of energy is the Giga-electron-Volt, GeV . This is equal to $10^9 eV$, i.e. 1 billion electron-Volts. An electron-Volt is equal to $1.602 \times 10^{-19} J$. To be able to perform any kind of unit conversion we can use the following table:

Energy:	$1 GeV$	$=$	$1.602 \times 10^{-10} J = 1.5637 \times 10^{38} Mpc^{-1}$
Mass:	$1 Kg$	$=$	$5.6095 \times 10^{26} GeV = 8.7714 \times 10^{64} Mpc^{-1}$
Temperature:	$1 K$	$=$	$8.61698 \times 10^{-14} GeV = 1.34744 \times 10^{25} Mpc^{-1}$
Length:	$1 m$	$=$	$5.0677 \times 10^{15} GeV^{-1} = 3.2409 \times 10^{-23} Mpc$
Time:	$1 s$	$=$	$1.51925 \times 10^{24} GeV^{-1} = 9.7160 \times 10^{-15} Mpc = 2.997925 \times 10^8 m$

For example, consider the Hubble constant $H_0 = 100h \text{ km/s/Mpc}$ (where h is a number of order 1). Now H_0 has units of $[Length]^{-1}$ or equivalently $[Mass]$. Let us answer the following questions

- What is H_0 in Mpc^{-1} ? We have

$$\begin{aligned} H_0 &= 100h \times 10^3 \times 3.2409 \times 10^{-23} / (9.7160 \times 10^{-15}) \times Mpc^{-1} \\ &= 3.336h \times 10^{-4} Mpc^{-1} \end{aligned}$$

- What is H_0 in GeV ? We find $H_0 = 2.1334h \times 10^{-42} GeV$

You can then answer quickly things like, "What is the size of the observable Universe?". Well, it should be about $1/H_0$ (more-or-less) so it is $\sim 1/3 \times 10^4 Mpc \sim 3000 Mpc$. This is in fact very close to the correct exact answer! Unit conversions need practice but once you master it you will realize how important it is to have this power!

So from now on no more c , k_B or $\hbar$. They are banned!

3.2 Elements of Newtonian fluid dynamics

3.2.1 Newtonian fluids

Strictly speaking, we should be trying to describe how all the particles in the late Universe, propagate, collide and collect together into clumped structures. As you can imagine this is practically impossible as the number of individual particles is humongous. Instead we shall resort to an approximation, but a very good one in this case: we shall describe baryons and dark matter as two non-interacting (except via gravity) effective fluids. For our description to hold, we shall assume that any small volume element in the fluid is always large enough that it contains a large number of particles. In that case, we may talk of the energy density of a fluid $\rho(t, \vec{x})$ at point O (with coordinates t and $\vec{x}$) as representing the average mass density of all the particles in a small volume δV around point O . For each separate fluid the particles are identical. If the mass of one particle is m and we have N particles within δV then the energy density inside V is $\rho_{\delta V} = \frac{Nm}{\delta V}$. We will assume that the energy density of the fluid at point O is

$$\rho = \lim_{\delta V \rightarrow 0} \frac{Nm}{\delta V} \quad (3.2)$$

By moving the point O and the volume δV at every other point in space we can construct the fluid energy density $\rho(t, \vec{x})$ (see figure 3.1). Of course as the volume region shrinks to zero, the number of particles will also decrease and the assumption here is that their ratio will have a well-defined

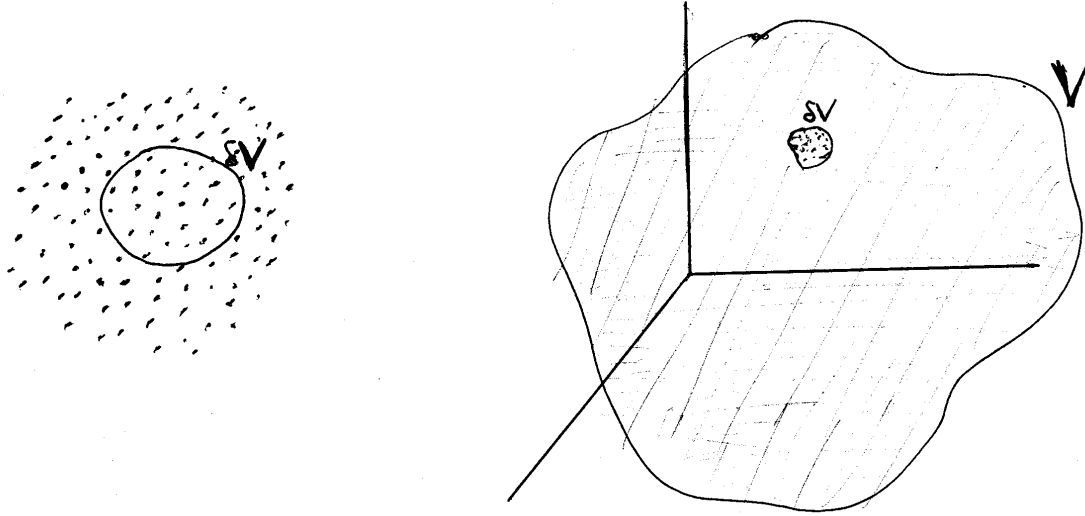


Figure 3.1: Collection of particles as an effective fluid. Left: large numbers of particles are collected within small volume elements δV . Right: The small volume element δV is much smaller than the total volume of any given fluid region fluid V .

meaning in this limit. We will not be concerned whether this procedure can be done or whether it makes sense, and we shall simply assume that it is possible. We proceed in a similar fashion with other possible variables, for instance the velocity $\vec{v}(t, \vec{x})$ of a fluid packet and the pressure $P(t, \vec{x})$. Each separate fluid is thus described by mass density $\rho(t, \vec{x})$, velocity $\vec{v}(t, \vec{x})$ and pressure $P(t, \vec{x})$.

3.2.2 Newtonian gravity for continuous media

Newton's law of gravity is usually considered to act between individual, discrete, point masses: $\vec{F} = -G \frac{m_1 m_2}{r^2} \hat{r}$ where r is the distance between the two masses m_1 and m_2 and the minus sign indicates that the force is in the opposite direction to the unit vector $\hat{r}$, i.e. it is an attractive force. Our first task is to reformulate this law for a continuous distribution of mass such as the fluid we described earlier. We shall do so by breaking the fluid into small regions of volume δV and mass δm and considering the force of gravity between all such regions.

Consider a fluid region as in figure 3.2. We first focus on a small part of the fluid at position $\vec{r}$ with mass $\delta m(\vec{r})$ and calculate the total gravitational force on δm due to the rest of the fluid. To do so, consider first the force due to another small part at position $\vec{r}_i$ with mass $\delta m_i(\vec{r}_i)$. Newton's law of gravity tells us that the force on δm due to δm_i is

$$\delta^2 \vec{F}_i = -G \frac{\delta m \delta m_i}{|\vec{r} - \vec{r}_i|^2} (\hat{r} - \hat{r}_i) \quad (3.3)$$

The δ^2 in front of $\vec{F}$ is to remind us that this is the force due to two infinitesimal masses. To calculate the total force acting on δm we must add-up the forces from all possible small masses

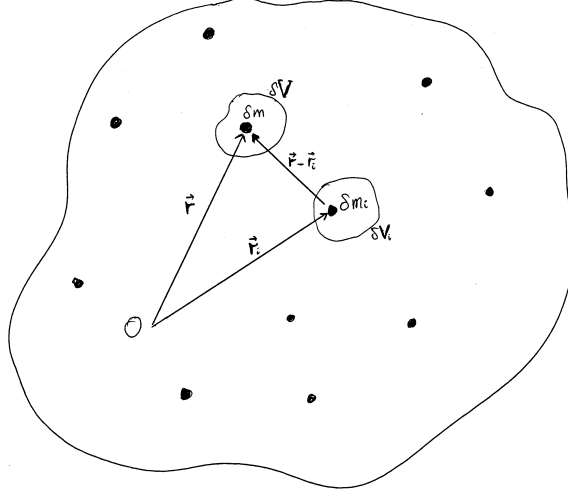


Figure 3.2: A small region of the fluid which is build-up from small masses δm_i . The position vector from the origin O to δm is $\vec{r}$ while the position vector from O to δm_i is $\vec{r}_i$. The vector from δm_i to δm is $\vec{r} - \vec{r}_i$. Each small mass δm is surrounded by a small volume δV .

δm_i . This gives us

$$\delta \vec{F} = \sum_i \delta^2 \vec{F}_i \quad (3.4)$$

$$= -G\delta m \sum_i \frac{\delta m_i}{|\vec{r} - \vec{r}_i|^2} (\hat{r} - \hat{r}_i) \quad (3.5)$$

Now we use the mass density of the fluid ρ to exchange a small mass δm with a small volume δV : $\delta m = \rho\delta V$. Likewise, we define the *force density* $\vec{f}$ as the force per unit volume: $\vec{f} = \frac{\delta \vec{F}}{\delta V}$. Our force law becomes

$$\delta \vec{F} = \vec{f} \delta V = -G\rho(\vec{r}) \delta V \sum_i \rho(\vec{r}_i) \frac{\delta V_i}{|\vec{r} - \vec{r}_i|^2} (\hat{r} - \hat{r}_i) \quad (3.6)$$

and after canceling δV we get

$$\vec{f} = -G\rho(\vec{r}) \sum_i \rho(\vec{r}_i) \frac{\delta V_i}{|\vec{r} - \vec{r}_i|^2} (\hat{r} - \hat{r}_i) \quad (3.7)$$

The final step is to take make δV infinitesimal, i.e. $\delta V \rightarrow dV = d^3r$, and convert the sum into an integral. We also let $\vec{r}_i \rightarrow \vec{r}'$. The final equation for the force density at position $\vec{r}$ is

$$\vec{f}(\vec{r}) = -G\rho(\vec{r}) \int d^3r' \rho(\vec{r}') \frac{(\hat{r} - \hat{r}')}{|\vec{r} - \vec{r}'|^2} \quad (3.8)$$

where the integral is carried out over all parts of the fluid. Equation (3.8) is the infinitesimal analogue of $\vec{F} = m\vec{a}$. In fact we can read-off the gravitational acceleration at position $\vec{r}$:

$$\vec{g}(\vec{r}) = -G \int d^3r' \rho(\vec{r}') \frac{(\hat{r} - \hat{r}')}{|\vec{r} - \vec{r}'|^2} \quad (3.9)$$

where we have changed notation from $\vec{a}$ to $\vec{g}$ to distinguish the gravitational acceleration from any other kind of acceleration.

3.2.3 The Newtonian potential Φ and the Poisson equation

We will now introduce a new concept: the gravitational potential $\Phi(\vec{r})$ at position $\vec{r}$. To do so, consider the (vectorial) term $\frac{\hat{r}-\hat{r}'}{|\vec{r}-\vec{r}'|^2}$ in (3.9). Since $\hat{r}$ is a unit vector, then this term is of the form $\sim 1/x^2$ which equals $-\frac{d}{dx}(\frac{1}{x})$. So it is suggestive that we may obtain our vectorial term via some kind of differentiation. There is one vector differential operator which may be of help: the *gradient operator* $\vec{\nabla}$. To remind you the $\vec{\nabla}$ operator in cartesian coordinates (also called "del" or "grad"), converts a scalar into a vector like this : $\vec{\nabla}f = (\frac{\partial f}{\partial x}, \frac{\partial f}{\partial y}, \frac{\partial f}{\partial z})$. But the grad operator is more general and in fact is defined for any coordinate system: $\vec{\nabla}f \rightarrow \vec{\nabla}_i f = \frac{\partial f}{\partial x^i}$.

Now back to our vectorial term $\frac{\hat{r}-\hat{r}'}{|\vec{r}-\vec{r}'|^2}$. We want to obtain this term via the grad operator acting on some scalar function. The only such scalar function we have available is $|\vec{r}-\vec{r}'|$ so let's apply the grad operator to $|\vec{r}-\vec{r}'|$. Our grad operator may treat either $\vec{r}$ or $\vec{r}'$ as the independent variable (but not both). Let us choose $\vec{r}$ (because we will eventually integrate over $\vec{r}'$). Let's do the calculation in cartesian coordinates for which $\vec{r} = (x, y, z)$ and $\vec{r}' = (x', y', z')$. Then $|\vec{r}-\vec{r}'| = \sqrt{(x-x')^2 + (y-y')^2 + (z-z')^2}$. The grad operator gives us

$$\vec{\nabla}|\vec{r}-\vec{r}'| = \begin{pmatrix} \frac{\partial}{\partial x} \\ \frac{\partial}{\partial y} \\ \frac{\partial}{\partial z} \end{pmatrix} \sqrt{(x-x')^2 + (y-y')^2 + (z-z')^2} \quad (3.10)$$

$$= \frac{1}{2\sqrt{(x-x')^2 + (y-y')^2 + (z-z')^2}} \begin{pmatrix} 2(x-x') \\ 2(y-y') \\ 2(z-z') \end{pmatrix} \quad (3.11)$$

$$= \frac{\vec{r}-\vec{r}'}{|\vec{r}-\vec{r}'|^2} \quad (3.12)$$

$$= \hat{r} - \hat{r}' \quad (3.13)$$

So we have just found that $\hat{r} - \hat{r}' = \vec{\nabla}|\vec{r}-\vec{r}'|$. Back to our vectorial term we get

$$\frac{\hat{r} - \hat{r}'}{|\vec{r} - \vec{r}'|^2} = \frac{\vec{\nabla}|\vec{r} - \vec{r}'|}{|\vec{r} - \vec{r}'|^2} \quad (3.14)$$

but remember that $\frac{d}{dx}(1/x) = -1/x^2$ so we get

$$\frac{\hat{r} - \hat{r}'}{|\vec{r} - \vec{r}'|^2} = -\vec{\nabla} \left(\frac{1}{|\vec{r} - \vec{r}'|} \right) \quad (3.15)$$

We now return back to our acceleration equation (3.9) and replacing the answer for our vectorial term we get

$$\vec{g}(\vec{r}) = G \int d^3r' \rho(\vec{r}') \vec{\nabla} \left(\frac{1}{|\vec{r} - \vec{r}'|} \right) \quad (3.16)$$

$$= G \vec{\nabla} \int d^3r' \frac{\rho(\vec{r}')}{|\vec{r} - \vec{r}'|} \quad (3.17)$$

where in the 2nd line we pulled $\vec{\nabla}$ out of the integral ($\vec{\nabla}$ acts on $\vec{r}$ and not on $\vec{r}'$). What we have gained via this long process is that now the integral is over a *scalar* rather than a vector. We call this scalar the Newtonian gravitational potential $\Phi(\vec{r})$. To be more specific, we re-insert the minus signs to signify that the acceleration is in the opposite direction to the vector $\vec{\nabla}$ (it is attractive) and define

$$\vec{g}(\vec{r}) = -\vec{\nabla}\Phi(\vec{r}) \quad (3.18)$$

$$\Phi(\vec{r}) = -G \int d^3r' \frac{\rho(\vec{r}')}{|\vec{r} - \vec{r}'|} \quad (3.19)$$

Recall your Electromagnetism? You were dealing with electric fields $\vec{E}$ which were obtained from an electric potential V by $\vec{E} = -\vec{\nabla}V$. In our case the analogue of the electric field is the gravitational acceleration $\vec{g}$ and the analogue of the electric potential is the Newtonian potential Φ . In fact (3.19) is the analogue of the integral formulation of Gauss law. But this means that just as there is a differential formulation of Gauss law, there should also be a differential formulation of (3.19). And yes there is! It is called the Poisson equation.

Recall that the differential formulation of Gauss law was related to the charge density σ by $\text{div}\vec{E} = \vec{\nabla} \cdot \vec{E} = \frac{\sigma}{\epsilon_0}$. So we should expect an analogous thing here. This means we need to calculate $\vec{\nabla} \cdot \vec{g} = -\nabla^2\Phi$. Now $\nabla^2\Phi$ gives

$$\nabla^2\Phi(\vec{r}) = -G \int d^3r' \rho(\vec{r}') \nabla^2 \left(\frac{1}{|\vec{r} - \vec{r}'|} \right) \quad (3.20)$$

Without proof, the term $\nabla^2 \left(\frac{1}{|\vec{r} - \vec{r}'|} \right)$ is equal to $-4\pi\delta^{(3)}(\vec{r} - \vec{r}')$ where $\delta^{(3)}$ is the three-dimensional Dirac delta-function. We may revisit the proof in a non-assessed problem sheet further down the course for those interested. Hence we find the Poisson equation

$$\nabla^2\Phi = 4\pi G\rho \quad (3.21)$$

The Poisson equation is a 2nd order linear differential equation which tells us how the Newtonian potential Φ responds to the presence of mass density ρ . Notice that in Newtonian gravity, the temporal response of Φ to any temporal changes in ρ is instantaneous. As we shall see later, this will change in General Relativity.

3.2.4 The continuity and Euler equations

So now we have at our disposal the fluid variables ρ , $\vec{v}$ and P . We also have the Poisson equation which relates the Newtonian potential Φ to the density ρ . But what do we do with all these variables? What we have left to do is to find how ρ and $\vec{v}$ respond to changes, i.e. how do they evolve dynamically.

The first equation we will derive is the *continuity equation*. To derive it consider the total mass within a region V . It is

$$M(t) = \int_V dV \rho(t, \vec{x}) \quad (3.22)$$

This mass is time-dependent because the mass density is time-dependent. The time-dependence of the mass density comes about because the particles inside V may be moving around so that their number changes from place to place and this changes ρ . But as long as there is no flow of fluid into

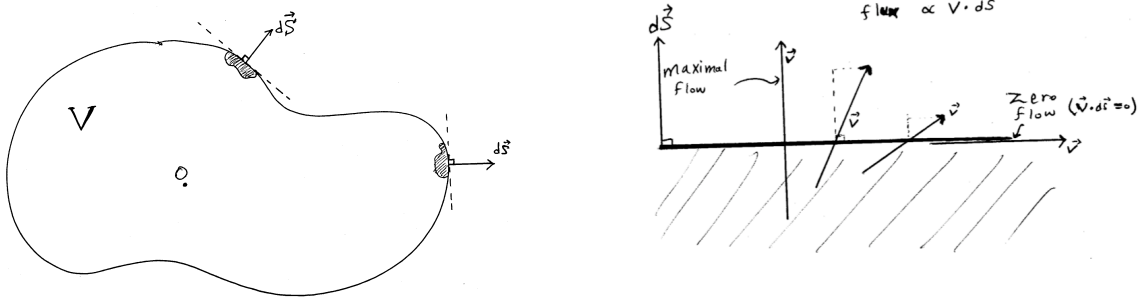


Figure 3.3: Left: A volume of fluid V surrounded by surface $\vec{S}$. The vector corresponding to an infinitesimal surface area is $d\vec{S}$ and is always perpendicular to the surface. Right: Flow through a surface with normal $d\vec{S}$. Maximal flow occurs when the velocity vector $\vec{v}$ is parallel to $d\vec{S}$. Zero flow is when they are perpendicular. In general the flux is proportional to $\vec{v} \cdot d\vec{S}$.

or out from the region V then the total mass M would be constant. But the particles may in fact move in or out of the region V and in that case the total mass within V will change with time due to the fluid flowing into or out from the region V . The rate of mass *decrease* per unit time is of course

$$-\frac{dM(t)}{dt} = - \int_V dV \frac{\partial}{\partial t} \rho(t, \vec{x}) \quad (3.23)$$

Now the rate of mass decrease should equal to the flux through the surface which bounds the volume V . Let the vector normal to the surface be $\vec{S}$ (see figure 3.3). Then the flux should be maximal if the particles flow at right angles to the surface or in other words in a direction parallel to $\vec{S}$. Likewise, the flux should be zero if the particles are moving tangentially to the surface or in other words in a direction perpendicular to $\vec{S}$. This means that for a small surface element $\delta\vec{S}$, the flux should be proportional to $\vec{v} \cdot \delta\vec{S}$. Now $\vec{v} \cdot \delta\vec{S}$ has units of $[Length]^2$ while flux, which is mass per unit time, should have units $[Mass][Length]^{-1}$ (remember time is length according to Einstein). Meanwhile ρ has units $[Mass][Length]^{-3}$, hence $\rho\vec{v} \cdot \delta\vec{S}$ has the required units of flux: $[Mass][Length]^{-1}$. Thus the differential flux through a differential surface element $d\vec{S}$ is $\rho\vec{v} \cdot d\vec{S}$, and integrating over the whole surface we get the total flux as

$$\int_S \rho \vec{v} \cdot d\vec{S} \quad (3.24)$$

Clearly then we must equate (3.23) to (3.24) to get

$$\int_V dV \frac{\partial}{\partial t} \rho(t, \vec{x}) = - \int_S \rho \vec{v} \cdot d\vec{S} \quad (3.25)$$

But we still have that annoying $d\vec{S}$. Fortunately, Stokes comes to rescue and by using his theorem we can convert the surface integral to a volume integral:

$$\int_S \rho \vec{v} \cdot d\vec{S} = \int_V \text{div}(\rho \vec{v}) dV = \int_V \vec{\nabla} \cdot (\rho \vec{v}) dV \quad (3.26)$$

Hence we find

$$\frac{\partial}{\partial t} \int_V \rho(t, \vec{x}) dV = - \int_V \vec{\nabla} \cdot (\rho \vec{v}) dV \quad (3.27)$$

but this must be valid for any volume V and therefore we get

$$\frac{\partial \rho}{\partial t} + \vec{\nabla} \cdot (\rho \vec{v}) = 0 \quad (3.28)$$

which is equivalent to

$$\frac{\partial \rho}{\partial t} + \rho \vec{\nabla} \cdot \vec{v} + \vec{v} \cdot \vec{\nabla} \rho = 0 \quad (3.29)$$

The above equation is known as the continuity equation. The continuity equation tells us how temporal changes in the mass density ρ happen due to the motions of the particles. It really is an equation for conservation of mass.

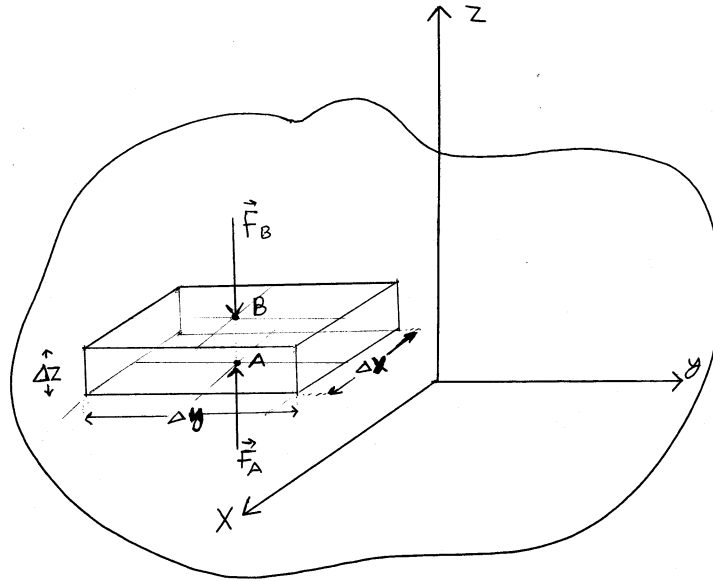


Figure 3.4: A slab of fluid with sides Δx , Δy and Δz being acted by pressure forces F_A and F_B at points A and B along the vertical (z) axis.

We clearly need one more equation which relates to temporal changes of the velocity, i.e. we need an equation for the acceleration $\vec{a}$ of fluid elements. The particles which comprise the fluid obey Newton's laws of motion, in particular the one that we are interested in is Newton II: $\vec{F} = m\vec{a}$. We have already found such an equation when we were discussing gravity. In particular we introduced the concept of the force density $\vec{f}$ according to which Newton II takes the form

$$\rho \vec{a} = \vec{f} \quad (3.30)$$

This is the equation we need for $\vec{a}$, but we still need to provide for the possible forces acting on the fluid. There can be various kinds of forces but here we are interested in two in particular. The first is the hydrodynamic force which arises because of pressure differences from different parts of

the fluid. The second is gravity and as we have already seen we have that $\vec{f}_{grav} = \rho\vec{g} = -\rho\vec{\nabla}\Phi$ (see (3.18)).

So what is the force due to pressure difference? Remember that pressure is force per unit area. Let us have a look at figure 3.4. Suppose we consider a small slab inside the fluid of size $\Delta x \times \Delta y \times \Delta z$. Consider the hydrodynamic pressure acting at points A and B and relate this pressure to the respective force. We have that $F_A = P_A \Delta x \Delta y$ and $F_B = P_B \Delta x \Delta y$. Hence the total force acting in the vertical direction is $\Delta F = F_A - F_B = (P_A - P_B) \Delta x \Delta y$ (remember that F_A is positive because it points to the positive z axis while F_B is negative because it points to the negative z axis). We now use a trick: Taylor expansion. If B is sufficiently close to A , i.e. Δz is small, then we can take P_B and Taylor expand it around point A . We find $P_B = P_A + \frac{\partial P}{\partial z} \Delta z$. Hence the total force is

$$\Delta F = -\frac{\partial P}{\partial z} \Delta x \Delta y \Delta z \quad (3.31)$$

$$= -\frac{\partial P}{\partial z} \Delta V \quad (3.32)$$

where we have collected $\Delta x \Delta y \Delta z = \Delta V$ as the volume of the slab. We finally obtain the force density as $\vec{f} = \frac{\Delta F}{\Delta V}$ and at the same time generalize the pressure difference to any direction (not just the vertical direction). For instance the pressure difference along the x -direction is $\frac{\partial P}{\partial x} \Delta x$ and along the y -direction is $\frac{\partial P}{\partial y} \Delta y$, hence the pressure difference along any direction is $(\frac{\partial P}{\partial x}, \frac{\partial P}{\partial y}, \frac{\partial P}{\partial z})$, i.e. it is $\vec{\nabla}P$: a vector. Thus putting things together, the force density due to hydrodynamic pressure is

$$\vec{f}_{hydro} = -\vec{\nabla}P \quad (3.33)$$

We now sum up our forces $\vec{f} = \vec{f}_{hydro} + \vec{f}_{grav}$ and into Newton-II to get

$$\rho\vec{a} = -\vec{\nabla}P - \rho\vec{\nabla}\Phi \quad (3.34)$$

We are almost done but not completely done! We need to relate the acceleration $\vec{a}$ to the velocity $\vec{v}$. Clearly

$$\vec{a} = \frac{d\vec{v}}{dt} \quad (3.35)$$

but remember that the velocity may change from point to point, i.e. it is $\vec{v}(t, \vec{x})$ so we have to be careful about what we mean by the *total derivative* $\frac{d}{dt}$. Consider a small variation of a variable $A(t, x)$ (which may be a scalar or a vector). We can express this variation in terms of variations of the time t and position (x, y, z) . We have

$$\delta A = \frac{\partial A}{\partial t} \delta t + \frac{\partial A}{\partial x} \delta x + \frac{\partial A}{\partial y} \delta y + \frac{\partial A}{\partial z} \delta z \quad (3.36)$$

Dividing by δt and taking the limit $\delta t \rightarrow 0$ we obtain the total time derivative

$$\frac{dA}{dt} = \lim_{\delta t \rightarrow 0} \frac{\delta A}{\delta t} = \frac{\partial A}{\partial t} + \frac{dx}{dt} \frac{\partial A}{\partial x} + \frac{dy}{dt} \frac{\partial A}{\partial y} + \frac{dz}{dt} \frac{\partial A}{\partial z} \quad (3.37)$$

But notice that the last three terms are a dot product of two vectors: $\vec{v} = (\frac{dx}{dt}, \frac{dy}{dt}, \frac{dz}{dt})$ and $\vec{\nabla}A = (\frac{\partial A}{\partial x}, \frac{\partial A}{\partial y}, \frac{\partial A}{\partial z})$. Hence we may write

$$\frac{dA}{dt} = \frac{\partial A}{\partial t} + \vec{v} \cdot \vec{\nabla}A \quad (3.38)$$

$\frac{\partial \rho}{\partial t} + \vec{v} \cdot \vec{\nabla} \rho = -\rho \vec{\nabla} \cdot \vec{v}$	Continuity equation
$\rho \left(\frac{\partial \vec{v}}{\partial t} + \vec{v} \cdot \vec{\nabla} \vec{v} \right) = -\vec{\nabla} P - \rho \vec{\nabla} \Phi$	Euler equation
$\nabla^2 \Phi = 4\pi G \rho$	Poisson equation
$\vec{\nabla} P = c_s^2 \vec{\nabla} \rho$	equation of state

Figure 3.5: The system of equations that describe Newtonian fluids

The operator

$$\frac{d}{dt} = \frac{\partial}{\partial t} + \vec{v} \cdot \vec{\nabla} \quad (3.39)$$

goes by many names: *total derivative*, *substantial derivative*, *material derivative*, *convective derivative* and possibly many others. It may act on any scalar or vector.

We now use it into our acceleration equation (3.34) to get our 2nd fluid equation:

$$\rho \left(\frac{\partial \vec{v}}{\partial t} + \vec{v} \cdot \vec{\nabla} \vec{v} \right) = -\vec{\nabla} P - \rho \vec{\nabla} \Phi. \quad (3.40)$$

The above equation is called the Euler equation. It tells us how changes in the velocity $\vec{v}$ come about due to pressure gradients and gravitational potentials in the fluid. It is the equation which describes motions in the fluid.

But now let's count variables. We have ρ , $\vec{v}$, Φ and P : 4 variables. But we only have three equations which means that we can't solve the continuity and Euler equations without further assumptions or equations. One such further assumption is to specify an equation of state. We define a *speed of sound* c_s so that

$$\vec{\nabla} P = c_s^2 \vec{\nabla} \rho \quad (3.41)$$

The speed of sound encapsulates properties of the fluid under studied and is determined by the microphysics involved. It is a given function that may depend on the other variables of the fluid the fluid, e.g. $c_s = c_s(\rho, \vec{v}, \Phi)$. However, in many cases it is assumed to be a constant. A summary of all the equations that are used to describe Newtonian fluids is shown in figure 3.5.

3.2.5 Evolution of small fluctuations in a Newtonian fluid

Having established the formalism for describing fluids in Newtonian theory we are ready for our first application. We want to investigate how small initial fluctuations behave under the pull of gravity and how they may be counter-balanced by internal pressure. This was done first by Jeans in 1902 when he considered the problem of formation of gaseous nebulae³.

Since we are to study fluctuations, we must specify a background for which our variables will be fluctuating about. Our background variables will be denoted by a "bar" over them. They are: $\bar{\rho}$, $\vec{\bar{v}}$ and $\bar{\Phi}$. To solve the background equations we further assume that on the average the fluid is homogeneous (which means we can drop the $\vec{x}$ dependence for the background variables) and at rest. The second assumption means that $\vec{\bar{v}} = 0$. Hence, the continuity equation gives $\dot{\bar{\rho}} = 0$ so $\bar{\rho}$

³J.H. Jeans, *The stability of spherical nebulae*, Phil. Trans. Roy. Soc. (London) **199**, 1 (1902).

is constant both in space and in time. Hence, the Newtonian potential can at best be a constant which we may set to zero (as the zero point of the potential is arbitrary). Finally, as you may easily check, the Euler equation (3.40) is identically satisfied and gives nothing new. Hence, the only non-vanishing background variable is a constant mass density $\bar{\rho}$.

There is a source of worry, however, which is that the Poisson equation is inconsistent. The reason is that $\nabla^2\Phi = 0$ while $\rho \neq 0$. This is known as the "Jeans swindle". We shall not concern ourselves with this here but if you are interested you may read <http://xxx.lanl.gov/abs/astro-ph/9910247>.

Let us now proceed to fluctuations. We define the total mass density by Taylor expanding $\rho(t, \vec{x})$ to linear order as

$$\rho(t, \vec{x}) = \bar{\rho} + \delta\rho(t, \vec{x}) \quad (3.42)$$

where now the fluctuation $\delta\rho$ depends on time and space. We will find it useful to further define the density *contrast* δ as

$$\rho = \bar{\rho}(1 + \delta) \quad (3.43)$$

so that $\delta = \frac{\delta\rho}{\bar{\rho}}$. Since the variables $\vec{v}$ and Φ vanish for the background we shall assume that they are already fluctuations and so we shall not use a "delta" in front of them.

Consider first the continuity equation. We shall consider each term separately and then add them together:

$$\frac{\partial\rho}{\partial t} = \frac{\partial\delta\rho}{\partial t} = \bar{\rho}\dot{\delta} \quad (3.44)$$

$$\vec{v} \cdot \vec{\nabla}\rho = \vec{v} \cdot \vec{\nabla}\delta\rho = 0 \quad (\text{vanishes because this is a higher than linear order term}) \quad (3.45)$$

$$\rho\vec{\nabla} \cdot \vec{v} = \bar{\rho}\vec{\nabla} \cdot \vec{v} \quad (3.46)$$

So, after cancelling $\bar{\rho}$ the continuity equation reduces to

$$\dot{\delta} + \vec{\nabla} \cdot \vec{v} = 0 \quad (3.47)$$

Now to the Euler equation. Keeping terms up-to linear order we find that it reduces to

$$\frac{\partial\vec{v}}{\partial t} = -\frac{1}{\bar{\rho}}\vec{\nabla}P - \vec{\nabla}\Phi \quad (3.48)$$

and after invoking the speed of sound we get the perturbed Euler equation

$$\frac{\partial\vec{v}}{\partial t} = -c_s^2\vec{\nabla}\delta - \vec{\nabla}\Phi \quad (3.49)$$

We now need to eliminate the potential using the Poisson equation. First, we hit the perturbed Euler equation (3.49) with $\vec{\nabla}$ to get

$$\frac{\partial}{\partial t} [\vec{\nabla} \cdot \vec{v}] = -c_s^2\nabla^2\delta - \nabla^2\Phi \quad (3.50)$$

$$= -c_s^2\nabla^2\delta - 4\pi G\bar{\rho}\delta \quad (3.51)$$

But then we can also eliminate $\vec{\nabla} \cdot \vec{v}$ using (3.47) and the final result is a 2nd order linear partial differential equation for δ :

$$\ddot{\delta} - c_s^2\nabla^2\delta - 4\pi G\bar{\rho}\delta = 0 \quad (3.52)$$

So let us now proceed to solve this equation. To do so we shall use Fourier transforms.

3.3 Math recap: Fourier transforms

We shall be using Fourier transforms throughout the course as they usually simplify hard problems. One such problem is the conversion of linear partial differential equations into ordinary differential equations. Most probably you already know Fourier transforms but let's briefly recap a few things about them that we will need for this course. See also handout 1 for more discussion.

3.3.1 Fourier transforms in 3 dimensions

A function $f(x)$ may be expanded in a Fourier integral as

$$f(x) = \int_{-\infty}^{\infty} \frac{dk}{2\pi} e^{ikx} \tilde{f}(k) \quad (3.53)$$

where $\tilde{f}(k)$ is called the Fourier-transformed function of $f(x)$. The above relation is invertible and we may write

$$\tilde{f}(k) = \int_{-\infty}^{\infty} dx e^{-ikx} f(x) \quad (3.54)$$

The two functions $f(x)$ and $\tilde{f}(k)$ form a Fourier transform pair. The factor 2π in the denominator of the integration measure in (3.53) is purely conventional. Other conventions include the "symmetric" convention where we put a $\sqrt{2\pi}$ in the denominator of both (3.53) and (3.54) and the "angular" convention which is obtained by defining a new variable $q = \frac{k}{2\pi}$ so that the 2π factor appears in the two exponentials instead.

Let us now pass to three dimensions. Fourier transforms in three dimensions are very similar to the one-dimensional ones with very minor modifications. Firstly the functions depend on position vectors $\vec{x}$ in real space and $\vec{k} = (k_x, k_y, k_z)$ in Fourier space. Secondly the integration measure in Fourier space obtains a factor $(2\pi)^3$ in the denominator. Thus the Fourier transforms corresponding to the Fourier pair $f(\vec{x})$ and $\tilde{f}(\vec{k})$ are

$$f(\vec{x}) = \int \frac{d^3k}{(2\pi)^3} e^{i\vec{k}\cdot\vec{x}} \tilde{f}(\vec{k}) \quad (3.55)$$

and

$$\tilde{f}(\vec{k}) = \int d^3x e^{-i\vec{k}\cdot\vec{x}} f(\vec{x}) \quad (3.56)$$

Inside the Fourier transform the two vectors $\vec{x}$ and $\vec{k}$ have reciprocal role. In particular small k corresponds to large $|\vec{x}|$, i.e. large scales, while large k corresponds to small $|\vec{x}|$, i.e. small scales.

3.3.2 The three-dimensional Dirac δ -function

Let's use (3.56) into (3.55). We find that

$$f(\vec{x}) = \int d^3y \int \frac{d^3k}{(2\pi)^3} e^{i\vec{k}\cdot(\vec{x}-\vec{y})} \tilde{f}(\vec{k}) f(\vec{y}) \quad (3.57)$$

But this means that the integral over $\vec{k}$ must result to the Dirac δ -function:

$$\delta^{(3)}(\vec{x}-\vec{y}) = \int \frac{d^3k}{(2\pi)^3} e^{i\vec{k}\cdot(\vec{x}-\vec{y})} \quad (3.58)$$

since $\delta^{(3)}(\vec{x} - \vec{y})$ has the property that

$$f(\vec{x}) = \int d^3y \delta^{(3)}(\vec{x} - \vec{y}) f(\vec{y}) \quad (3.59)$$

From the definition of the Dirac δ -function we see that it is the Fourier transform of 1. Thus $\delta^{(3)}(\vec{x})$ and 1 form a Fourier transform pair. In other words

$$\int d^3y e^{-i\vec{k}\cdot\vec{y}} \delta^{(3)}(\vec{y}) = 1 \quad (3.60)$$

3.3.3 Solving linear partial differential equations with Fourier transforms

Our discussion above opens up a new way for solving linear partial differential equations such as the ones we deal with in cosmological perturbation theory. Consider for example the following equation for a function $A(t, \vec{x})$

$$\ddot{A} + H(t)\dot{A} + \nabla^2 A = 0 \quad (3.61)$$

We may Fourier transform A to get $\tilde{A}(\eta, \vec{k})$. Then any time derivative acting on A can be written as acting on $\tilde{A}$ (try it). The action of the Laplacian ∇^2 is also straightforward. We can take the Laplacian inside the integral to get

$$\begin{aligned} \nabla^2 A(t, \vec{x}) &= \int \frac{d^3k}{(2\pi)^3} \nabla^2 e^{i\vec{k}\cdot\vec{x}} \tilde{A}(t, \vec{k}) \\ &= \int \frac{d^3k}{(2\pi)^3} e^{i\vec{k}\cdot\vec{x}} (-k^2) \tilde{A}(t, \vec{k}) \end{aligned}$$

Therefore our differential equation can be written as

$$\int \frac{d^3k}{(2\pi)^3} e^{i\vec{k}\cdot\vec{x}} \left[\ddot{\tilde{A}} + H(t)\dot{\tilde{A}} - k^2 \tilde{A} \right] = 0 \quad (3.62)$$

and dropping the integral

$$\ddot{\tilde{A}} + H(t)\dot{\tilde{A}} - k^2 \tilde{A} = 0 \quad (3.63)$$

Using Fourier transforms we have converted the partial differential equation into a (infinite) set of ordinary differential equations for each wavenumber k . Once the solutions of these ordinary differential equations is found, we can Fourier transform back to get the original function $A(t, \vec{x})$.

3.4 The Jeans instability

Now that we familiarized ourselves with Fourier transforms we can proceed to solve the 2nd order partial differential equation for the density contrast δ , so let's re-write it. It is

$$\ddot{\delta} - c_s^2 \nabla^2 \delta - 4\pi G \bar{\rho} \delta = 0 \quad (3.64)$$

We Fourier transform $\delta(t, \vec{x})$ to $\delta(t, \vec{k})$ (we drop the "tilde" as there is no confusion). The differential equation turns into

$$\ddot{\delta} + (k^2 c_s^2 - 4\pi G \bar{\rho}) \delta = 0 \quad (3.65)$$

which is the equation for simple harmonic motion. We know how to solve this and in fact we identify three cases. Let's define

$$\omega^2 = k^2 c_s^2 - 4\pi G \bar{\rho} \quad (3.66)$$

Then

$$\delta(t, \vec{k}) = \delta_0(\vec{k}) \cos(\omega t) + \delta_1(\vec{k}) \sin(\omega t) \quad \text{if } \omega^2 > 0 \quad (3.67)$$

$$\delta(t, \vec{k}) = \delta_0(\vec{k}) + \delta_1(\vec{k}) t \quad \text{if } \omega^2 = 0 \quad (3.68)$$

$$\delta(t, \vec{k}) = \delta_0(\vec{k}) e^{|\omega|t} + \delta_1(\vec{k}) e^{-|\omega|t} \quad \text{if } \omega^2 < 0 \quad (3.69)$$

where $\delta_0(\vec{k})$ and $\delta_1(\vec{k})$ are completely arbitrary functions of $\vec{k}$. As you can see the solutions we have just found separate into two distinct classes (barring the marginal case $\omega = 0$):

- Stable oscillatory solutions
- Unstable solutions with one exponentially growing mode.

The marginal case for which $\omega = 0$ defines a special value for k which we will denote k_J . It is

$$k_J = \frac{\sqrt{4\pi G \rho_0}}{c_s} \quad (3.70)$$

which is called the Jeans wavenumber. Modes for which $k > k_J$ lead to oscillatory and stable behaviour while modes for which $k < k_J$ lead to exponential growth. This exponential growth is called the Jeans instability.

Let us discuss what we have just found mathematically in terms of physics. From k_J we define a wavelength $\lambda_J = \frac{2\pi}{k_J}$. It is

$$\lambda_J = c_s \sqrt{\frac{\pi}{G \rho_0}} \quad (3.71)$$

and we call this the Jeans length. The Jeans length is proportional to the speed of sound. But the speed of sound came from the pressure of the fluid. Now what we have just found is that scales which are smaller than the Jeans length undergo oscillations. These oscillations are supported by the fluid pressure which on scales smaller than λ_J is stronger than gravity and holds the system against collapse. However, for scales larger than the Jeans length gravity dominates the pressure and the system undergoes gravitational collapse. Inevitably the Jeans length sets the largest size for a bound object. Objects larger than λ_J have to collapse under their own gravity. One special case comes to mind: the case for which $c_s = 0$. In that case $\lambda_J = 0$ which means that all scales are unstable as the fluid in this case has no pressure to counteract gravity. Cold Dark Matter comes very close to being such a fluid cosmologically.

Before concluding this section let us find the solutions in real space. They are the Fourier transforms of the k -space solutions found above, i.e.

$$\delta(t, \vec{x}) = \int \frac{d^3 k}{(2\pi)^3} e^{i\vec{k} \cdot \vec{x}} \left[\delta_+(\vec{k}) e^{i\omega t} + \delta_-(\vec{k}) e^{-i\omega t} \right] \quad (3.72)$$

which can be thought of as an infinite set of linear combinations of plane wave solutions.

3.5 Newtonian gravitational collapse in an expanding universe

Having discussed fluctuations in a static fluid let's move to the next level of difficulty which in turn is closer to a realistic Universe. We shall consider the evolution of fluctuations in an expanding fluid.

3.5.1 Setting up the system: the background equations

As in the case of the static fluid, we assume that the background solution is homogeneous and therefore the mass density has no spatial dependence, i.e. $\bar{\rho} = \bar{\rho}(t)$ only. However, we no longer assume that the fluid is static. Rather we assume that the background solution is given by Hubble's law:

$$\vec{v} = H(t)\vec{x} \quad (3.73)$$

where $\dot{\vec{x}} = 0$. This last point needs some clarification. Remember that

$$\vec{v} = \frac{d\vec{x}}{dt} = \dot{\vec{x}} + \vec{v} \cdot \vec{\nabla} \vec{x} \quad (3.74)$$

where the 2nd equality follows from application of the convective derivative. Now for any vector $\vec{A}$ we have that

$$\vec{A} \cdot \vec{\nabla} \vec{x} = \vec{A} \quad (3.75)$$

since

$$\vec{A} \cdot \vec{\nabla} \vec{x} = \left(A_x \frac{\partial}{\partial x} + A_y \frac{\partial}{\partial y} + A_z \frac{\partial}{\partial z} \right) \begin{pmatrix} x \\ y \\ z \end{pmatrix} \quad (3.76)$$

$$= \begin{pmatrix} A_x \\ A_y \\ A_z \end{pmatrix} \quad (3.77)$$

$$= \vec{A}. \quad (3.78)$$

This holds in particular for the velocity: $\vec{v} \cdot \vec{\nabla} \vec{x} = \vec{v}$. Hence, we must have that $\dot{\vec{x}} = 0$.

You would be right to suspect that (3.73) violates the principle of isotropy as it picks out a preferred direction given by $\vec{x}$. We can assume that this direction is radially outwards but this leaves us with a preferred location: the centre. Evidently this model is flawed as a model for the Universe. We will not be concerned with this as it is only an approximate model which will be superseded later by the fully relativistic theory of fluctuations.

Let us now proceed to solve the background system of equations. Consider first the continuity equation (3.29). Since $\vec{\nabla} \bar{\rho} = 0$ we get

$$\frac{\partial \bar{\rho}}{\partial t} + \bar{\rho} \vec{\nabla} \cdot \vec{v} = 0 \quad (3.79)$$

But $\vec{\nabla} \cdot \vec{v} = H \vec{\nabla} \cdot \vec{x}$, and furthermore $\vec{\nabla} \cdot \vec{x} = 3$ (check this out as an exercise), hence the background continuity equation becomes

$$\dot{\bar{\rho}} + 3H\bar{\rho} = 0 \quad (3.80)$$

We recognise this equation as the General Relativistic energy conservation equation for pressureless matter. But we have only used Newtonian theory to derive it which shows how good our approximation is. However, at this point we don't have a Friedman equation yet so we can't solve the above equation without postulating $H(t)$.

Substituting (3.73) into the Euler equation gives

$$\bar{\rho} \left(\dot{H}\vec{x} + H^2\vec{x} \cdot \vec{\nabla} \vec{x} \right) = -\bar{\rho} \vec{\nabla} \Phi \quad (3.81)$$

From (3.75) we get $\vec{x} \cdot \vec{\nabla} \vec{x} = \vec{x}$, hence, the Euler equation gives us $\vec{\nabla} \bar{\Phi}$ as

$$\vec{\nabla} \bar{\Phi} = - \left(\dot{H} + H^2 \right) \vec{x} \quad (3.82)$$

We hit it with $\vec{\nabla}$ to get

$$\nabla^2 \bar{\Phi} = - \left(\dot{H} + H^2 \right) \vec{\nabla} \cdot \vec{x} = -3 \left(\dot{H} + H^2 \right) \quad (3.83)$$

and using the Poisson equation we get

$$\dot{H} + H^2 + \frac{4\pi G}{3} \bar{\rho} = 0 \quad (3.84)$$

Notice that there is no Jeans swindle in an expanding Universe and we can consistently solve the background equations, including for the potential. We now turn to fluctuations around this background solution.

3.5.2 Equations for fluctuations

As for the static case we Taylor expand as

$$\rho = \bar{\rho}(1 + \delta) \quad (3.85)$$

$$\vec{v} = \vec{\bar{v}} + \vec{u} = H\vec{x} + \vec{u} \quad (3.86)$$

$$\Phi = \bar{\Phi} + \phi \quad (3.87)$$

Without proof, the continuity and Euler equations give

$$\dot{\delta} + \vec{\nabla} \cdot \vec{u} + H\vec{x} \cdot \vec{\nabla} \delta = 0 \quad (3.88)$$

and

$$\dot{\vec{u}} + H \left(\vec{u} + \vec{x} \cdot \vec{\nabla} \vec{u} \right) = -c_s^2 \vec{\nabla} \delta - \vec{\nabla} \phi \quad (3.89)$$

respectively. Furthermore, the perturbed Poisson equation gives

$$\nabla^2 \phi = 4\pi G \bar{\rho} \delta \quad (3.90)$$

(The derivation of the above three equations will be in the problem sheet).

At this point, to make further progress it is easier to switch to Fourier space. We have to be careful as the coordinate $\vec{x}$ is the physical coordinate while for Fourier space we should be using comoving coordinates (so that the expansion of the Universe is factored out). Comoving coordinates are those coordinates so that

$$\frac{d\vec{r}}{dt} = 0 \quad (3.91)$$

Clearly $\vec{x}$ is not comoving (in fact it is what is called a Eulerian coordinate). Applying the convective derivative we find that the comoving coordinate must satisfy

$$\dot{\vec{r}} + H\vec{x} \cdot \vec{\nabla} \vec{r} = 0. \quad (3.92)$$

Let's try a simple ansatz:

$$\vec{r} = \frac{\vec{x}}{a} \quad (3.93)$$

where $a(t)$ is an arbitrary function of time. Substituting (3.93) into (3.92) we find that for (3.93) to be a solution we must have that

$$H = \frac{\dot{a}}{a} \quad (3.94)$$

Hence, a has the interpretation of the "scale factor" but notice that we are dealing with Newtonian theory which doesn't have a metric!

Back to the task: Fourier transforms. The Fourier transform is to be used in conjunction with co-moving coordinates. The Fourier transform of δ , say, is

$$\delta = \int \frac{d^3k}{(2\pi)^3} e^{i\vec{k}\cdot\vec{r}} \tilde{\delta} \quad (3.95)$$

and substituting (3.93) we get ⁴

$$\delta = \int \frac{d^3k}{(2\pi)^3} e^{\frac{i\vec{k}\cdot\vec{x}}{a}} \tilde{\delta} \quad (3.96)$$

Leaving the proof to the problem set, the Fourier space equations are (dropping the "tilde")

$$\dot{\delta} + i\vec{k} \cdot \vec{\theta} = 0 \quad (3.97)$$

$$\dot{\vec{\theta}} + 2H\vec{\theta} = -\frac{i\vec{k}}{a^2} [c_s^2\delta + \phi] \quad (3.98)$$

$$-k^2\phi = 4\pi G a^2 \bar{\rho}\delta \quad (3.99)$$

where we have defined a rescaled velocity perturbation $\vec{\theta}$ by

$$\vec{u} = a\vec{\theta} \quad (3.100)$$

The new velocity vector $\vec{\theta}$ is called the *peculiar velocity field*.

Eliminating $\vec{k} \cdot \vec{\theta}$ as well as ϕ we find a 2nd order equation for δ

$$\ddot{\delta} + 2H\dot{\delta} + \left(\frac{c_s^2 k^2}{a^2} - 4\pi G \bar{\rho} \right) \delta = 0 \quad (3.101)$$

This is the equation for a *damped* harmonic oscillator with time-varying "mass". The damping factor comes from the Hubble expansion.

Equation (3.101) is very similar to the one we derived for the case of a static fluid. There are only two main differences:

- The term $H\dot{\delta}$. The expansion of the Universe gives rise to a damping term. This will have dramatic consequence on the solutions, in particular on the growing modes.
- The background density $\bar{\rho}$ is time dependent.

⁴You can now see why we need to define the Fourier transform in co-moving coordinates. If we take the total derivative $\frac{d\delta}{dt}$ then this is mapped into $\frac{d\tilde{\delta}}{dt}$ in Fourier space because $\frac{d\vec{r}}{dt} = 0$. On the other hand had we defined the Fourier transform using the Eulerian coordinate $\vec{x}$, i.e. as $\delta = \int \frac{d^3k}{(2\pi)^3} e^{i\vec{k}\cdot\vec{x}} \tilde{\delta}$ then taking $\frac{d\delta}{dt}$ we would pick up a term from $e^{i\vec{k}\cdot\vec{x}}$ and the mapping would have been $\frac{d\delta}{dt} \rightarrow \frac{d\tilde{\delta}}{dt} + i\vec{k} \cdot \vec{v} \tilde{\delta}$ which is not desirable.

3.5.3 The Jeans length in an expanding Universe

Let's first ignore the damping factor and proceed with the Jeans analysis in this case. As in the case of the static fluid we define the Jeans wavenumber as

$$k_J = a \frac{\sqrt{4\pi G \bar{\rho}}}{c_s} \quad (3.102)$$

Once again modes for which $k > k_J$ lead to oscillatory and stable behaviour (these will no longer be cosines and sines because of the damping factor but they will still be oscillatory). But how about modes for which $k < k_J$? Clearly these modes will not be oscillatory, but will they be unstable? Will they grow or decay? We cannot answer these questions without further assumptions about $H(t)$. The reason is that the damping factor can have a dramatic effect on whether a mode with $k < k_J$ is growing or decaying!

Just like the case of the static fluid we can define the Jeans length as $\lambda_J = \frac{2\pi}{k_J}$. This is the co-moving Jeans length (the physical Jeans length is found by multiplying by a) and is

$$\lambda_J^{(com)} = \frac{c_s}{a} \sqrt{\frac{\pi}{G \bar{\rho}(t)}} \quad (3.103)$$

The physics contained in the above equation are the same as in the static case albeit with one important difference: the Jeans length is time dependent. This means that a particular fluctuation mode can switch between a periods of oscillatory behaviour and periods of growth (or decay).

Before proceeding to solve (3.101) let us note another important fact. Although the background around which δ is fluctuating was considered to be that of pressureless matter, it turns out that (3.101) is valid for *any* cosmological background, including radiation domination or cosmological constant. To derive this fact we will need the relativistic theory of perturbations. However, (3.101) describes only the fluctuations of non-relativistic matter and cannot describe the fluctuations of radiation. We shall find the equivalent equation for radiation when we consider the relativistic theory.

3.5.4 Solutions during matter domination

We shall consider only solutions for which the scales involved are outside the Jeans length. This means that $k < k_J$ and therefore we can ignore the $k^2 c_s^2$ term. This assumption also corresponds to that for completely pressureless matter (such as cold dark matter) for all scales ($c_s = 0$ for dark matter). For these modes the density contrast evolves as

$$\ddot{\delta} + 2H\dot{\delta} - 4\pi G \bar{\rho} \delta = 0 \quad (3.104)$$

Furthermore, during matter domination $a = (t/t_0)^{2/3}$ and so $H = \frac{2}{3t}$. Moreover $4\pi G \bar{\rho} = 3H^2/2 = \frac{2}{3t^2}$ and our differential equation becomes

$$\ddot{\delta} + \frac{4}{3t}\dot{\delta} - \frac{2}{3t^2}\delta = 0 \quad (3.105)$$

To solve this equation we will try an educated guess that the solutions are given by powerlaws $\delta \propto t^n$. The reason is that the power of t that appears in each term in (3.105) is always matched by the number of time derivatives (since $\frac{d}{dt} \sim 1/t$). Substituting $\delta = t^n$ into (3.105) we get a quadratic equation for n which is

$$3n^2 + n - 2 = 0 \quad (3.106)$$

It has solutions $n = \frac{2}{3}$ and $n = -1$. The later corresponds to a decaying solution (as $1/t$) while the former corresponds to a growing solution. Thus the density contrast of pressureless matter during the matter era evolves as

$$\delta(t, \vec{k}) = \delta_0(\vec{k})(t/t_0)^{2/3} = \delta_0(\vec{k})a \quad (3.107)$$

This is how dramatic the damping effect of the expansion can be. It has converted an exponentially growing mode in the static case into a powerlaw growing mode for the expanding case. We will see below that for the case of radiation domination the effect is even more dramatic.

For the case of completely pressureless matter (for which $c_s = 0$ always) the Jeans length is zero and the above equation is valid for all k -modes. Therefore the solution in real space (at comoving position $\vec{r}$) is

$$\delta(t, \vec{r}) = a \delta_0(\vec{r}) = \int \frac{d^3k}{(2\pi)^3} e^{i\vec{k}\cdot\vec{r}} \delta_0(\vec{k}) \quad (3.108)$$

3.5.5 Solutions during radiation domination: the Mészáros effect.

During radiation domination, $a = (t/t_0)^{1/2}$, hence $H = \frac{1}{2t}$ and our differential equation becomes

$$\ddot{\delta} + \frac{1}{t}\dot{\delta} - 4\pi G\bar{\rho}\delta = 0 \quad (3.109)$$

where $\bar{\rho}$ is the background mass density of matter and does not include that of radiation. Since the background is assumed to be dominated by radiation, it means that the Friedmann equation is driven not by matter but by radiation. Therefore $4\pi G\bar{\rho} \ll 3H^2$ and we may ignore this term. Our differential equation then becomes

$$\ddot{\delta} + \frac{1}{t}\dot{\delta} = 0 \quad (3.110)$$

We may easily integrate this equation to get that

$$\delta(t, \vec{k}) = \delta_0(\vec{k}) + \frac{1}{2}\delta_1(\vec{k})\ln\frac{t}{t_0} = \delta_0(\vec{k}) + \delta_1(\vec{k})\ln a \quad (3.111)$$

Thus during radiation domination, the growth of matter fluctuations is at best logarithmic with time (or with the scale factor). This is called the Mészáros effect which has the consequence that during the radiation era matter fluctuations cannot grow enough to produce significant structure and only during the matter era where the growth is a powerlaw can significant structure form.

3.6 Peculiar velocities

So far we have been talking only about what happens to the density contrast δ . But our fluid equations can also give us information about the peculiar velocity $\vec{\theta}$.

Let us decompose the vector $\vec{\theta}$ into a scalar θ and a pure vector $\vec{\theta}_{rot}$ as follows

$$\vec{\theta} = \vec{\nabla}\theta + \vec{\theta}_{rot} \quad (3.112)$$

so that the pure vector $\vec{\theta}_{rot}$ obeys $\vec{\nabla} \cdot \vec{\theta}_{rot} = 0$. This last relation means that $\vec{\theta}_{rot} = 0$ is a curl mode: $\vec{\theta}_{rot} = \vec{\nabla} \times \vec{\theta}_{curl}$. What we have done here is to split the 3 independent components of $\vec{\theta}$ into 1 component (scalar θ) plus 2 independent components in $\vec{\theta}_{rot}$. The reason that $\vec{\theta}_{rot}$ has only two independent components can best be seen in Fourier space. The relation $\vec{\nabla} \cdot \vec{\theta}_{rot} = 0$ becomes $\vec{k} \cdot \vec{\theta}_{rot} = k_x\vec{\theta}_{rot,x} + k_y\vec{\theta}_{rot,y} + k_z\vec{\theta}_{rot,z} = 0$ so that we can solve for one of the components

of $\vec{\theta}_{rot}$ in terms of the other two. This decomposition is called the scalar-vector decomposition. We shall encounter this again further below when we consider the relativistic theory of cosmological perturbations and where we will be dealing with a scalar-vector-tensor decomposition.

We call θ a compressional mode and $\vec{\theta}_{rot}$ a rotational mode (curl implies rotation). Now consider again our perturbed Euler equation (3.98) and perform the scalar-vector decomposition on $\vec{\theta}$. We get

$$i\vec{k} \left[\dot{\theta} + 2H\theta + \frac{1}{a^2} (c_s^2 \delta + \phi) \right] + \dot{\vec{\theta}}_{rot} + 2H\vec{\theta}_{rot} = 0 \quad (3.113)$$

Acting again with $\vec{k}$, we kill the rotational part to get

$$\dot{\theta} + 2H\theta + \frac{1}{a^2} (c_s^2 \delta + \phi) = 0 \quad (3.114)$$

which when re-inserted into (3.113) gives

$$\dot{\vec{\theta}}_{rot} + 2H\vec{\theta}_{rot} = 0 \quad (3.115)$$

Notice also that the continuity equation depends only on the compressional part:

$$\dot{\delta} = k^2 \theta \quad (3.116)$$

We have managed to separate the equations obeyed by the rotational part from those obeyed by the compressional part. First let's have a look at the rotational part. We can solve (3.115) and the solution is

$$\vec{\theta}_{rot}(t, \vec{k}) = \frac{1}{a^2} \vec{\theta}_{rot}^{(in)}(\vec{k}) \quad (3.117)$$

We see that the rotational peculiar velocity decays as $1/a^2$. This means that we can safely neglect the rotational part from now on as it will be virtually unobservable in the late Universe.

Now let's consider the compressional part of the peculiar velocity. Since we have already found the solutions to δ we can simply read-off the evolution of θ from (3.116). It is customary to turn the time derivative into a derivative with respect to the scale factor. Equation (3.116) then gives θ as

$$\theta = \frac{H f(a)}{k^2} \delta \quad (3.118)$$

where $f(a)$ is called the "growth factor" and is given by

$$f(a) = \frac{d \ln \delta}{d \ln a} \quad (3.119)$$

In a Universe dominated by pressureless matter $f = 1$. If a cosmological constant is present then f may be approximated (Peebles 1980) by $f = \Omega_m^{0.6}$. Better approximations can be found by Taylor expanding all the relevant equations in powers of Ω_Λ . If we parameterize f as $f = \Omega_m^\gamma$ then for a Λ CDM Universe one finds that

$$\gamma = \frac{6}{11} + \frac{15}{2057} \Omega_\Lambda + O(\Omega_\Lambda^2). \quad (3.120)$$

Notice that the formula for θ contains a factor k^2 in the denominator. This means that it is dominated by larger scales ($k \rightarrow 0$) than the density contrast and therefore the deviations from the Hubble flow given by θ (if can be accurately measured) provide a better probe of inhomogeneities than the large scale clustering given by δ .

3.7 Cosmological perturbation theory

We are now ready for the next level of difficulty: the relativistic theory of cosmological perturbations.

Simply put, Cosmological Perturbation Theory is a form of Taylor expansion around a Friedman universe. This means that we have a known background metric given by the Robertson-Walker metric. For simplicity we shall assume a flat Universe. Furthermore when dealing with perturbation theory we shall be using conformal time coordinates. The background metric in these coordinates is

$$ds^2 = \bar{g}_{\mu\nu} dx^\mu dy^\nu = a^2(\eta) [-d\eta^2 + \gamma_{ij} dx^i dx^j] \quad (3.121)$$

where η is the conformal time and γ_{ij} is the Euclidean 3-dimensional metric in arbitrary coordinates. For example, in cartesian coordinates $\gamma_{ij} dx^i dx^j = dx^2 + dy^2 + dz^2$ and in spherical coordinates $\gamma_{ij} dx^i dx^j = dr^2 + r^2 d\vartheta^2 + r^2 d\varphi^2$. The 3-metric γ_{ij} will be used to raise and lower spatial indices, e.g. $v_i = \gamma_{ij} v^j$ and $v^i = \gamma^{ij} v_j$.

At the same time, we have the background density for fluids $\bar{\rho}(t)$ and background pressure $\bar{P}$ (and corresponding equation of state w). We shall adopt the following convention. If no subscript appears for a fluid variable, e.g. $\bar{\rho}$ then that variable corresponds to the total quantity (in this case total density) summed over all fluids. If a subscript appears then it is going to have the following meaning : "r" : radiation (photons plus neutrinos), "m" : matter (baryons plus cold dark matter), "b" : baryons, "c": cold dark matter, "γ": photons, "ν": neutrinos and finally "Λ": cosmological constant. If a different subscript appears, e.g. $\bar{\rho}_I$ then that usually means that the variable involved is for some arbitrary fluid. Usually this is used when we sum over all fluids, e.g. $\bar{\rho} = \sum_I \bar{\rho}_I$.

We have two sets of background equations. The first set comes from the background Einstein equations

$$\bar{G}^\mu{}_\nu = 8\pi G \bar{T}^\mu{}_\nu \quad (3.122)$$

Let us first define the *conformal Hubble parameter* $\mathcal{H}$ as

$$\mathcal{H} = \frac{a'}{a} \quad (3.123)$$

where a prime denotes differentiation with respect to the conformal time η . The conformal Hubble parameter is related to the normal Hubble parameter that you already know by

$$\mathcal{H} = aH \quad (3.124)$$

From the $\mu = \nu = 0$ component we get the first Friedman equation ($T^0_0 = -\bar{\rho}$)

$$3\mathcal{H}^2 = 8\pi G a^2 \bar{\rho} = 8\pi G a^2 \sum_I \bar{\rho}_I \quad (3.125)$$

and from the $\mu = \nu = i$ component (diagonal spatial components) we get the second Friedman equation ($T^i_j = \bar{P} \delta^i_j$)

$$-2\mathcal{H}' - \mathcal{H}^2 = 8\pi G a^2 \bar{P} = 8\pi G a^2 \sum_I \bar{P}_I \quad (3.126)$$

Due to the homogeneity and isotropy of the background, the $\mu = 0, \nu = i$ components as well as the off-diagonal $\mu = i, \nu = j, i \neq j$ components vanish.

The second set of equations comes from the conservation of the energy-momentum tensor of the fluid : $\nabla_\mu T^\mu_\nu = 0$. Setting $\nu = 0$ we get

$$\dot{\bar{\rho}}_I + 3\mathcal{H}(\bar{\rho}_I + \bar{P}_I) = 0 \quad (3.127)$$

for each fluid "I". Once again, due to the homogeneity and isotropy of the background the component $\nu = i$ vanishes. The equation above is the analogue of the continuity equation that we used in the Newtonian treatment, only now it includes pressure. Here it is derived directly from $\nabla_\mu T^\mu_\nu = 0$.

Compared to the Newtonian treatment notice that there is no background velocity for the fluid: it is zero. The reason is that in the Newtonian treatment we had to impose the Hubble expansion by hand by assuming that $\vec{v} = H\vec{x}$. However, for the Friedman Universe this is already taken care for and it is already part of the metric: the Hubble expansion is provided by gravity. Another way to think of this is that if a fluid has a non-zero background velocity, then it will violate the homogeneity and isotropy of the Universe by picking out a preferred direction along $\vec{v}$.

3.7.1 Setting up the perturbations

We now perturb our background variables. First we perturb the metric around the background $\bar{g}_{\mu\nu}$. The full Einstein metric $g_{\mu\nu}(t, \vec{x})$ to first order in perturbation theory is given by

$$g_{\mu\nu} = \bar{g}_{\mu\nu} + \delta g_{\mu\nu} \quad (3.128)$$

where $\delta g_{\mu\nu} \ll \bar{g}_{\mu\nu}$. If we insist that $\delta g_{\mu\nu}$ is small then we proceed to calculate the Christoffel symbols to 1st order in $\delta g_{\mu\nu}$, then the Ricci and scalar curvature tensors and finally the Einstein tensor, always keeping at most 1st order in $\delta g_{\mu\nu}$, i.e. terms which go as $(\delta g)^2$ or higher are ignored. When this procedure is followed, we get that the Einstein tensor is perturbed as $G^\mu_\nu = \bar{G}^\mu_\nu + \delta G^\mu_\nu$.

In a similar fashion, we assume that the full energy-momentum tensor is also perturbed as $T^\mu_\nu = \bar{T}^\mu_\nu + \delta T^\mu_\nu$. The perturbed Einstein equations then read

$$\delta G^\mu_\nu = 8\pi G \delta T^\mu_\nu \quad (3.129)$$

The derivation of the perturbed Einstein equations will be dealt with in a non-assessed problem set so here we will give the answer after we have considered a few more simplifications in the next subsection. The perturbed equations obtained will be a set of linear partial differential equations so once again we will make heavy use of Fourier transforms. In particular we shall be expanding all relevant variables as

$$A(\eta, \vec{r}) = \int \frac{d^3k}{(2\pi)^3} e^{i\vec{k}\cdot\vec{r}} \tilde{A}(\eta, \vec{k}) \quad (3.130)$$

We shall use the notation $\vec{r} \leftrightarrow x^i$ and $\vec{k} \leftrightarrow k_i$ so that $\vec{k} \cdot \vec{r} = k_i x^i$. Furthermore, every time we have a spatial derivative $\vec{\nabla}_i$ we convert it into ik_i where k_i is the Fourier wavevector. This means that $\nabla^2 \rightarrow -k^2$ where $k^2 = \gamma^{ij} k_i k_j$. We shall be using $\vec{\nabla}_i$ and k_i interchangeably throughout the course.

3.7.2 The scalar-vector-tensor decomposition

It turns out that things equations simplify if we use the so called 3+1 decomposition of all tensors. That is we split all tensor components into a "time"-part and a "space"-part. In the following

discussion we shall consider vectors and tensors which are small fluctuations around a given background field. If we have a four vector $v_\mu(\eta, \vec{r})$ we will consider the "time" component $v_0(\eta, \vec{r})$ and "space" component $v_i(\eta, \vec{r})$ separately as we have done earlier for the Friedman Universe. The part $v_i(\eta, \vec{r})$ may then be considered as a 3-dimensional spatial vector (actually a set of them labelled by η) while the part $v_0(\eta, \vec{r})$ may be considered as a scalar (as it has no spatial index). The part $v_i(\eta, \vec{r})$ may even be further decomposed into a "longitudinal" part and a "transverse" part:

$$v_i(\eta, \vec{r}) = \vec{\nabla}_i v(\eta, \vec{r}) + \hat{v}_i(\eta, \vec{r}) \quad (3.131)$$

with the condition that $\gamma^{ij} \vec{\nabla}_i \hat{v}_j = \vec{\nabla}^i \hat{v}_i = 0$.

Consider first the variable v . It has no index and is therefore a "scalar mode". It acquires an index through the gradient operator $\vec{\nabla}_i$ to become part of v_i . Note that $\nabla^2 v = \vec{\nabla}^i v_i$ or in Fourier space, $v = -\frac{i}{k^2} k^i v_i$.

Consider now the part $\hat{v}_i$. Since $\vec{\nabla}^i \hat{v}_i = 0$, this means that it can be written as a curl mode, i.e. $\hat{v}_i = (\vec{\nabla} \times \vec{S})_i$ for some other 3-vector S_i . This in turn means that $\hat{v}_i$ is a "pure" vector perturbation in the sense that it cannot be obtained from a scalar mode via some gradient operator. The vector perturbation $\hat{v}_i$ is also orthogonal to the direction specified by $\vec{\nabla}_i$. This means that it falls on a 2-dimensional subspace. If we choose as a basis on this 2-dimensional subspace two vectors m_i and n_i we can decompose $\hat{v}_i = \hat{v}^+ m_i + \hat{v}^- n_i$. We say that $\hat{v}_i$ has two polarizations: $\hat{v}^+$ and $\hat{v}^-$, or that the vector perturbation $\hat{v}_i$ contains two vector modes.

To summarize, we see that we can decompose the 4 components of v_μ into 2 *scalar modes*: v_0 and v and 2 *vector modes* $\hat{v}_i$ which obey $\vec{\nabla}^i \hat{v}_i = 0$. Similar arguments hold for a contravariant vector field v^μ .

Now consider the next level of complexity: a 2nd rank symmetric tensor $h_{\mu\nu}$, e.g. the metric tensor. Performing the 3 + 1 decomposition first, we need to consider the following parts: h_{00} , h_{0i} and h_{ij} . Clearly h_{00} is a scalar mode and h_{0i} contains a scalar mode and two vector modes, just as we discussed above. So we are left with h_{ij} . Since it is symmetric we can take its trace $h = \gamma^{ij} h_{ij}$ which gives us 1 scalar mode. We subtract the trace to obtain a traceless symmetric tensor $h_{ij} - \frac{1}{3} h \gamma_{ij}$. We can form yet another scalar mode by contracting it with $\vec{\nabla}^i \vec{\nabla}^j$ i.e. $\vec{\nabla}^i \vec{\nabla}^j h_{ij} - \frac{1}{3} \nabla^2 h$. We would find it very useful to introduce the traceless operator

$$D_{ij} = \vec{\nabla}_i \vec{\nabla}_j - \frac{1}{3} \nabla^2 \gamma_{ij} \quad (3.132)$$

i.e. so that $\gamma^{ij} D_{ij} = 0$. Then $\vec{\nabla}^i \vec{\nabla}^j h_{ij} - \frac{1}{3} \nabla^2 h = D^{ij} h_{ij}$. This defines a 2nd scalar mode ν such that $D^{ij} D_{ij} \nu = D^{ij} h_{ij} = \frac{2}{3} \nabla^4 \nu$.

So now out of the 6 components of h_{ij} we found 2 scalars, therefore we have 4 dof to go. We subtract the two scalars to form $\tilde{h}_{ij} = h_{ij} - \frac{1}{3} h \gamma_{ij} - D_{ij} \nu$. Clearly then $\gamma^{ij} \tilde{h}_{ij} = \vec{\nabla}^i \vec{\nabla}^j \tilde{h}_{ij} = 0$ (check it yourself) so this new auxiliary tensor $\tilde{h}_{ij}$ does not contain any more scalar modes. We may then guess that it should contain some vector modes, but how many? Suppose that it does contain two vector modes f_i , so that $\vec{\nabla}^i f_i = 0$, (remember our discussion above concerning the two vector-mode polarizations of a vector perturbation f_i). Now we want to build $\tilde{h}_{ij}$ from them. We cannot have a combination $f_i f_j$ because that would be 2nd order. Therefore the only possibility left is to have the combination $\frac{1}{2} (\vec{\nabla}_i f_j + \vec{\nabla}_j f_i) = \vec{\nabla}_{(i} f_{j)}$. But f_i contribute only two dof and we need two more. Is it possible that there are more vector modes in $\tilde{h}_{ij}$? The answer is NO! To see this consider $\vec{\nabla}^j \tilde{h}_{ij}$. This would contain $\frac{1}{2} \nabla^2 f_i$ so any other vector mode can be absorbed into f_i .

In fact the two dof left are part of what we call a purely *tensor* perturbation χ_{ij} . The tensor mode obeys the transverse-traceless conditions $\gamma^{ij}\chi_{ij} = \vec{\nabla}^j\chi_{ij} = 0$. As with the vector perturbations, the tensor perturbation χ_{ij} also falls on a two dimensional subspace and therefore also contains two polarizations, i.e. two tensor modes. At this point we have succeeded in identifying all modes present in a tensor $h_{\mu\nu}$, that is the 10 components of $h_{\mu\nu}$ are decomposed into 4 *scalar* modes, 4 *vector modes* and 2 *tensor modes*.

We shall not need any other types of tensors for this course but I leave it as an exercise to find out how to decompose an anti-symmetric 2nd rank tensor $F_{\mu\nu}$. How does a general 2nd rank tensor decompose?

3.7.3 The general form of $\delta g_{\mu\nu}$ and $\delta T_{\mu\nu}$

We are now in position to discuss the perturbations to the metric $\delta g_{\mu\nu}$ and to the energy-momentum tensor $\delta T_{\mu\nu}$. Following our discussion in the last subsection, we identify 4 scalar modes in the metric perturbation, 4 vector modes and 2 tensor modes. Remember the "rotational" part of the peculiar velocity $\vec{\theta}_{rot}$? That was an example of a vector mode and we found that it decayed as $1/a^2$. It turns out that usually vector modes tend to decay and are therefore not that important in cosmology. We shall therefore ignore them in this course.

In General Relativity the metric perturbation is not unique but is subject to something called a "gauge transformation" (this is similar to an infinitesimal coordinate transformation). Due to these gauge transformations we are free to set two scalar modes to zero (but we can't eliminate the tensor modes). In this course we will always work in a specific gauge which is called the Conformal Newtonian Gauge. As the name implies, in this gauge, the metric perturbations and the equations they obey look as close as possible to their Newtonian analogues. In particular, we will have two scalar modes which are the Newtonian potentials Φ and Ψ and two tensor modes $h_{ij}^{(T)}$. Thus, ignoring the vector modes, the metric perturbation in the Conformal Newtonian Gauge takes the following form

$$\delta g_{\mu\nu} = a^2 \begin{pmatrix} -2\Psi & 0 \\ 0 & -2\Phi\gamma_{ij} + h_{ij}^{(T)} \end{pmatrix} \quad (3.133)$$

or in alternative form

$$ds^2 = a^2 \left[-(1 + 2\Psi)d\eta^2 + (1 - 2\Phi)\gamma_{ij}dx^i dx^j + h_{ij}^{(T)}dx^i dx^j \right] \quad (3.134)$$

Notice that in General Relativity we have two potentials, not one!

The energy-momentum tensor has a similar decomposition. It is easier to define our variables by using the mixed tensor T^μ_ν and we shall also pull out a factor $\bar{\rho}$. Consider the general form for the energy-momentum tensor

$$T^\mu_\nu = (\rho + P)u^\mu u_\nu + P\delta^\mu_\nu + \sigma^\mu_\nu$$

where u^μ is the 4-velocity (normalized as $g_{\mu\nu}u^\mu u^\nu = -1$) and σ^μ_ν is a tensor describing anisotropic stress (shear). The shear tensor is something new which did not exist on a Friedman background, simply because the Friedman spacetime has no anisotropy. But it does exist at the perturbative level and so σ^μ_ν is treated as a 1st order quantity. To proceed we define the energy density fluctuation as $\delta\rho$. If we further define the energy density contrast as

$$\delta = \frac{\delta\rho}{\bar{\rho}} \quad (3.135)$$

then the total energy density is $\rho = \bar{\rho}(1 + \delta)$. Similarly we define the pressure fluctuation as δP and the pressure contrast Π

$$\Pi = \frac{\delta P}{\bar{P}} \quad (3.136)$$

where now we normalized Π to the energy density rather than the pressure. The reason is that the background pressure can be zero. Then the total pressure is $P = \bar{P} + \delta P = \bar{\rho}(w + \Pi)$.

Finally we need the velocity perturbation. Remember that the velocity is normalized as $u_\mu u_\nu g^{\mu\nu} = -1$. This means that δu_0 component is not free but is fixed in terms of the metric fluctuation as $\delta u_0 = a(1 + \Psi)$. The component u_i is free and represents the 3-velocity fluctuation⁵. We pull out a normalization factor a as in the case of δu_0 and let

$$u_i = a\vec{\nabla}_i u \quad (3.137)$$

where u is the scalar part of the velocity fluctuation (we are ignoring vector modes).

We can now proceed and find the form of δT^μ_ν . It is

$$\delta T^0_0 = -\bar{\rho}\delta \quad (3.138)$$

$$\delta T^0_i = -\bar{\rho}(1 + w)\vec{\nabla}_i u \quad (3.139)$$

$$\delta T^i_0 = \bar{\rho}(1 + w)\vec{\nabla}^i u \quad (3.140)$$

$$\delta T^i_j = \bar{\rho} \left\{ \frac{1}{3}\Pi\delta^i_j + (1 + w) \left[D^i_j \sigma + \sigma^{(T) i}_j \right] \right\} \quad (3.141)$$

where σ is the scalar anisotropic stress and $\sigma^{(T)}_{ij}$ is the tensor anisotropic stress.

3.7.4 Einstein and fluid equations for scalar modes

We first consider only scalar modes (remember that we can deal with scalar, vector and tensor mode separately). In this case the metric is

$$ds^2 = -a^2(1 + 2\Psi)d\eta^2 + a^2(1 - 2\Phi)\gamma_{ij}dx^i dx^j \quad (3.142)$$

We find the Einstein equations (after some calculation that is left to the non-assessed problem sheet) as

$$\delta G^0_0 \quad 2\nabla^2\Phi - 6\mathcal{H}(\Phi' + \mathcal{H}\Psi) = 8\pi Ga^2 \sum_I \bar{\rho}_I \delta_I \quad (3.143)$$

$$\delta G^0_i : \quad 2(\Phi' + \mathcal{H}\Psi) = 8\pi Ga^2 \sum_I (\bar{\rho}_I + \bar{P}_I) u_I \quad (3.144)$$

$$\delta G^i_i : \quad \Phi'' + \mathcal{H}\Psi' + 2\mathcal{H}\Phi' + \left(2\mathcal{H}' + \mathcal{H}^2 + \frac{1}{3}\nabla^2 \right) \Psi - \frac{1}{3}\nabla^2\Phi = 4\pi Ga^2 \sum_I \bar{\rho}_I \Pi_I \quad (3.145)$$

and

$$\delta G^i_j \quad i \neq j : \quad \Phi - \Psi = 8\pi Ga^2 \sum_I (\bar{\rho}_I + \bar{P}_I) \sigma_I \quad (3.146)$$

⁵We use the word momentum to stress that we are perturbing u_μ and not u^μ which is the velocity. The covariant variable u_μ is up-to a multiplicative factor given by the mass, equal to the canonical momentum.

These equations look strikingly similar to the Poisson equation (take (3.143) for example). The biggest difference is that gravity is now sourced by velocities, pressures and shear in addition to the density. The other difference is that the potentials obey differential equations in time as well as space. This is also a relativistic effect as time and space are treated equally.

The fact that we have time derivatives on the potentials, however, is misleading. In fact it turns out that both Φ and Ψ are not independent dynamical degrees of freedom. This means that we cannot set initial conditions for them independently from the other variables. We can see this as follows.

We combine (3.143) and (3.144) and we can find Φ in terms of the matter variables as

$$\nabla^2 \Phi = 4\pi G a^2 \sum_I \bar{\rho}_I [\delta_I + 3\mathcal{H}(1 + w_I)u_I] \quad (3.147)$$

while Ψ is then obtained using (3.146). The advantage of the Newtonian gauge is now clearly seen: the potentials are non-dynamical (but they are time dependent) and are completely fixed by the evolution of the matter fields. Furthermore, (3.147) looks very similar to the Poisson equation in Newtonian gravity, only now it is sourced by the velocity as well. If there is no matter present then we find that $\nabla^2 \Phi = \nabla^2 \Psi = 0$.

Apart from the Einstein equations, we also need the evolution equations for each fluid. These are given by

$$\nabla_\mu T^\mu_\nu = 0 \quad (3.148)$$

for each fluid. Once again we leave the calculation to the non-assessed problem sheet and here we quote the answer. The two evolution equations obtained from (3.148) are the relativistic analogue of the continuity equation

$$\delta' = 3\mathcal{H}(w\delta - \Pi) + (1 + w)(\nabla^2 u + 3\Phi') \quad (3.149)$$

and the relativistic analogue of the Euler equation

$$u' = -\mathcal{H}(1 - 3w)u - \frac{w'}{1 + w}u + \frac{1}{1 + w}\Pi + \frac{2}{3}\nabla^2 \sigma + \Psi \quad (3.150)$$

where it is understood that all fluid variables refer to a single fluid. Notice that only δ and u are dynamical variables. The pressure fluctuation Π and shear σ are not determined by (3.148) and therefore (3.149) and (3.150) are not closed unless we have a way of determining Π and σ . Usually Π is determined by an equation of state and we single out the following cases:

- pressureless matter, for instance Cold Dark Matter and Baryons: $\Pi = 0$.
- relativistic fields, for instance photons and massless neutrinos $\Pi = \frac{1}{3}\delta$.

The shear σ is zero for both CDM and baryons but not for photons and neutrinos. Rather it is determined from the Boltzmann equation which we shall briefly touch when we consider the Cosmic Microwave Background. For massive neutrinos things are more complicated and we shall not deal with it in this course.

3.7.5 Einstein and fluid equations for tensor modes

For tensor modes, the metric takes the form

$$ds^2 = -a^2 d\eta^2 + a^2 \left(\gamma_{ij} + h_{ij}^{(T)} \right) dx^i dx^j \quad (3.151)$$

Dropping the indices on $h_{ij}^{(T)}$ we find that the metric tensor mode evolves as

$$h^{(T)''} + 2\mathcal{H}h^{(T)'} - \nabla^2 h^{(T)} = 16\pi G a^2 \sum_I (\rho_I + P_I) \sigma_I^{(T)} \quad (3.152)$$

and so it is sourced by the tensor mode of the anisotropic stress $\sigma^{(T)}$. In this case (3.148) does not provide us with any evolution equations. Rather $\sigma^{(T)}$ is given by the Boltzmann equation.

Unlike the Newtonian potentials Φ and Ψ , the tensor gravitational perturbation $h^{(T)}$ is a fully dynamical quantity. This means that to determine its evolution we have to specify initial conditions for it, independently of the initial conditions specified for the matter fields. The tensor mode $h^{(T)}$ is what we call graviton and is the part of the metric responsible for gravitational waves. The tensor modes do not participate in structure formation, only scalar modes do. However, the tensor modes imprint themselves on the Cosmic Microwave Background anisotropy spectrum and are therefore detectable.

3.7.6 Evolution of the potential for fluids with zero shear

Before considering the more general case where the Universe contains different species, let's focus on the single fluid case.

We shall also assume that the equation of state w is *constant* and equal to the speed of sound c_s^2 so that $\Pi = w\delta$. This is valid for pressureless matter and for radiation. Furthermore we shall assume that the shear is negligible, an assumption valid for pressureless matter but also quite good in the case of radiation. For zero shear, the Einstein equation (3.146) says that the two Newtonian potentials are equal $\Psi = \Phi$ (and we choose to use Φ).

Since we have a single fluid, we can use the background Friedman equations to trade ρ for $\mathcal{H}$. Thus the gravitational equation (3.143) gives the potential Φ as

$$-2k^2\Phi - 6\mathcal{H}(\Phi' + \mathcal{H}\Phi) = 3\mathcal{H}^2\delta \quad (3.153)$$

while (3.147) gives

$$\Phi = -\frac{3\mathcal{H}^2}{2k^2} [\delta + 3(1+w)\mathcal{H}u] \quad (3.154)$$

Finally, (3.145) gives

$$\Phi'' + 3\mathcal{H}\Phi' + (2\mathcal{H}' + \mathcal{H}^2)\Phi = \frac{3}{2}\mathcal{H}^2w\delta \quad (3.155)$$

and we may eliminate δ and u to get a single equation for Φ .

$$\Phi'' + 3\mathcal{H}(1+w)\Phi' + wk^2\Phi = 0 \quad (3.156)$$

Let us now try to solve these equations. Unfortunately to find the full solution is not possible without introducing special functions. So we shall find the solution under two approximations: super-horizon scales and sub-horizon scales.

First consider super-horizon scales. By super-horizon scales we mean that the wavelength $2\pi/k$ of the perturbations is larger than the horizon. Now the horizon is $\sim \frac{1}{\mathcal{H}}$, hence by super-horizon scales we mean that

$$k < \mathcal{H} \quad \text{super-horizon condition} \quad (3.157)$$

Obviously the above condition is time dependent, i.e. a particular Fourier mode with wavenumber k starting outside the horizon will subsequently enter the horizon because $\mathcal{H}$ decreases with time.

First take (3.156) and impose (3.157). This means that we can set the k^2 term to zero so for super-horizon scales (3.156) becomes

$$\Phi'' + 3\mathcal{H}(1+w)\Phi' = 0 \quad (3.158)$$

Clearly, one solution is that Φ is constant. To find the other solution, notice that (3.156) has the same form as the equation for energy conservation. Hence, the other solution is found from $\Phi' = \Phi_1 a^{-3(1+w)}$ which is a decaying solution and we will ignore it. Therefore on super-horizon scales, and as long as w is constant (not during the transition between matter and radiation) the potential Φ stays constant in time (but not in space):

$$\Phi(t, \vec{k}) = \Phi_{sup}(\vec{k}) \quad (3.159)$$

Now let's find δ for super-horizon scales. We insert (3.159) into (3.153) which gives $-2k^2\Phi - 6\mathcal{H}^2\Phi = 3\mathcal{H}^2\delta$. For super-horizon scales the term $k^2\Phi$ is much smaller than the term $\mathcal{H}^2\Phi$ so we ignore it. Therefore cancelling $\mathcal{H}$ we get that on super-horizon scales, the total density contrast is also constant in time and is related to Φ_{sup} by

$$\delta_{sum}(\vec{k}) = -2\Phi_{sup}(\vec{k}) \quad (3.160)$$

Finally, we use the above relation into (3.154) to get $2(k^2 - 3\mathcal{H}^2)\Phi_{sup} = -9\mathcal{H}^3(1+w)u$. Once again we can ignore the k^2 term and solve for u to get

$$u_{sup}(\vec{k}) = \frac{2}{3(1+w)\mathcal{H}}\Phi_{sup} \quad (3.161)$$

In this case u_{sup} is not constant in time. We can get $\mathcal{H}$ from the Friedman equation. It is given by $\mathcal{H} = \frac{2}{1+3w} \frac{1}{\eta}$ hence

$$u_{sup}(\vec{k}) = \frac{1+3w}{3(1+w)}\Phi_{sup} \eta \quad (3.162)$$

We find that u_{sup} increases linearly with η on super-horizon scales.

Let us pause for a moment. We have found the following:

- On super-horizon scales all fluctuations are given by the *same* initial condition which we have expressed as Φ_{sup} . This are called *curvature* or *adiabatic* initial conditions.
- Our solutions are fairly general as the only assumption about the background dynamics is that w is constant. This means that the same solutions hold for both the radiation era and the matter era. The only thing that changes is the value of w .
- Since as $\eta \rightarrow 0$ we have that $u_{sup} \rightarrow 0$, adiabatic initial conditions are equivalent to saying that there is no initial velocity in the fluid as we approach the big bang.
- Even though the background density diverges as $\eta \rightarrow 0$ (because $a \rightarrow 0$), the fluctuations remain regular and finite.
- Actually, the solution $\Phi_{sup} = \text{const}$ is also valid for non-relativistic matter on ALL scales! The reason is that setting $w = 0$ in (3.156) has the same effect as $k = 0$.

Let us exploit the last fact even more to investigate the evolution of *pressureless matter* on sub-horizon scales. It is a very easy step: Since $\Phi_{sup} = \text{const}$ is a solution for non-relativistic matter on subhorizon scales, then $\Phi' = 0$. Then consider the fluid equations (3.149) and (3.150) for $w = \Pi = \sigma = 0$ (pressureless matter) and also use $\Phi' = 0$. They become:

$$\delta' = -k^2 u \quad (3.163)$$

and

$$u' = -\mathcal{H}u + \Phi \quad (3.164)$$

respectively. These are identical (up-to coordinate transformation to cosmic time t) to the continuity and Euler equation we have already found in the Newtonian treatment. Thus the Newtonian treatment is a very good approximation even in relativistic cosmology for a Universe which contains only pressureless matter. The solution for δ (which we have already found in the Newtonian section) is now easily found in the matter era once we impose $\Phi = \text{const}$ in (3.153). For then we get $-2(k^2 + 3\mathcal{H}^2)\Phi = 3\mathcal{H}^2\delta$ and for sub-horizon scales $k^2 > \mathcal{H}^2$ so that we read-off δ as

$$\delta = -\frac{2k^2}{3\mathcal{H}^2}\Phi = -\frac{k^2\eta^2}{6}\Phi \propto a \quad (3.165)$$

Finally let's consider radiation on sub-horizon scales. The equation for the potential (3.156) acquires a "mass term" $\frac{1}{3}k^2\Phi$ hence we expect the potentials to be oscillating with a decaying amplitude (due to the damping term). What happens physically is the the Jeans length for radiation is the horizon. Therefore we expect the radiation density contrast inside the horizon to oscillate and quickly become subdominant.

3.7.7 Simplified equations for matter and radiation

For the rest of the course we shall only be interested in a Universe which contains pressureless matter and radiation (and a cosmological constant but this doesn't contribute to perturbations). We shall neglect the radiation shear. The shear comes mainly from the neutrino contribution (the photon shear is suppressed for reasons we shall see next weak) and is also smaller than the other variables. So let's adapt our general equations to this particular case.

For zero shear, the Einstein equation (3.146) says that the two Newtonian potentials are equal $\Psi = \Phi$ (and we choose to use Φ). The other Einstein equations give Φ and its first derivative as

$$\Phi = -\frac{4\pi G a^2}{k^2} [\bar{\rho}_m (\delta_m + 3\mathcal{H}u_m) + \bar{\rho}_r (\delta_r + 4\mathcal{H}u_r)] \quad (3.166)$$

and

$$\Phi' + \mathcal{H}\Phi = 4\pi G a^2 (\bar{\rho}_m u_m + \frac{4}{3}\bar{\rho}_r u_r) \quad (3.167)$$

respectively.

The fluid equations for pressureless matter become

$$\delta'_m = -k^2 u_m + 3\Phi' \quad (3.168)$$

and

$$u'_m = -\mathcal{H}u_m + \Phi \quad (3.169)$$

	super-horizon radiation era	super-horizon matter era	sub-horizon radiation era	sub-horizon matter era
Φ	const	const	oscillate, decay	const
δ_m	const	const	const (+ log)	grow as η^2
u_m	grow as η	grow as η	decay	grow as η

Table 1: Summary of solutions for the potential Φ , matter density contrast δ_m and velocity u_m .

while the equations for radiation are

$$\delta'_r = -\frac{4}{3}k^2 u_r + 4\Phi' \quad (3.170)$$

and

$$u'_r = \frac{1}{4}\delta_r + \Phi \quad (3.171)$$

Now let's consider solutions. Fortunately we can use the solutions for a single fluid that we have found in the previous subsection. In particular for super-horizon scales the potential will be constant $\Phi_{sup}(\vec{k})$ and for the *total* energy density we have $\delta = -2\Phi_{sup}$. Now $\delta = \Omega_m \delta_m + \Omega_r \delta_r$ but only ONE of the two fluids, either matter or radiation will be dominating. Therefore if we are in the radiation era, $\delta \approx \delta_r$ and if we are in the matter era then $\delta \approx \delta_m$. So given one of the δ how do we get the other one? We use (without proof) the so called adiabatic condition:

$$\delta_m = \frac{3}{4}\delta_r \quad (3.172)$$

Therefore on super-horizon scales we isolate the following two cases:

$$\text{Radiation era} \quad \delta_r = -2\Phi_{sup} \quad \delta_m = \frac{3}{4}\delta_r = -\frac{3}{2}\Phi_{sup} \quad (3.173)$$

$$\text{Matter era} \quad \delta_m = -2\Phi_{sup} \quad \delta_r = \frac{4}{3}\delta_m = -\frac{8}{3}\Phi_{sup} \quad (3.174)$$

On sub-horizon scales, the matter fluctuation δ will evolve as in the Newtonian case and so during radiation era $\delta_m \approx \text{constant}$ (the log-term is present only if there is initial δ' which we shall ignore) and during the matter era $\delta_m \propto a$.

A summary of the solutions found is given in table-1.

3.8 Probes of Large Scale Structure

So far we have been developing the theory of cosmological perturbations. It's time that we start applying it to real world physics. What we would like to do now is to understand the formation of structure in detail but at the same time create a set of observables that can be used to test whether our cosmological model of structure formation is correct.

3.8.1 The process of structure formation

We have seen in the previous subsection that the density contrast of pressureless matter inside the horizon, grows at best logarithmically with time during the radiation era, but grows as a powerlaw during the matter era. Outside the horizon the picture is different: the density contrast stays

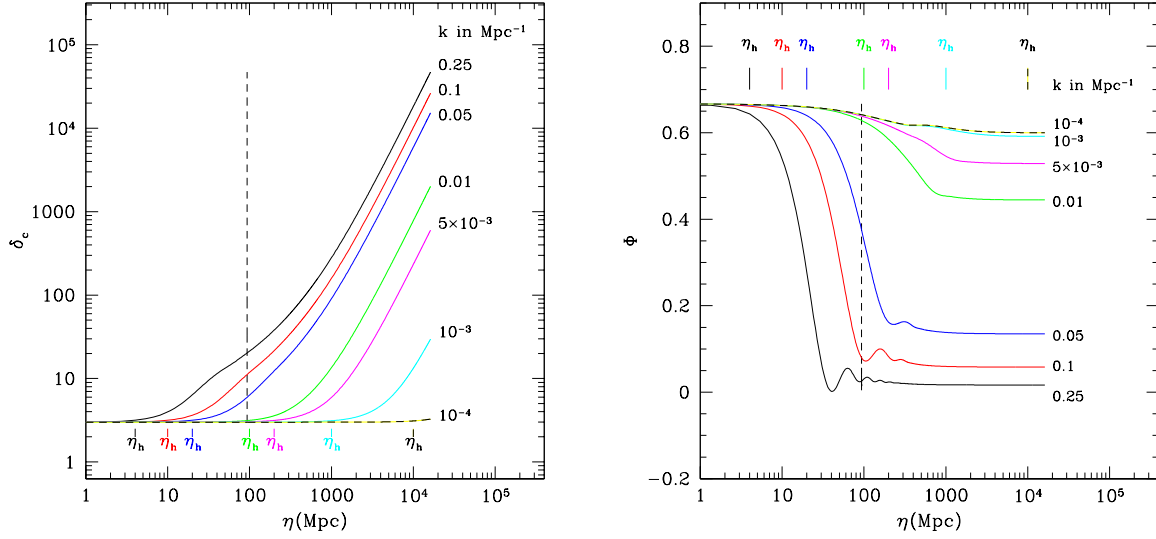


Figure 3.6: Left: The density contrast δ_c for pressureless matter in a Universe containing photons and matter. The radiation-matter equality η_{eq} is shown by a vertical dashed line. Horizon crossing is indicated by a vertical line and η_h for each k -mode. We see the effects derived in the lectures: (1) all modes stay constant outside the horizon, (2) modes entering the horizon in the radiation era grow logarithmically and then as a powerlaw in the matter era, (3) modes entering the horizon in the matter era grow as a power law.

Right: The potential Φ for the same model as in the left panel. Once again we see the effects derived in the lectures: (1) all modes stay constant outside the horizon, (2) sub-horizon modes in the radiation era oscillate and decay in amplitude, (3) sub-horizon modes in the matter era stay constant.

constant during both the radiation and the matter eras. Now for a fixed wavenumber k , a given perturbation mode $\delta(\vec{k})$ starts initially outside the horizon (for $\eta < k^{-1}$) and then at some point it crosses the horizon at $\eta_h \sim k^{-1}$. Depending on the value of k , horizon crossing may happen either during the radiation era or the matter era. Therefore a given perturbation mode will go through either two or all three of the evolutionary phases we found. Cold dark matter is exactly pressureless (actually its Jeans length is tiny compared with cosmological scales) and therefore follows this kind of picture. The left panel of figure 3.6 displays this behaviour for a number of k -modes. The right panel displays the gravitational potential Φ for the same model and the same k -modes.

Let's now talk a bit about baryons because it turns out that baryons are not always pressureless (they are for the background but not at the fluctuation level). Baryons for our purposes are composed of the light elements, i.e. roughly $\sim 76\%$ Hydrogen, 24% Helium and tiny fractions for the rest. To put it differently, baryons are composed of protons ($\sim 76\%$) and Helium nuclei ($\sim 24\%$). Both of these are charged, therefore baryons interact electromagnetically. This means that baryons are coupled with photons. How strong is the coupling depends on the temperature in the Universe which in turn depends on the number density of photons and the number density of baryons. We shall study this in more detail later but for the time being it suffices to say that in the

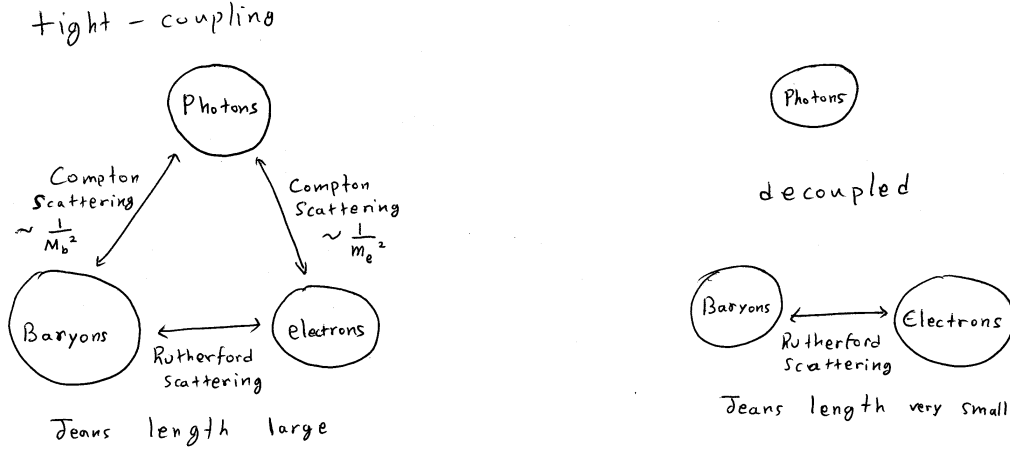


Figure 3.7: Schematic picture of the interaction of photons, baryons and electrons. When the temperature of the Universe was high, photons baryons and electrons were tightly-coupled to each other (Left). During that time the Universe was ionized. As the temperature drops, the photons decouple from the baryons and the electrons (which remain tightly coupled). During that time the Universe is composed of neutral atoms.

early Universe when the temperature was high, baryons were ionized. Thus both baryons and the free electrons were strongly interacting with the photons through Compton scattering. We say that during this period electrons and baryons are tightly-coupled with the photons to give the photon-baryon fluid. Baryons and electrons are also tightly-coupled to each other via Rutherford-Coulomb scattering. In direct contrast with Compton scattering, Rutherford scattering keeps the baryons and photons tightly-coupled during the entire history of the Universe and therefore we may assume that $\delta_b = \delta_e$. So we only need to calculate the evolution of one of them (baryons or electrons) and the other will follow. Now back to Compton scattering. The Compton scattering cross-section is inversely proportional to the square of the mass of the particle involved. Since baryons are at least 2000 times heavier than electrons (for the case of Hydrogen and much more for Helium) we may safely ignore the Compton scattering of baryons and focus on electrons. This means that we will calculate the evolution of electrons and baryons follow. This schematic picture is displayed in figure 3.7.

Now what happens during tight coupling is that the photons (which are relativistic and hence their Jeans length is the horizon) exchange momentum with the electrons (which take the baryons with them). This way the electrons are forced to move at relativistic speeds and so are the baryons. Hence the baryon speed of sound is close to the speed of light and this means that the baryon Jeans length is also close to the horizon. Actually it is slightly smaller than the horizon but still we can safely say that the baryon Jeans length is very large. This in turn means that the baryonic fluid has substantial pressure during that time. After baryons decouple, however, they are no longer disturbed by the photons and so they rapidly cool down and become non-relativistic. Their sound

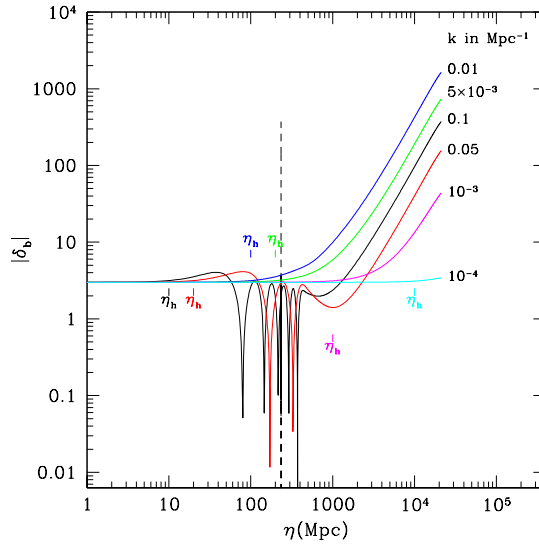


Figure 3.8: The density contrast δ_b for baryons in a Universe containing only photons and baryons. The radiation-matter equality η_{eq} is shown by a vertical dashed line. Horizon crossing is indicated by a vertical line and η_h for each k -mode. We see the effects derived in the lectures: (1) all modes stay constant outside the horizon, (2) modes entering the horizon before decoupling and inside the Jeans length oscillate then grow after decoupling, (3) modes entering the horizon outside the Jeans length grow as a powerlaw in the matter era. Notice how the modes which underwent oscillations have the final growth reduced due to the time lost oscillating (black and red).

speed tends to zero and so is their Jeans length. This means that during this time they start to behave as exactly pressureless matter.

As we have already seen, if matter has a small pressure then the picture of structure growth we described earlier is altered for scales smaller than the Jeans length: once a mode enters the horizon, if $k > k_J$ then δ will undergo damped oscillations which will persist until the Jeans length (which in an expanding Universe is time-dependent) becomes sufficiently small so that $k < k_J$ and δ starts evolving as for the case of pressureless matter. The oscillating phase takes its toll on the final amplitude of δ after the growing period. Since now the time for which the mode can grow is reduced, the final amplitude of δ will be smaller than the case for which no oscillations take place. It turns out that baryons follow this kind of picture. You can see these effects in figure 3.8.

Now for the big question. When we observe rotation curves of galaxies, it looks like that the gravitational field is greatly enhanced compared to the prediction from Newtonian gravity alone. This is inferred by observing the velocity of stars around galaxies. The most popular paradigm to solve this puzzle is that galaxies are immersed on a much bigger "halo" of cold dark matter. So the gravitational potential is enhanced because it is sourced by much more matter which does not interact with light but does interact with gravity. But if this picture is correct then we should see similar effects in different systems, e.g. in cosmology. Observations of large scale structure is one such place where we should expect to see something similar. The question again here is the

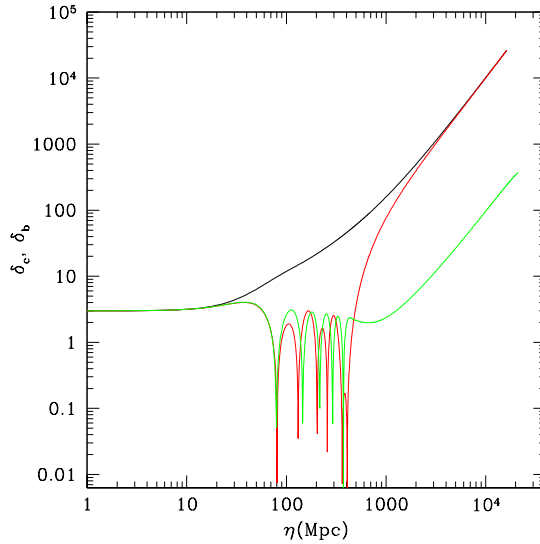


Figure 3.9: The evolution of the density contrast of CDM δ_c and baryons δ_b for $k = 0.1 Mpc^{-1}$ for two different Universes. (1) A Universe with both baryons (red) and CDM (black) (and photons) and a (2) a Universe with only baryons (green) (and photons). Notice how if CDM is present, then baryons (red) after decoupling fall into the potential wells created by CDM and thus δ_b traces δ_c . Thus CDM helps baryons grow as if they did not undergo through tight-coupling.

following. Is the gravitational potential sourced by baryons alone (visible matter) or is there an additional contribution by some new invisible degree of freedom like dark matter?

What we have discussed so far gives us clues on how to answer this question. Baryons undergo oscillations inside their Jeans length and at the same time (a further effect called diffusion or Silk damping further suppresses growth inside the diffusion length and we shall discuss this later) their growth is delayed which reduces the final amplitude for their density contrast δ_b relative to a case with no oscillations. A pressureless fluid like dark matter has a miniscule Jeans length and therefore has no oscillations. The growth due to dark matter will therefore be larger than for baryons. So if we can trace the underlying density contrast using observations then we should be able to distinguish these two cases. If we observe large oscillations in the density field and suppression of growth on small scales then we know that the Universe only has baryons and no dark matter. If we observe very little oscillations (baryons are there so they will leave the oscillation imprint) and no significant suppression of growth on small scales then there should be a sizable component of dark matter present.

Fortunately we have tracers of the underlying density field, in fact trillions of them: galaxies! But since galaxies are made of baryons, this begs the question: are galaxies tracing the baryon density δ_b or the dark matter density (if it exists) δ_c ? The answer is: both! Physically what happens is that if CDM is present then it will start sourcing potential wells Φ at a much earlier time than baryons do. After baryons decouple from the photons, they will fall into these potential wells, and so their density contrast will grow initially faster than a powerlaw until it catches up

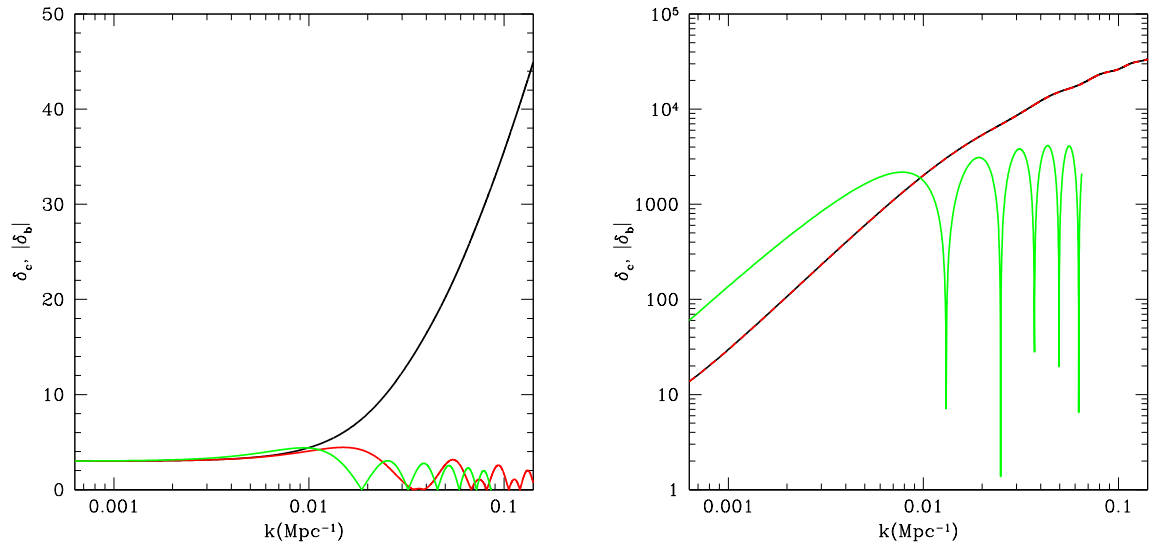


Figure 3.10: Left: The density contrasts of CDM and baryons at decoupling. We display two Universes: one containing only baryons (green curve) and one containing both CDM (black) and baryons (red). At decoupling the CDM is separated from the baryons in both Universes and has already grown substantially more while the baryons were spending their time oscillating.

Right: The density contrasts of CDM and baryons at the present time. We display the same two Universes as on the left. What happens here is that if CDM is present then the baryon density contrast tracks the CDM density contrast and also transfers the oscillations to it (but very reduced). On the contrary in the baryon only case, δ_b is suppressed on small scales and still retains large oscillation pattern. On large scales it appears that it has grown more than the CDM case because the Universe is older for the baryon-only case, thus baryons had more time to grow.

with CDM. After that both CDM and baryons grow with the same powerlaw index and so δ_b traces δ_c . This effect is shown in figure 3.9.

But what about the oscillations? The fact that baryons catch up with CDM has a further effect, this time on CDM. Although their density is much smaller than CDM, when $\delta_c \sim \delta_b$ the baryons will back-react on the potential Φ and will also contribute to its source. Thus the initial oscillation in k will be transferred to the CDM as well. This small effect, called Baryon Acoustic Oscillations (BAO) can be detected and in fact is becoming one of the main observational probes of the matter distribution and of dark energy (more on this later!). On the left of figure 3.10 we see the density contrasts of CDM and baryons at decoupling. Clearly CDM has grown more than baryons and displays no oscillations. On the right of figure 3.10 we see the same models only this time the density contrasts are evaluated at the present time. If the universe contains only baryons, then δ_b is suppressed on small scales and retains its large oscillatory pattern. On the contrary if the Universe contains CDM in addition to baryons, $\delta_b \sim \delta_c$ and the oscillation pattern is transferred to CDM as well (but it's a very small effect as you can see). The fact that δ_b in the baryon-only Universe is higher on large scales than the CDM Universe is because the baryon-only Universe is

about 2.5 times older than the CDM Universe and so baryons had more time to grow. Even then, their growth on small scales is still smaller than the CDM Universe.

In what follows we shall see how to use observations of the distribution of galaxies to get information about the underlying cosmological model.

3.8.2 Observing large scale structure

No matter which Universe we leave in, either baryon only or baryon plus CDM, in both cases we expect that tracing δ_b at the present time, we should be tracing the underlying density field. We have already mentioned that this may be achieved by observing galaxies. But we don't observe the mass density of the galaxy directly. What we do observe is in fact the number density of galaxies in a small patch of the sky, at specific redshift. So we select a patch of the sky at direction $\hat{n}$ and then select galaxies which have similar redshifts and group them together at a common redshift z . We perform this for many different patches of the sky at different directions but always keeping the same redshift z . For each patch we count the number of galaxies and call it $N(z, \hat{n})$. We can then take the average from all patches and construct the average number of galaxies observed at redshift z , $\bar{N}(z)$. Finally construct the galaxy number density fluctuation

$$\delta_N(z, \hat{n}) = \frac{N(z, \hat{n}) - \bar{N}(z)}{\bar{N}(z)}. \quad (3.175)$$

If all the galaxies in the patch are similar morphologically then very likely they have similar masses to some extent. So we can convert the galaxy number density to a galaxy mass density δ_s in redshift space:

$$\delta_s(z, \hat{n}) = b_1 \delta_N(z, \hat{n}) \quad (3.176)$$

where b_1 is a number which may depend on the types of galaxies we are considering. I have denoted the galaxy mass density in redshift-space as $\delta_s(z, \hat{n})$ to distinguish it from the actual galaxy mass density $\delta_g(r, \hat{n})$ in real space (more on this below). The next step is to relate the galaxy mass density δ_s to the underlying density field $\delta(\vec{r})$. This introduces two further problems.

The first problem is that we don't observe the spatial distribution of galaxies at a given time η , i.e. we don't observe $\delta_g(\eta, \vec{r}) = \delta_g(\eta, r, \hat{n})$. Rather we observe only their angular distribution on the sky $\hat{n}$ and a combination of r and η : we only observe galaxies on our past light-cone and so $\eta = r$. This information is encapsulated into the redshift of the galaxy and so we say that we observe $\delta_s(z, \hat{n})$. But then the redshift contains two contributions. One contribution is the cosmological redshift coming from the Hubble expansion, and the other contribution is coming from the fact that galaxies have a peculiar velocity which contributes to their redshift. The net effect of observing in redshift rather than in real space is called a *redshift-space distortion*. It turns out that we can quantify this effect and thus relate $\delta_s(z, \hat{n})$ to $\delta_g(r, \hat{n})$ and in turn to $\delta_g(\vec{k})$ which is the galaxy density contrast in k -space. We shall describe this further below.

The second problem is that galaxies don't simply form from the underlying baryon density field out of the blue. Galaxy formation is a rather complicated process, in fact it is a rather non-linear process which is still not completely understood. This is true even if we are observing scales in the linear regime. Thus galaxies are not expected to trace the underlying density field exactly although it is still expected that overdense regions should contain many more galaxies than underdense regions. To express our ignorance regarding the process of galaxy formation we introduce a new variable called the "bias" b . The bias is assumed to be a constant in its simpler form but more

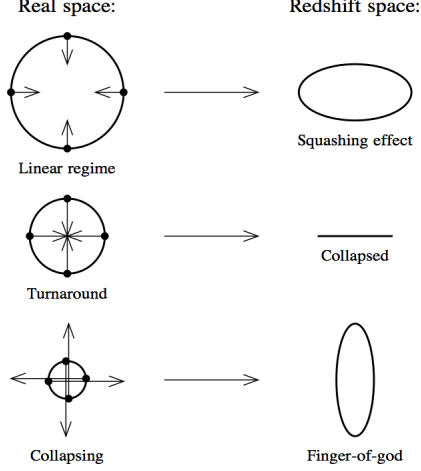


Figure 3.11: Redshift distortions: Far away galaxy groups are just beginning to collapse into bound objects and individual members have a coherent velocity pointing inwards. In redshift space this results to a squashing of the observed group along the line-of-sight. Objects closer to us have already formed bound structures and are virialized. The velocities of individual members are randomly oriented and the effect of averaging gives an overall stretching along the line-of-sight called the "Finger of God". From A. J. Hamilton, "Linear redshift distortions: a review", astro-ph/9708102.

generally (and there is observational evidence for this) it can be a function of scale, $b = b(k)$. We then relate the observed to the underlying density field by

$$\delta_g(\vec{k}) = b(k)\delta(\vec{k}) \quad (3.177)$$

Not let us go back to the first problem: redshift-space distortions. Let us assume that we are observing galaxies at relatively low redshift so that Hubble's law is a very good approximation to the cosmological redshift. Then we have that

$$\bar{z}(\eta) = \bar{z}(r) = H_0|\vec{r}| = H_0r \quad (3.178)$$

where the position of a galaxy in real space is $\vec{r} = r\hat{n}$. The total redshift z is equal to the cosmological one plus the Doppler shift due to the peculiar velocity

$$z = \bar{z} + \delta z = H_0r + \vec{v} \cdot \hat{n} \quad (3.179)$$

The reason that $\delta z = \vec{v} \cdot \hat{n}$ is because only the component of the peculiar velocity along the line-of-sight $\hat{n}$ is contributing to the redshift. Now let us define a redshift-distance $\vec{s}$. The direction of $\vec{s}$ is kept the same as in real space : $\hat{n}$. The magnitude of $\vec{s}$ is $s = |\vec{s}|$ so that $\vec{s} = s\hat{n}$. We define s to be due to the total redshift z in a similar way to Hubble's law

$$z = H_0s \quad (3.180)$$

Then we can relate the redshift distance s to the real distance r as

$$\vec{s} = \left[r + \frac{\delta z}{H_0} \right] \hat{n} = \left[1 + \frac{\vec{v} \cdot \hat{n}}{H_0r} \right] \vec{r} \quad (3.181)$$

Since we are observing galaxies in redshift space rather than real space, the effect is to distort the appearance of galaxies and galaxy clusters compared to their real space distribution, hence the name redshift-space distortions.

Now let us describe what is the effect of redshift-space distortions on galaxies. This is schematically shown in figure 3.11. Consider a spherical region (for instance a cluster of galaxies) in real space which begins to collapse and therefore the velocity field of the object is pointing from all directions towards its centre. Consider now the line-of-sight to the object. The part of the object which is closer to us has a peculiar velocity which points away from us and thus contributes an additional redshift $\delta z > 0$ on top of the cosmological redshift. The part of the object which is further away from us has a peculiar velocity pointing towards us and thus contributes a blueshift $\delta z < 0$ which has the effect of diminishing the total redshift. Finally the parts of the object which are perpendicular to the line-of-sight will not receive any correction to their redshift because the peculiar velocity will be pointing to a direction perpendicular to the line-of-sight. If we now consider the object in redshift space, it will thus appear squashed along the line-of-sight and the squashing factor depends on the peculiar velocity. This is known as the Kaiser effect ⁶. Usually regions which begin to collapse occur in the earlier stages of the Universe and are thus further away from us than regions which have already collapsed and virialized.

The effect of the peculiar velocity on collapsed virialized objects can be different. For such objects, for example galaxy groups which are usually closer to us, the effect of the peculiar velocity is to introduce random corrections to the redshift, due to the random velocities of the galaxies within the group. This has the effect of stretching the observed (in redshift space) galaxy distribution along the line-of-sight: the "Fingers of God" phenomenon (even if the underlying distribution in real space is spherical). The "Fingers of God" is a non-linear phenomenon and we shall not attempt to describe it in detail.

Back to the Kaiser effect. Kaiser realised that the number of galaxies within a volume V remains the same whether observed in real or in redshift space. If $n_s(\vec{s})$ is the number density of galaxies in redshift space and $n_r(\vec{r})$ in real space then we have that

$$n_s d^3 s = n_r d^3 r \quad (3.182)$$

Now $d^3 s = s^2 ds d\Omega_{\hat{n}}$ and $d^3 r = r^2 dr d\Omega_{\hat{n}}$ so that

$$n_s = n_r \frac{r^2}{s^2} \frac{1}{ds/dr} \quad (3.183)$$

$$= n_r \frac{1}{\left(1 + \frac{\vec{v} \cdot \hat{n}}{H_0 r}\right)^2} \frac{1}{1 + \frac{1}{H_0} \frac{\partial}{\partial r}(\vec{v} \cdot \hat{n})} \quad (3.184)$$

It turns out that the correction term due to the derivative i.e. $\frac{1}{H_0} \frac{\partial}{\partial r}(\vec{v} \cdot \hat{n})$ is more important than the correction term $\frac{\vec{v} \cdot \hat{n}}{H_0 r}$ appearing in the above expression. The reason is that if we consider these terms in Fourier space, then a term with the derivative goes as $\sim k \vec{v} \cdot \hat{n}$ so that it is larger than the term without a derivative by a factor kr . Why larger? Because $kr \gg 1$. This is because r is of the order of the size of the survey that is observing the galaxies while k is the wavenumber of the Fourier modes that are being measured by the survey. Only small wavelength (large k) Fourier modes are well measured since only then do we have a large sample of them within the size of the survey. In otherwords, we can only hope to measure those Fourier modes for which $k > \sim 1/r$ so

⁶Nick Kaiser "Clustering in real and in redshift space", Mon.Not.Roy.Astron.Soc. 227, 1 (1987) .

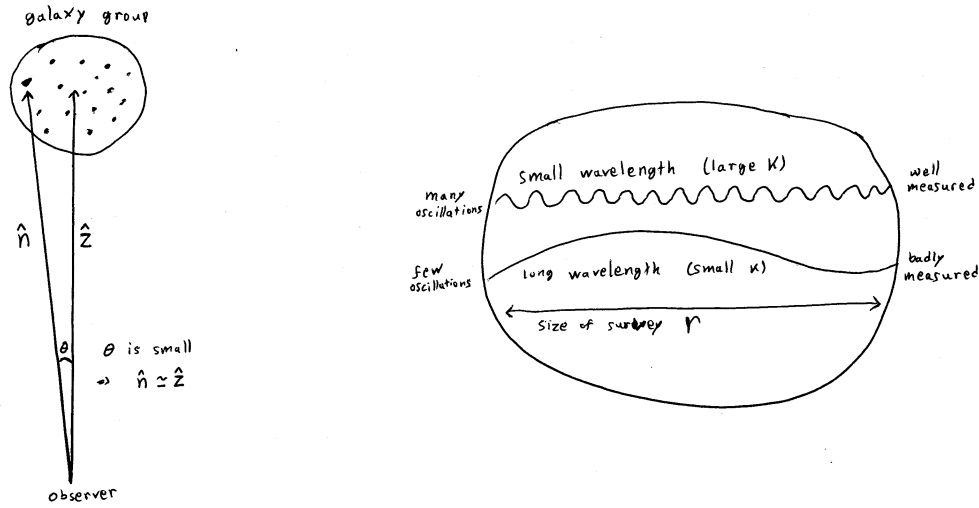


Figure 3.12: Observing galaxies. Left: Observing a distant galaxy group. The vector to an individual member is $\hat{n}$ and the vector to the centre of the group is $\hat{z}$. Since the group is very far away we may assume the distant observer approximation $\hat{n} \approx \hat{z}$.

Right: Observing within a region of size r . Small wavelength (large k) modes are well observed because there are many of them while large wavelength (small k) are not. The size r of the survey sets a limit on the largest wavelengths that can be observed.

that $kr \gg 1$. See figure 3.12. Therefore we can ignore the term $\frac{\vec{v} \cdot \hat{n}}{H_0 r}$ and since $\vec{v}$ is small we get that

$$n_s = n_r \left[1 - \frac{1}{H_0} \frac{\partial}{\partial r} (\vec{v} \cdot \hat{n}) \right] \quad (3.185)$$

Now we need to relate the number densities of the galaxies to the mass densities. We start from (3.175) where we may replace the actual number N with the number density n_s by dividing with the volume of the patch we are observing so that solving for $n_s(z, \hat{n})$ we get

$$n_s(z, \hat{n}) = \bar{n}(z) [1 + \delta_s(z, \hat{n})] \quad (3.186)$$

A similar relation holds in real space which relates $n_r(r, \hat{n}) = n_g(\vec{r})$ with $\delta_g(\vec{r})$:

$$n_r(\vec{r}) = \bar{n}(z) [1 + \delta_g(\vec{r})] \quad (3.187)$$

Relating the two using (3.185) we find that

$$\delta_s(\vec{r}) = \delta_g(\vec{r}) - \frac{1}{H_0} \frac{\partial}{\partial r} (\vec{v} \cdot \hat{n}) \quad (3.188)$$

where we replaced $\delta_s(z, \hat{n})$ with $\delta_s(r, \hat{n})$ because the difference between z and r is small and so is δ . Thus the redshift-space density contrast is equal to the real-space density contrast plus a correction due to the peculiar velocity. Now we Fourier transform $\delta_s(\vec{r})$ to $\delta_s(\vec{k})$. Then $\delta_s(\vec{k})$ is the inverse Fourier transform of $\delta_s(\vec{r})$, i.e.

$$\delta_s(\vec{k}) = \int d^3r e^{-i\vec{k} \cdot \vec{r}} \delta_s(\vec{r}) \quad (3.189)$$

$$= \int d^3r e^{-i\vec{k} \cdot \vec{r}} \left[\delta_g(\vec{r}) - \frac{1}{H_0} \frac{\partial}{\partial r} (\vec{v} \cdot \hat{n}) \right] \quad (3.190)$$

To proceed further, we make the distant observer approximation, that is if the group of galaxies we are observing is very far, then we may take $\hat{n}$ to point exactly to the centre of the group which is at direction $\hat{z}$, rather to individual galaxies. See figure 3.12. Then

$$\delta_s(\vec{k}) = \int d^3r e^{-i\vec{k} \cdot \vec{r}} \left[\delta_g(\vec{r}) - \frac{1}{H_0} \frac{\partial}{\partial r} (\vec{v} \cdot \hat{z}) \right] \quad (3.191)$$

and now $\hat{z}$ is treated as a fixed vector which is not affected by the Fourier transform. We now perform a further Fourier transform, this time on the peculiar velocity $\vec{v}$:

$$\vec{v}(\vec{r}) = \int \frac{d^3k}{(2\pi)^3} e^{i\vec{k} \cdot \vec{r}} \vec{v}(\vec{k}) \quad (3.192)$$

$$= \int \frac{d^3k}{(2\pi)^3} e^{i\vec{k} \cdot \hat{n}} \vec{v}(\vec{k}) \quad (3.193)$$

Acting with $\frac{\partial}{\partial r}$ brings down a term $i\vec{k} \cdot \hat{n}$ which once again assuming the distant observer approximation we let $i\vec{k} \cdot \hat{n} \approx i\vec{k} \cdot \hat{z}$ so that

$$\delta_s(\vec{k}) = \delta_g(\vec{k}) - \frac{1}{H_0} \int d^3r e^{-i\vec{k} \cdot \vec{r}} \int \frac{d^3k'}{(2\pi)^3} e^{i\vec{k}' \cdot \vec{r}} (i\vec{k}' \cdot \hat{z}) \hat{z} \cdot \vec{v}(\vec{k}') \quad (3.194)$$

where we have performed the first integral to recover $\delta_g(\vec{k})$. We may now perform the integral over $\vec{r}$ which gives us a δ function:

$$\delta_s(\vec{k}) = \delta_g(\vec{k}) - \frac{1}{H_0} \int d^3k' (i\vec{k}' \cdot \hat{z}) \hat{z} \cdot \vec{v}(\vec{k}') \delta^{(3)}(\vec{k}' - \vec{k}) \quad (3.195)$$

so that

$$\delta_s(\vec{k}) = \delta_g(\vec{k}) - \frac{i\vec{k} \cdot \hat{z}}{H_0} \hat{z} \cdot \vec{v}(\vec{k}) \quad (3.196)$$

The above equation relates the redshift-space density contrast to the real-space density contrast and an arbitrary velocity field. However, we have seen that peculiar velocities may be treated as irrotational, i.e. they can be written in terms of a scalar mode: $\vec{v} = \vec{\nabla}v$. In Fourier space the relation is $\vec{v} = i\vec{k}v$ so that our expression becomes

$$\delta_s(\vec{k}) = \delta_g(\vec{k}) + \frac{\mu^2}{H_0 k^2} v(\vec{k}) \quad (3.197)$$

where $\mu = \hat{k} \cdot \hat{z}$ is the cosine of the angle between $\vec{k}$ and $\hat{z}$. Now on sub-horizon scales, the velocity may be given in terms of the density contrast and the growth factor $f = \frac{d \ln \delta}{d \ln a}$ via (??) so that introducing the bias factor using (3.177) we find our final expression

$$\delta_s(\vec{k}) = b [1 + \mu^2 \beta] \delta(\vec{k}) \quad (3.198)$$

where

$$\beta = \frac{f}{b} \quad (3.199)$$

We have succeeded in relating the observed galaxy density contrast in redshift space to the underlying density contrast in real space (both Fourier transformed). In doing so we have introduced an additional parameter, the bias b , which is not a part of the cosmological model but models our ignorance about galaxy formation.

To make inferences about the density field $\delta(\vec{k})$ from observations of $\delta_s(\vec{k})$ we need to use statistics. The reason is that $\delta(\vec{k})$ is not a fixed quantity but may vary randomly. Let us first distinguish $\delta(\vec{k})$ from either $\delta_b(k)$ or $\delta_c(k)$ that we find by solving the Einstein and fluid equations as we have already done earlier in the course. First of all notice that I use $\delta(\vec{k})$ rather than $\delta(k)$. The reason is that there is more information in $\delta(\vec{k})$ than in $\delta_b(k)$ (which does not depend on the direction of $\vec{k}$).

The picture is as follows. We start from an initial density field in the early Universe $\delta_0(\vec{r})$ at some initial time η_{in} . This field has a completely unknown spatial dependence and thus should be treated as a random variable. Fourier transforming we form $\delta_0(\vec{k})$ which is also a set of random variables for each vector $\vec{k}$ (so an infinite number of random variables). These are drawn from an unknown probability distribution. As Ed may discuss when he introduces inflation, inflation typically predicts that the probability distribution is very close to Gaussian (observations confirm this). Now we need to propagate the initial random variable $\delta_0(\vec{k})$ to the present time to get $\delta(\vec{k})$. We do so by propagating each individual $\vec{k}$ -mode separately and we may write

$$\delta(\vec{k}) = T(k) \delta_0(\vec{k}) \quad (3.200)$$

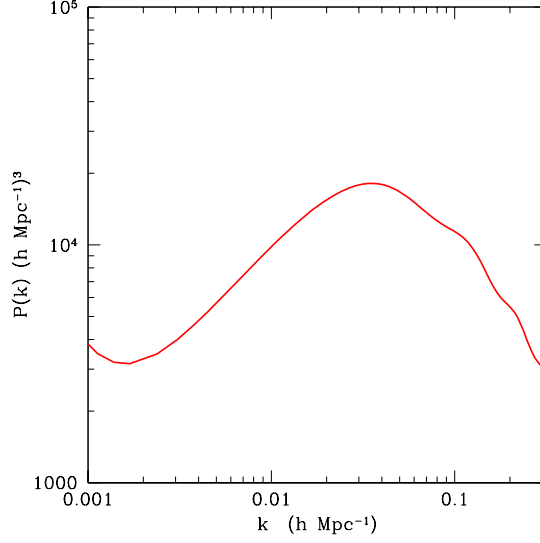


Figure 3.13: The matter power spectrum $P(k)$ of a universe containing baryons and cold dark matter (no cosmological constant) for a scale invariant initial power spectrum ($n = 1$). Observe the baryon acoustic oscillations imprinted on small scales. Normalization is arbitrary.

where $T(k)$ is called the transfer function. So what is this transfer function? It is none other than the solution $\delta_b(k)$ or $\delta_c(k)$ that we have already found! More precicely

$$T(k) = \Omega_b \delta_b(k) + \Omega_c \delta_c(k) \quad (3.201)$$

Since $\delta(\vec{k})$ (or $\delta_0(\vec{k})$) are random variables, we need to use statistics to describe them. The simplest thing we can do is to create the 2-point correlation between different $\vec{k}$ -modes (the 1-point correlation is just the average which is by definition zero). The two-point function is related to a function $P(k)$ (no direction dependence) as

$$\langle \delta(\vec{k}) \delta(\vec{k}') \rangle = (2\pi)^3 P(k) \delta^{(3)}(\vec{k} - \vec{k}') \quad (3.202)$$

The function $P(k)$ is called the power spectrum. Similarly we may compute the power spectrum of the initial distribution $\delta_0(\vec{k})$:

$$\langle \delta_0(\vec{k}) \delta_0(\vec{k}') \rangle = (2\pi)^3 P_0(k) \delta^{(3)}(\vec{k} - \vec{k}') \quad (3.203)$$

where now $P_0(k)$ is called the initial power spectrum (as given by inflation for example). The two power spectra are then related via the transfer function as

$$P(k) = P_0(k) |T(k)|^2 \quad (3.204)$$

Generically $P_0(k)$ is an arbitrary unknown function which encapsulates our ignorance of initial conditions. A theory of initial conditions should be able to predict precicely what its form is. Inflationary theories predict a rather simple form for $P_0(k)$. It is given as a powerlaw

$$P_0(k) = A_0 k^n \quad (3.205)$$

We call A_0 the initial amplitude of the perturbations and n the spectral index. The value of A_0 is measured to be around 10^{-12} . The special case $n = 1$ is of particular significance. It is called the Harrison-Zel'dovich scale-invariant spectrum. If the initial power spectrum has this form (i.e. $n = 1$) then the fluctuations have equal power at every scale (they are scale-invariant). This will be discussed further when Ed considers inflation so be patient. Observations show that $n \approx 0.95$. The power spectrum of the baryon + CDM model we discussed earlier in the course is shown in figure 3.13 for a scale-invariant initial power spectrum.

The goal now is to relate the power spectrum in real space, i.e. $P(k)$ to the power spectrum in redshift space. We use (3.198) and take the 2-point function of both the galaxy redshift-space distribution δ_s and the density field δ . To do so we need to calculate things like $\langle \mu^2 \rangle = \frac{1}{3}$ and $\langle \mu^4 \rangle = \frac{1}{5}$. The final answer is

$$P_s(k) = b^2 \left[1 + \frac{2\beta}{3} + \frac{\beta^2}{5} \right] P(k) \quad (3.206)$$

and we are done. Given a cosmological model we can calculate the transfer function $T(k)$, compute $P(k)$ assuming an initial power spectrum and finally compute $P_s(k)$ by supplying the bias b and the growth factor f . We then compare $P_s(k)$ with galaxy observations and test the theory.

Before finishing we mention a few more facts without details. First, just like we can have power spectra of the galaxy-galaxy 2-point function, we may also have power spectra of the correlation between galaxy and velocity ($P_{gv}(k)$) as well as the 2-point function of the velocity field $P_{vv}(k)$. Both of these can also be measured and provide additional and complementary information on the underlying cosmological model. Furthermore, we may also construct 3-point functions or 4-point functions etc. If the initial power spectrum is Gaussian, it may be shown that these do not provide any additional information than the 2-point function (in fact the odd-point functions vanish for Gaussian probability distributions). Thus measuring these n -point functions may provide us with information about the statistical distribution of the initial density fluctuation, whether it is Gaussian or not. This is currently a very active and popular field of research.

4 The Cosmic Microwave Background

The Cosmic Microwave Background radiation is one of the cleanest probes of the Universe. It consists of thermalized radiation, a remnant echo from the Big Bang. The CMB was discovered accidentally by Penzias and Wilson in 1964 while looking for microwave radiation from the galaxy. It is perhaps the most important discovery concerning cosmology since Edwin Hubble demonstrated the expansion of the Universe. Penzias and Wilson, who received the Nobel Prize in Physics in 1978 for their astonishing discovery, detected a small dull "noise" in their antenna which looked like it was the same in every direction they pointed and was oblivious to whether it was day or night.

The current temperature of the CMB is $2.725K$ as measured by the COBE Firas instrument and corresponds to microwave radiation. We are constantly bombarded by around 10 trillion CMB photons per second per square centimeter. In fact, the CMB contributes a few percent of TV "snow". As we shall see further below, the CMB is not completely isotropic but it has small ripples, anisotropies, which are only one part in 100 000. The first measurement of these anisotropies by the COBE satellite gave Mather and Smoot the Nobel prize in 2006 (2nd one to the CMB).

4.1 Photons in the Universe

4.1.1 The formation of the CMB and its spectrum

To understand when the CMB was formed let us briefly recap the history of the early Universe as is shown on figure 4.1. We think that there was a period of exponential expansion of the Universe what we call inflation. We don't know exactly when inflation ended but it must have been at a very high energy scale, e.g. close to the grand unification scale and definitely above the electroweak scale. During the electroweak era, the electromagnetic and weak forces were unified into a single force: the electroweak force. Thus during this time photons did not even exist. As the Universe cooled down, the electroweak symmetry broke and the three linear combinations of the electroweak gauge bosons became the massive weak bosons while a fourth linear combination gave rise to a new massless particle of spin 1: the photon. Thus, photons and with them electromagnetism was formally created at the end of the electroweak era, around $t = 10^{-12}s$. However, photons did not come to power immediately but had to wait their turn for three more eras.

The Universe went through the quark era, where the Universe was dominated by free quarks until $t = 10^{-6}s$ followed by the hadron era when quarks got confined into hadrons forever. The hadron era came to an end at about $t = 1s$, when all hadrons annihilated leaving only free protons and free neutrons (the neutron lifetime is about $15min$ which when compared to a few seconds means it is pretty much a stable particle). At that point, a new class of particles came to dominate: the leptons. Leptons consist of the electron, muon and tauon and their corresponding anti-particles as well as their corresponding neutrinos and anti-neutrinos. The lepton era also came to pass at around $t = 100s$ when the last surviving leptons, electrons and positrons, annihilate, leaving a tiny fraction of electrons (to match the protons and keep overall neutrality of the Universe) and at the same time making a billion new friends: the photons. This is the first time in the history of the Universe that photons come to dominate the background energy density and it is here that the CMB was initially formed.

As we have already mentioned, the CMB spectrum is thermal with a current temperature $T_0 = 2.725K$. This means that the intensity of CMB radiation has a Planck spectrum, i.e. the CMB intensity $I(\nu)$ at a frequency ν is

$$I(\nu) = \frac{4\pi\nu^3}{e^{2\pi\nu/T} - 1} \quad (4.1)$$

The intensity measures the energy of photons, per unit area per unit time, i.e. power per unit area. Integrating the intensity over all solid angles and over all frequencies gives the Stefan-Boltzmann law

$$P_{CMB} = \sigma T^4 \quad (4.2)$$

which relates the total power of the CMB to the fourth power of the temperature, where $\sigma = \frac{2\pi^5}{15}$ is the Stefan-Boltzmann constant. It turns out that the expansion of the Universe preserves a Planckian spectrum. Thus, since we observe the CMB to have a Planckian spectrum today, then it always had a Planckian spectrum. This is the best evidence we have that the Universe has been at thermal equilibrium all the way up until the creation of the CMB at the end of the lepton era.

4.1.2 Kinetic theory and the CMB

To describe the CMB we use Kinetic theory. In fact, you have already done that when you learned about the thermal history of the Universe. Since photons are bosons, they obey a Bose-Einstein

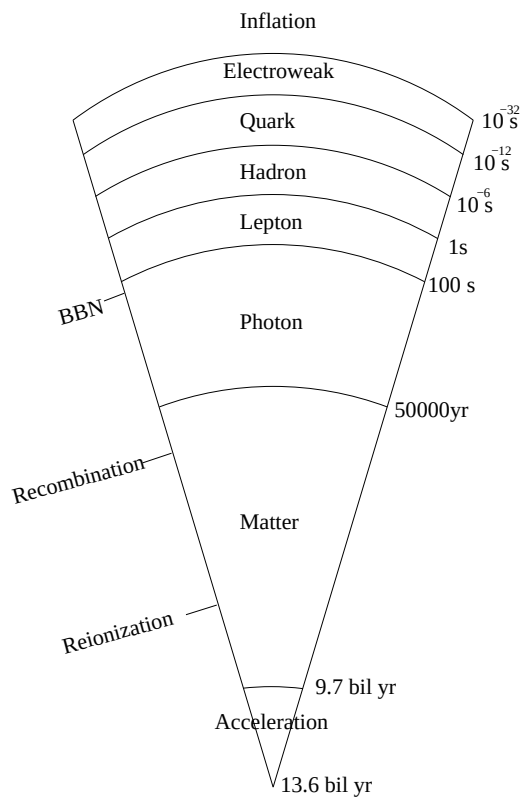


Figure 4.1: Brief history of the particle eras in the early Universe. The CMB is created around 100s after inflation at the end of the lepton era, when electrons and positrons annihilate. This is the first time that photons dominate the Universe.

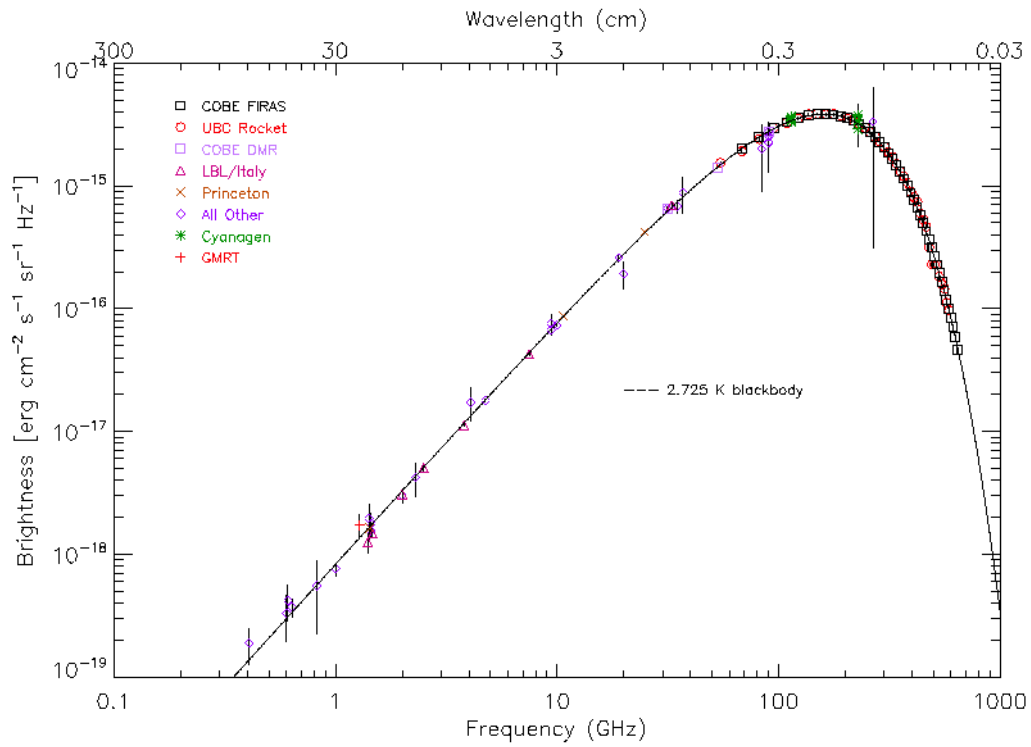


Figure 4.2: The CMB spectrum measure across a wide range of frequencies with unrepresented accuracy. The CMB spectrum is the best example of a Planckian spectrum in the Universe (better than the sun).

distribution function given by (2.1) which we rewrite again here in natural units and setting at the same time the chemical potential to zero, the degeneracy to 2 (photons have two polarizations) and using the fact that photons are massless so that $E = p$. The final distribution is

$$\bar{f}(t, p) = \frac{2}{(2\pi)^3} \frac{1}{\exp[p/T] - 1} \quad (4.3)$$

The bar on $\bar{f}$ denotes the fact that this is the background Friedmannian distribution function. Notice that $\bar{f}$ depends only on time and the *magnitude* of the momentum p but does not depend on the spatial position $\vec{r}$ nor the direction of momentum. Expressing the momentum in terms of a photon's frequency by $p = 2\pi\nu$ leads directly to the Planckian spectrum (4.1). Let's refresh what the distribution function can do.

The distribution function contains all available information regarding photons. From it we can get things like the average photon energy density, photon velocity or photon pressure in the Universe. To get these quantities we multiply the distribution function by the quantity of interest and then integrate over all momenta. For instance, let's calculate the average energy density of photons. The energy of a single photon is $E = p$ because photons are massless. Thus the average energy density is

$$\bar{\rho}_\gamma = \frac{2}{(2\pi)^3} \int d^3p \frac{p}{\exp[p/T] - 1} \quad (4.4)$$

$$= \frac{2}{(2\pi)^3} \int dp d\Omega_p \frac{p^3}{\exp[p/T] - 1} \quad (4.5)$$

$$= \frac{1}{\pi^2} \int dp \frac{p^3}{\exp[p/T] - 1} \quad (4.6)$$

$$= \frac{1}{\pi^2} T^4 \int_0^\infty dx \frac{x^3}{e^x - 1} \quad (4.7)$$

$$= \frac{\pi^2}{15} T^4 \quad (4.8)$$

$$= \frac{\pi^2 T_0^4}{15} \frac{1}{a^4} \quad (4.9)$$

Similarly, the pressure is evaluated as

$$\bar{P}_\gamma = \frac{2}{(2\pi)^3} \int d^3p \frac{p^2}{3E} \frac{1}{\exp[p/T] - 1} \quad (4.10)$$

$$= \frac{1}{3} \bar{\rho}_\gamma \quad (4.11)$$

The above integrals giving $\bar{\rho}$ and $\bar{P}$ are actually special cases of a more general relation. Given $\bar{f}$ we can calculate the full energy-momentum tensor as follows

$$T^\mu_\nu = \frac{2}{(2\pi)^3} \int d^3p \frac{p^\mu p_\nu}{E} \frac{1}{\exp[p/T] - 1} \quad (4.12)$$

letting $\mu = \nu = 0$ you can "re-derive" the expression for $\bar{\rho}$ while letting $\mu = \nu = i$ the can get the expression for $\bar{P}$.

How about things like T^0_i ? This would give us an average velocity but we already expect that this velocity has to be zero in a Friedmann universe because of isotropy and homogeneity. Inserting

$\mu = 0$ and $\nu = i$ in (4.12) we have to do an integral over a photon direction $\hat{n}$ and since the distribution function does not depend on $\hat{n}$ we are left with the angular integral $\int d\Omega_{\hat{n}} \hat{n}$ which evaluates to zero.

Now the energy momentum tensor obeys a conservation law: $\nabla_{\mu} T^{\mu}_{\nu} = 0$. For the case of the Friedmann background this leads to

$$\dot{\bar{\rho}}_{\gamma} + 4H\bar{\rho}_{\gamma} = 0 \quad (4.13)$$

But both T^{μ}_{ν} and $\bar{\rho}_{\gamma}$ are obtained from the distribution function via (4.12) so how is this consistent with the conservation law? The answer is that the distribution function obeys a differential equation called the Boltzmann equation. For the Friedmann Universe the Boltzmann equation is

$$\frac{\partial \bar{f}}{\partial t} - Hp \frac{\partial \bar{f}}{\partial p} = 0 \quad (4.14)$$

Notice that the Boltzmann equation is not only a differential equation with respect to time but also with respect to momentum. This was to be expected as the distribution function depends on both t and p . The Boltzmann equation then implies conservation of energy-momentum $\nabla_{\mu} T^{\mu}_{\nu} = 0$.

Finally let's evaluate the Boltzmann equation for the Bose-Einstein distribution function of photons. Taking the derivative with time we get (we ignore the factor $\frac{2}{(2\pi)^3}$ as it will cancel out in the Boltzmann equation)

$$\frac{\partial \bar{f}}{\partial t} = \frac{\exp[p/T]}{(\exp[p/T] - 1)^2} \frac{p}{T^2} \frac{\partial T}{\partial t} \quad (4.15)$$

while taking the derivative wrt to momentum

$$\frac{\partial \bar{f}}{\partial p} = -\frac{\exp[p/T]}{(\exp[p/T] - 1)^2} \frac{1}{T} \quad (4.16)$$

Putting things together we find that in a Friedmann Universe the Boltzmann equation is equivalent to a differential equation for the temperature T :

$$\frac{\partial T}{\partial t} + HT = 0 \quad (4.17)$$

We can then solve this equation to get

$$T = \frac{T_0}{a} \quad (4.18)$$

which is the familiar expression for the radiation temperature.

4.2 CMB anisotropies

So far we have been dealing with a completely isotropic CMB. We have mentioned in the beginning of this chapter, however, that the CMB is anisotropic to 1 part in 100000. This is a very small anisotropy but by studying it we can get a better understanding of the physical processes in the early Universe. By detecting it we can greatly constrain the cosmological model and bring it in line with observations.

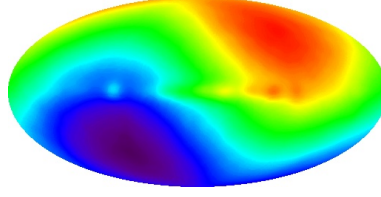


Figure 4.3: The CMB dipole anisotropy. Red means hotter than $2.725K$ and blue means cooler.

4.2.1 Quantifying the anisotropies

To observe the CMB anisotropy we have to observe photons coming from different parts of the sky. By observing many photons from the same part of the sky at a variety of frequencies we can obtain a distribution of photon energies. The intensity of the photons will then give us a temperature. If this photon temperature is different at different parts of the sky then we will know that the CMB is anisotropic. So what is the $2.725K$ that we have attributed to the CMB? It is the average temperature which is found by averaging the temperature of the photons from different parts of the sky. So suppose that the photon temperature we observe today in direction $\hat{n}$ is $T(\hat{n})$. The average temperature is

$$T_0 = \langle T(\hat{n}) \rangle = \frac{1}{4\pi} \int d\Omega_{\hat{n}} T(\hat{n}) \quad (4.19)$$

In the same way we formed the density contrast $\delta = \frac{\delta\rho}{\rho}$ we can do the same with the photon temperature and form the temperature anisotropy in direction $\Theta(\hat{n})$

$$\Theta(\hat{n}) = \frac{T(\hat{n}) - T_0}{T_0} \quad (4.20)$$

Clearly then the average of the temperature anisotropy vanishes: $\langle \Theta(\hat{n}) \rangle = 0$.

The average temperature T_0 is called the temperature monopole and what we have done in (4.20) is to subtract this monopole. We will now focus on $\Theta(\hat{n})$.

It turns out that $\Theta(\hat{n})$ has a dipole contribution which is shown in figure 4.3. The physical reason for having a CMB dipole anisotropy is due to our motion with respect to the rest frame of the CMB, more precisely due to the motion of our galaxy with respect to the CMB rest frame. The CMB dipole is thus due to the Doppler shift experienced by photons. As we move through the CMB, observing photons coming opposite to our motion will give them a blueshift resulting in a slight increase of their temperature, while observing photons from the same direction as our motion will give them a redshift resulting in a slight decrease of their temperature. The CMB dipole is about 1000 times smaller than the monopole and corresponds to the galaxy moving at a speed of $627 \pm 22 km/s$ in the direction of galactic longitude $l = 276 \pm 3^\circ$ and $b = 30 \pm 3^\circ$.

Subtracting the dipole from $\Theta(\hat{n})$ leaves further anisotropies which cannot be due to our motion. What is left are intrinsic anisotropies due to the various interactions experienced by the photons during their travel in time and space. It is this part which is only 1 part in 100000 and the latest observations of it by the WMAP satellite is shown in figure 4.4. This is the interesting part of the CMB anisotropy and we will devote the rest of the CMB part of the course to studying it in more detail.

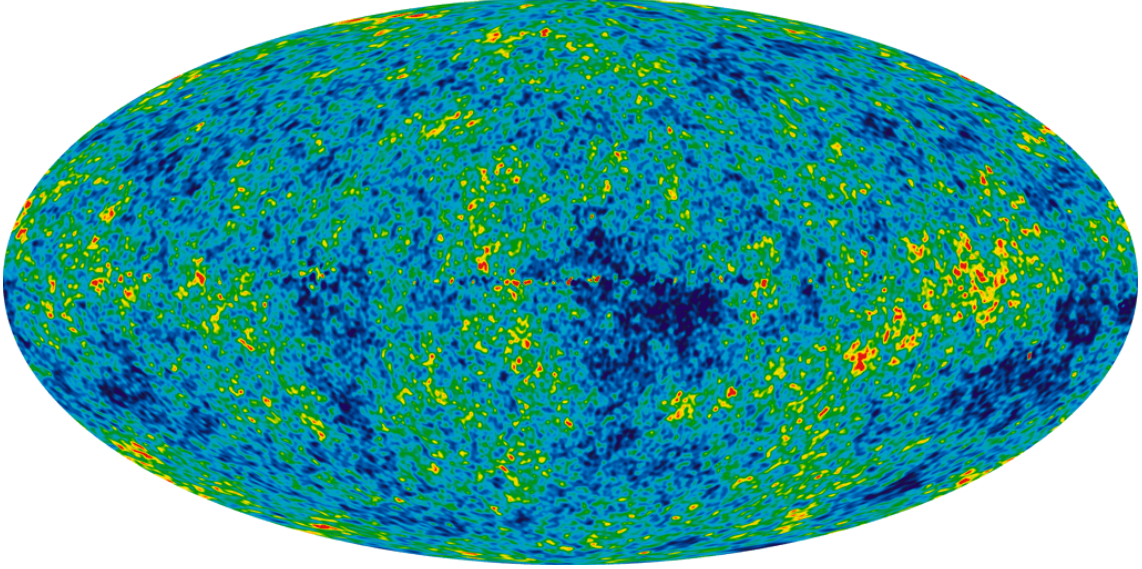


Figure 4.4: The CMB sky as seen by the WMAP satellite after 7 years of observation. Both the monopole and the dipole are subtracted but also further contamination due to astrophysical processes occurring in the galaxy (the so called foregrounds). Red means hotter than $2.725K$ and blue means cooler.

4.2.2 The CMB sky and expansion in angular eigenmodes

The CMB temperature anisotropy map is a rather complicated pattern. To extract useful information from it we need to simplify somehow. There is a neat way of doing so. Rather than using the temperature anisotropy directly $\Theta(\hat{n})$ (which varies continuously on the sky as $\hat{n}$ varies) we can use an infinite set of *discrete moments* that we call $a_{\ell m}$. The continuous variation is taken up by a set of special functions called spherical harmonics.

We express the temperature variation as the following sum over the complex coefficients $a_{\ell m}$ and the spherical harmonics $Y_{\ell m}(\hat{n})$:

$$\Theta(\hat{n}) = \sum_{\ell m} a_{\ell m} Y_{\ell m}(\hat{n}) \quad (4.21)$$

with inverse

$$a_{\ell m} = \int d\Omega_{\hat{n}} \Theta(\hat{n}) Y_{\ell m}^*(\hat{n}) \quad (4.22)$$

The coefficients $a_{\ell m}$ are constants. The index ℓ takes values from $0, 1, 2 \dots \infty$ while the index m takes values from $-\ell, -\ell + 1, \dots, -1, 0, 1, \dots, \ell - 1, \ell$. In doing this transform we have isolated all information about the temperature anisotropies into the $a_{\ell m}$ coefficients while the continuous variation with angle is taken up by a set of known functions which can be pre-calculated on a computer. Moreover, the spherical harmonics obey nice mathematical relations which can make tedious integrations tractable. Another way of seeing this expansion is that it is the analogue of an angular Fourier transform where the Fourier coefficients are now the $Y_{\ell m}$'s rather than $e^{i\vec{k} \cdot \vec{x}}$.

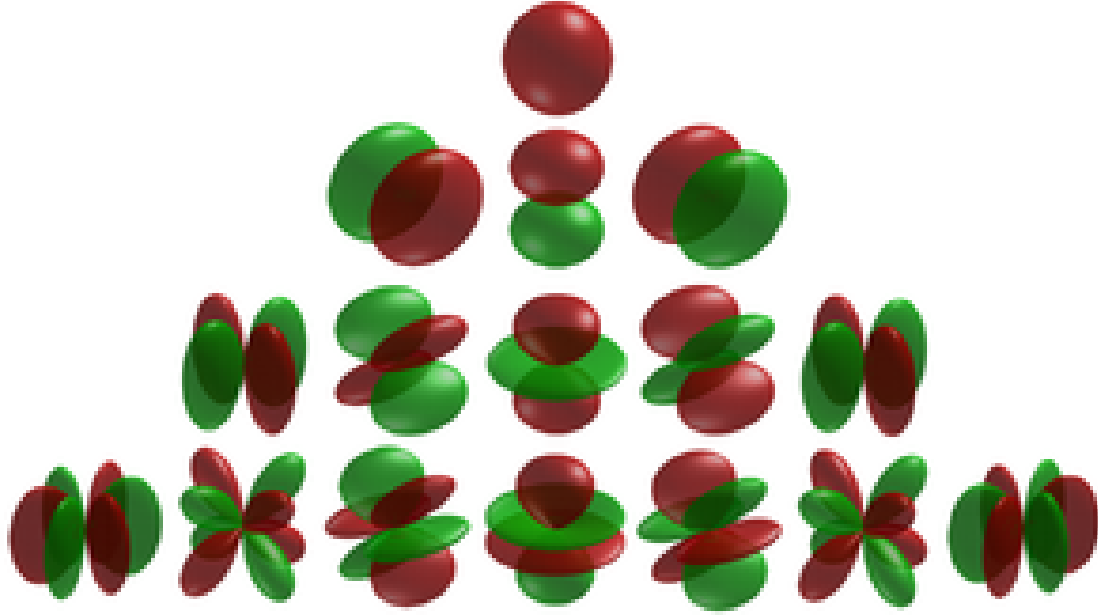


Figure 4.5: Graphical representation of the first few spherical harmonics. Top to bottom Y_{00} (monopole), Y_{1m} for $m = -1 \dots 1$ (dipole), Y_{2m} for $m = -3 \dots 3$ (quadrupole) and Y_{3m} for $m = -5 \dots 5$ (octopole).

4.2.3 Special functions: Spherical Harmonics

You have encountered spherical harmonics as the angular wavefunctions of the Schroedinger equation in the case of the hydrogen atom. But they are more general than the hydrogen atom. They offer the means to expand functions which depend on the angular variables θ and ϕ .

They are determined as solutions from the angular part of the Laplace operator:

$$\left[\frac{1}{\sin \theta} \frac{\partial}{\partial \theta} \left(\sin \theta \frac{\partial}{\partial \theta} \right) + \frac{1}{\sin^2 \theta} \frac{\partial^2}{\partial \phi^2} \right] Y_{\ell m}(\theta, \phi) = -\ell(\ell + 1) Y_{\ell m}(\theta, \phi) \quad (4.23)$$

The index ℓ takes values from the set $\{0, 1, 2, \dots\}$ and m in the set $\{-\ell, -\ell + 1, \dots, 0, \dots, \ell - 1, \ell\}$. Note further that the spherical harmonics have in general complex values and their ϕ dependence comes always as $e^{im\phi}$.

The first few Spherical Harmonics are

$$Y_{00}(\theta, \phi) = \frac{1}{\sqrt{4\pi}} \quad \text{"}s\text{-state"} \quad (4.24)$$

$$Y_{10}(\theta, \phi) = \sqrt{\frac{3}{4\pi}} \cos \theta \quad \text{"}p\text{-state"} \quad (4.25)$$

$$Y_{1,\pm 1}(\theta, \phi) = \mp \sqrt{\frac{3}{8\pi}} \sin \theta e^{\pm i\phi} \quad \text{"}p\text{-state"} \quad (4.26)$$

$$Y_{20}(\theta, \phi) = \sqrt{\frac{5}{16\pi}} (3 \cos^2 \theta - 1) \quad \text{"}d\text{-state"} \quad (4.27)$$

$$Y_{2,\pm 1}(\theta, \phi) = \mp \sqrt{\frac{15}{8\pi}} \cos \theta \sin \theta e^{\pm i\phi} \quad \text{"}d\text{-state"} \quad (4.28)$$

$$Y_{2,\pm 2}(\theta, \phi) = \sqrt{\frac{15}{32\pi}} \sin^2 \theta e^{\pm 2i\phi} \quad \text{"}d\text{-state"} \quad (4.29)$$

As we have mentioned above, we can expand any function of θ and ϕ as a series of spherical harmonics. We shall also find it convenient to collect θ and ϕ into a single variable $\hat{\omega} = \{\theta, \phi\}$.

$$f(\hat{\omega}) = f(\theta, \phi) = \sum_{\ell m} f_{\ell m} Y_{\ell m}(\hat{\omega}) \quad (4.30)$$

where here and thereafter we write $\sum_{\ell m}$ to mean $\sum_{\ell=0}^{\infty} \sum_{m=-\ell}^{\ell}$. The variables $f_{\ell m}$ are the coefficients of the series and are the analogue of the Fourier coefficients only now we are dealing with an angular transform. The inverse relation is

$$f_{\ell m} = \int d^2 \hat{\omega} Y_{\ell m}^*(\hat{\omega}) f(\hat{\omega}) \quad (4.31)$$

where $d^2 \hat{\omega} = \sin \theta d\theta d\phi$ is the two dimensional volume element.

The spherical harmonics obey the orthonormality relations

$$\int d^2 \hat{\omega} Y_{\ell m}^*(\hat{\omega}) Y_{\ell' m'}(\hat{\omega}) = \delta_{\ell \ell'} \delta_{m m'} \quad (4.32)$$

and

$$\sum_{\ell m} Y_{\ell m}^*(\hat{\omega}) Y_{\ell m}(\hat{\omega}') = \delta^{(2)}(\hat{\omega} - \hat{\omega}') = \frac{\delta^{(\theta)}(\theta - \theta') \delta^{(\phi)}(\phi - \phi')}{\sin \theta} \quad (4.33)$$

Note that since θ and ϕ are angular variables which take values in $[0, \pi)$ and $[0, 2\pi)$ respectively, the Dirac δ -functions have support in the same range

$$\int_0^\pi \delta^{(\theta)}(\theta - \theta') d\theta' = 1 \quad (4.34)$$

$$\int_0^{2\pi} \delta^{(\phi)}(\phi - \phi') d\theta' = 1 \quad (4.35)$$

Likewise the 2-dimensional angular Dirac δ -function has support on the sphere i.e. $\int \delta^{(2)}(\hat{\omega} - \hat{\omega}') d^2 \hat{\omega} = 1$

4.2.4 Special functions: Legendre Polynomials

The 2nd kind of special functions we will frequently use are the Legendre Polynomials $P_\ell(\mu)$ where once again $\ell = \{0, 1, 2, \dots\}$. The continuous variable μ takes values on the unit circle, i.e. in the range $[-1, 1]$. The Legendre polynomials are closely related to the spherical harmonics. They are used when expanding in an angular variable in one dimension or in an axisymmetric situation in two dimensions.

The Legendre polynomials are solutions to the differential equation

$$(1 - \mu^2) \frac{d^2 P_\ell}{d\mu^2} - 2\mu \frac{dP_\ell}{d\mu} + \ell(\ell + 1)P_\ell = 0 \quad (4.36)$$

The first few are

$$P_0(\mu) = 1 \quad (4.37)$$

$$P_1(\mu) = \mu \quad (4.38)$$

$$P_2(\mu) = \frac{1}{2}(3\mu^2 - 1) \quad (4.39)$$

$$P_3(\mu) = \frac{1}{2}(5\mu^3 - 3\mu) \quad (4.40)$$

$$P_4(\mu) = \frac{1}{8}(35\mu^4 - 30\mu^2 + 3) \quad (4.41)$$

In general we also have $P_\ell(\mu) = (-1)^\ell P_\ell(-\mu)$ and all of them obey $P_\ell(1) = 1$. The general form of the Legendre polynomial of order ℓ may be obtained from

$$P_\ell = \frac{1}{2^\ell \ell!} \frac{d^\ell}{d\mu^\ell} (\mu^2 - 1)^\ell \quad (4.42)$$

A function $f(\mu)$ is expanded in Legendre polynomials as

$$f(\mu) = \sum_{\ell} i^\ell (2\ell + 1) f_\ell P_\ell(\mu) \quad (4.43)$$

with inverse

$$f_\ell = \frac{(-i)^\ell}{2} \int_{-1}^1 d\mu P_\ell(\mu) f(\mu) \quad (4.44)$$

The Legendre polynomials obey the orthogonality relations

$$\int d\mu P_\ell(\mu) P_{\ell'}(\mu) = \frac{2}{2\ell + 1} \delta_{\ell\ell'} \quad (4.45)$$

and

$$\sum_{\ell} \frac{2\ell + 1}{2} P_\ell(\mu) P_\ell(\mu') = \delta^{(\mu)}(\mu - \mu') \quad (4.46)$$

Since μ takes values in $[-1, 1]$ then the Dirac δ -function has support in the same range: $\int_{-1}^1 \delta^{(\mu)}(\mu - \mu') = 1$.

We have a number of recurrence relations between them

$$P_\ell = \frac{\mu}{2\ell + 1} [\ell P_{\ell-1} + (\ell + 1) P_{\ell+1}] \quad (4.47)$$

and their derivatives

$$\frac{dP_\ell}{d\mu} = \frac{\ell}{1-\mu^2} (P_{\ell-1} - \mu P_\ell) \quad (4.48)$$

Further relations are

$$\int d\mu \mu P_\ell(\mu) P_{\ell'}(\mu) = \frac{2}{2\ell+1} \left[\frac{\ell}{2\ell-1} \delta_{\ell',\ell-1} + \frac{\ell+1}{2\ell+3} \delta_{\ell',\ell+1} \right] \quad (4.49)$$

and

$$\begin{aligned} \int d\mu \mu^2 P_\ell(\mu) P_{\ell'}(\mu) &= \frac{2}{2\ell+1} \left[\frac{\ell(\ell-1)}{(2\ell-3)(2\ell-1)} \delta_{\ell',\ell-2} \right. \\ &\quad \left. + \frac{2\ell^2+2\ell-1}{(2\ell-1)(2\ell+3)} \delta_{\ell',\ell} + \frac{(\ell+1)(\ell+2)}{(2\ell+3)(2\ell+5)} \delta_{\ell',\ell+2} \right] \end{aligned} \quad (4.50)$$

4.2.5 Relations between the spherical harmonics and the Legendre polynomials

The spherical harmonics and Legendre polynomials could not look more different. Yet, they are very much related as it turns out.

The simplest relation is a formula connecting Legendre polynomials and spherical harmonics with $m = 0$. We have that

$$Y_{\ell 0}(\theta, \phi) = \sqrt{\frac{2\ell+1}{4\pi}} P_\ell(\cos \theta) \quad (4.51)$$

Further, more general relations between the Legendre polynomials and spherical harmonics also exist. If $\hat{n}$ and $\hat{n}'$ are two direction unit vectors, i.e. $\hat{n} = \hat{n}(\hat{\omega}) = (\sin \theta \cos \phi, \sin \theta \sin \phi, \cos \theta)$ and similarly for $\hat{n}'$ then

$$P_\ell(\hat{n} \cdot \hat{n}') = \frac{4\pi}{2\ell+1} \sum_m Y_{\ell m}^*(\hat{n}') Y_{\ell m}(\hat{n}) \quad (4.52)$$

where we have used $\hat{n}$ rather $\hat{\omega}$ in the argument of $Y_{\ell m}$. This is not completely correct as $\hat{n}$ and $\hat{\omega}$ are two completely different objects ($\hat{n}$ is a unit 3-vector while $\hat{\omega} = \{\theta, \phi\}$) but this introduces considerable simplicity without (hopefully) any confusion and we will abuse this notation throughout the course. We shall also use $d\Omega_{\hat{n}} = d^2\hat{\omega} = \sin \theta d\theta d\phi$ in a similar kind of abusive notation.

A further relation between the Legendre polynomials and spherical harmonics is the integral

$$\int d\Omega_{\hat{n}'} \int d\Omega_{\hat{n}} Y_{\ell' m'}^*(\hat{n}') Y_{\ell m}(\hat{n}) P_L(\hat{n} \cdot \hat{n}') = \frac{4\pi}{2L+1} \delta_{L\ell} \delta_{L\ell'} \delta_{mm'} \quad (4.53)$$

4.2.6 The angular power spectrum

Enough of math, back to the CMB.

To make inferences on the cosmological model based on observations of the CMB anisotropies, we need to use statistics. The temperature anisotropy in direction $\hat{n}$ is considered to be a random variable. Since the average of $\Theta(\hat{n})$ is zero by definition we need to use the next available statistical moment: the 2-point correlation function. We correlate the temperature anisotropy in direction $\hat{n}$ with the temperature anisotropy in direction $\hat{n}'$ and by varying $\hat{n}$ and $\hat{n}'$ over all possible directions we form the correlation

$$C(\hat{n}, \hat{n}') = \langle \Theta(\hat{n}) \Theta(\hat{n}') \rangle \quad (4.54)$$

We now use the expansion into spherical harmonics to get

$$C(\hat{n}, \hat{n}') = C(\hat{n} \cdot \hat{n}') = \sum_{\ell m} \sum_{\ell' m'} Y_{\ell m}(\hat{n}') Y_{\ell' m'}^*(\hat{n}') \langle a_{\ell m} a_{\ell' m'}^* \rangle \quad (4.55)$$

Now the correlation of the $a_{\ell m}$'s satisfies

$$\langle a_{\ell m} a_{\ell' m'}^* \rangle = C_\ell \delta_{\ell \ell'} \delta_{m m'} \quad (4.56)$$

Let's explain why the above relation has the form it has. As we have mentioned $\Theta(\hat{n})$ is a random variable. If we express it in $a_{\ell m}$ then the $a_{\ell m}$ coefficients become random variables too. First of all remember that $a_{\ell m}$ is complex, except the special case a_{00} which is real. Then the real part $Re[a_{\ell m}]$ and imaginary part $Im[a_{\ell m}]$ for each $a_{\ell m}$ is to be drawn from a probability distribution density $P[a_{\ell m}] da_{\ell m}$. The total probability distribution is the product of all these probability distributions, i.e.

$$dP = (a_{00} da_{00}) (Re[a_{10}] dRe[a_{10}]) (Im[a_{10}] dIm[a_{10}]) \dots \quad (4.57)$$

The correlation in (4.56) is evaluated as

$$\langle a_{\ell m} a_{\ell' m'}^* \rangle = \int a_{\ell m} a_{\ell' m'}^* dP \quad (4.58)$$

What (4.56) says is that

- Different $a_{\ell m}$'s are uncorrelated and therefore statistically independent variables. Put it differently, the correlation between two different $a_{\ell m}$'s which have distinct ℓ or distinct m is zero.
- The auto-correlation between two *same* $a_{\ell m}$'s (same ℓ and m) is C_ℓ . This auto-correlation is *independent* of m . The reason is that there is only one angle between the two directions $\hat{n}$ and $\hat{n}'$ or in other words, the two vectors $\hat{n}$ and $\hat{n}'$ always lie on a plane (the plane which contains them). Therefore we expect the correlation $C(\hat{n}, \hat{n}')$ to depend only on their dot-product $\hat{n} \cdot \hat{n}'$.

When (4.56) holds we say that the CMB is statistically isotropic (this is an assumption which can be put to the test, and this has been done with inconclusive results).

Using (4.56) we find that the correlation function becomes

$$C(\hat{n}, \hat{n}') = \sum_{\ell m} \sum_{\ell' m'} C_\ell Y_{\ell m}(\hat{n}') Y_{\ell' m'}^*(\hat{n}') \delta_{\ell \ell'} \delta_{m m'} \quad (4.59)$$

$$= \sum_{\ell m} C_\ell Y_{\ell m}(\hat{n}') Y_{\ell m}^*(\hat{n}') \quad (4.60)$$

$$= \frac{1}{4\pi} \sum_{\ell} (2\ell + 1) C_\ell P_\ell(\hat{n} \cdot \hat{n}') \quad (4.61)$$

If the angle between $\hat{n}$ and $\hat{n}'$ is θ then we have

$$C(\theta) = \frac{1}{4\pi} \sum_{\ell} (2\ell + 1) C_\ell P_\ell(\cos \theta) \quad (4.62)$$

The set of moments C_ℓ is called the *angular power spectrum*. The two lowest moments are special. The $\ell = 0$ moment C_0 is the monopole which is by definition zero as we have already subtracted

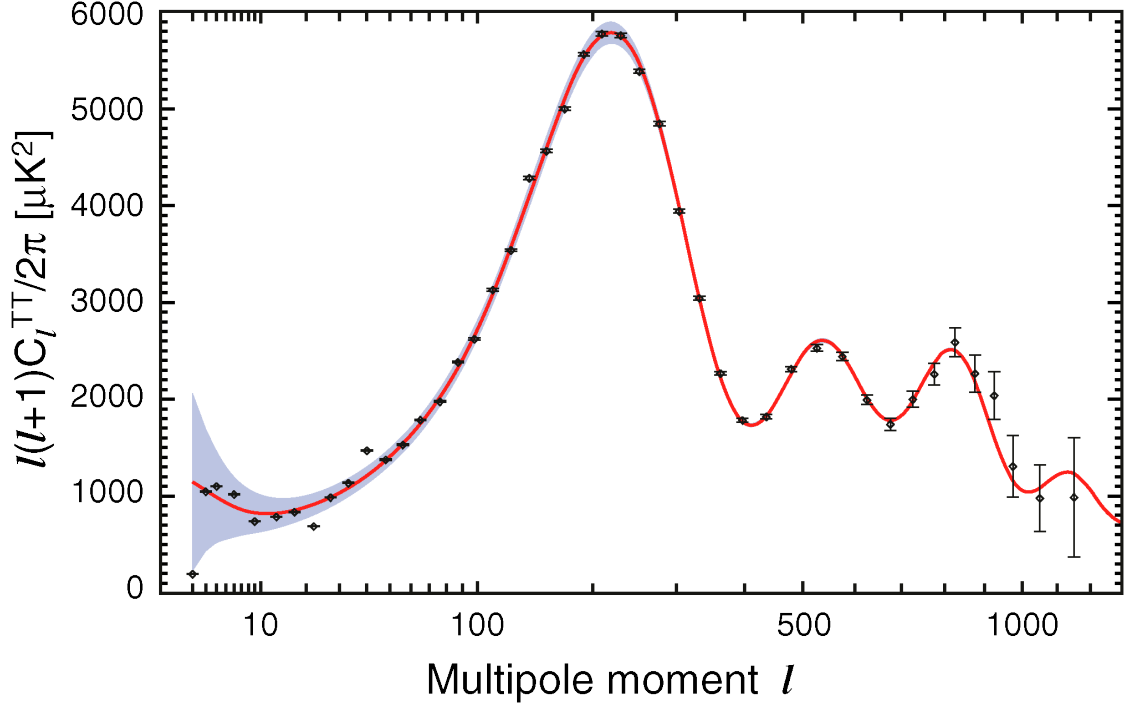


Figure 4.6: The CMB angular power spectrum C_ℓ as measured by the WMAP satellite after 7 years of observation. The grey shade is the cosmic variance around the best fit Λ CDM model.

the average temperature from the fluctuation $\Theta(\hat{n})$. The $\ell = 1$ moment is the dipole which we have also set to zero as we usually subtract it from the CMB sky map.

The fact that the temperature anisotropy is a random variable has one direct consequence. There is an intrinsic statistical error that cannot be removed by any means. It has to do with the fact that we have only one Universe, therefore only one CMB sky to observe. This means that there is only one monopole, 3 dipoles (because there are three m -moments for $\ell = 1$), 5 quadrupoles, and in general $2\ell + 1$ moments m for each ℓ . This introduces a fundamental uncertainty into any estimation of the number C_ℓ , in particular the error on C_ℓ is

$$\frac{\Delta C_\ell}{C_\ell} = \sqrt{\frac{2}{2\ell + 1}} \quad (4.63)$$

This fundamental uncertainty is called "Cosmic Variance". It means that if we measure C_ℓ then the measurement is uncertain by an amount $\pm \Delta C_\ell$, even if we have made the perfect measurement with zero experimental error! This is about 63% for $\ell = 2$, 22% for $\ell = 20$, 7% for $\ell = 200$ and smaller for larger ℓ . This is in a way bad news as it hinders the knowledge we can learn from the CMB. It also means that to gain the most amount of information we need to measure smaller angular scales where cosmic variance decreases. The good news is that the moments C_ℓ are statistically independent, so if we measure all of them, the total error from cosmic variance is multiplicative, e.g. if we measure $\ell = 20$ and $\ell = 21$ then the total error on both $\ell = 20$ and $\ell = 21$ drops to 3.3%. This means that the experimental power of the CMB is not in individual C_ℓ 's but in models with only a few parameters that can fit easily all of the C_ℓ 's.

The angular power spectrum as measured by the WMAP satellite after 7 years of observations

is shown in figure 4.6. The grey shade is the cosmic variance around the best fit Λ CDM model. Notice the tiny experimental error on each point, which however increases on smaller angular scales due to an experimental limit on the resolution of WMAP.

A few more things are in order.

- The number ℓ is inversely proportional to the angular scale observed. For large ℓ (typically $\ell > 10$ the relation $\ell \sim 2\pi/\theta$ approximately holds. Therefore large angular scales means small ℓ and vice versa. The scale of the first peak in the plot, thus corresponds to about 1.6° .
- We plot $\ell(\ell + 1)C_\ell$ rather than C_ℓ . This is because of two reasons. The first is because the factor $\ell(\ell + 1)$ becomes larger and larger for small scales which lifts the spectrum and makes it easier to display. The second has to do with the physics of CMB on large angular scales. As we shall see further below, in a matter dominated Universe (no cosmological constant present), the CMB spectrum for small ℓ obeys $\ell(\ell + 1)C_\ell = \text{const.}$
- The spectrum on the plot has units $(\mu K)^2$. This is because what is plotted is multiplied by the average CMB temperature squared. You can read-off the level of the temperature fluctuations from the plot and convince yourselves that they are indeed tiny compared to $2.725K$.

4.3 Generation of CMB anisotropies

Having described the variables that we use to observe the CMB anisotropies let us now discuss how they are formed, what are the physics behind their formation and what can we learn from them.

4.3.1 Tight-coupling, decoupling and recombination

We talked about how photons come to dominate the Universe at around $t = 100s$ at which point the Universe enters the radiation era. During that time photons interact with electrons and baryons (mainly protons and helium nuclei) via Compton scattering. At temperatures above $13.6eV$ the energy in photons is larger than the binding energy of hydrogen and so the Universe is kept ionized. However, even at temperatures less than $13.6eV$, there are still many high energy photons flying around that any neutral atom trying to form is instantly ionized once more. During this time we say that photons and electrons (and also baryons) are *tightly-coupled*. Eventually photons decouple from electrons (and baryons) at a point in time that we call *decoupling*. This happens when the electron number density n_e times the Thompson scattering cross-section σ_T (Thompson scattering is the classical limit of Compton scattering) becomes smaller than the Hubble rate H .

$$\text{Decoupling:} \quad n_e \sigma_T < H, \quad (4.64)$$

Soon after (but not at the same time) neutral atoms form at a temperature around $0.3eV$ (about $3500K$) and this event in the thermal history of the Universe is called *recombination*. During the times between decoupling and recombination photons and electrons are no longer tightly-coupled although the Universe is still ionized. This will lead to an important effect called *diffusion damping* (or Silk damping) that is observed on the CMB spectrum. After recombination the Universe is neutral and so photons can free-stream without being scattered off electrons. We call these times in the Universe the "Dark Ages". This phase persists all the way until there is light once again, that is when the first stars form, at which point we have a process called re-ionisation. The processes just described are shown in figure 4.7.

Tight-coupling and Free-streaming

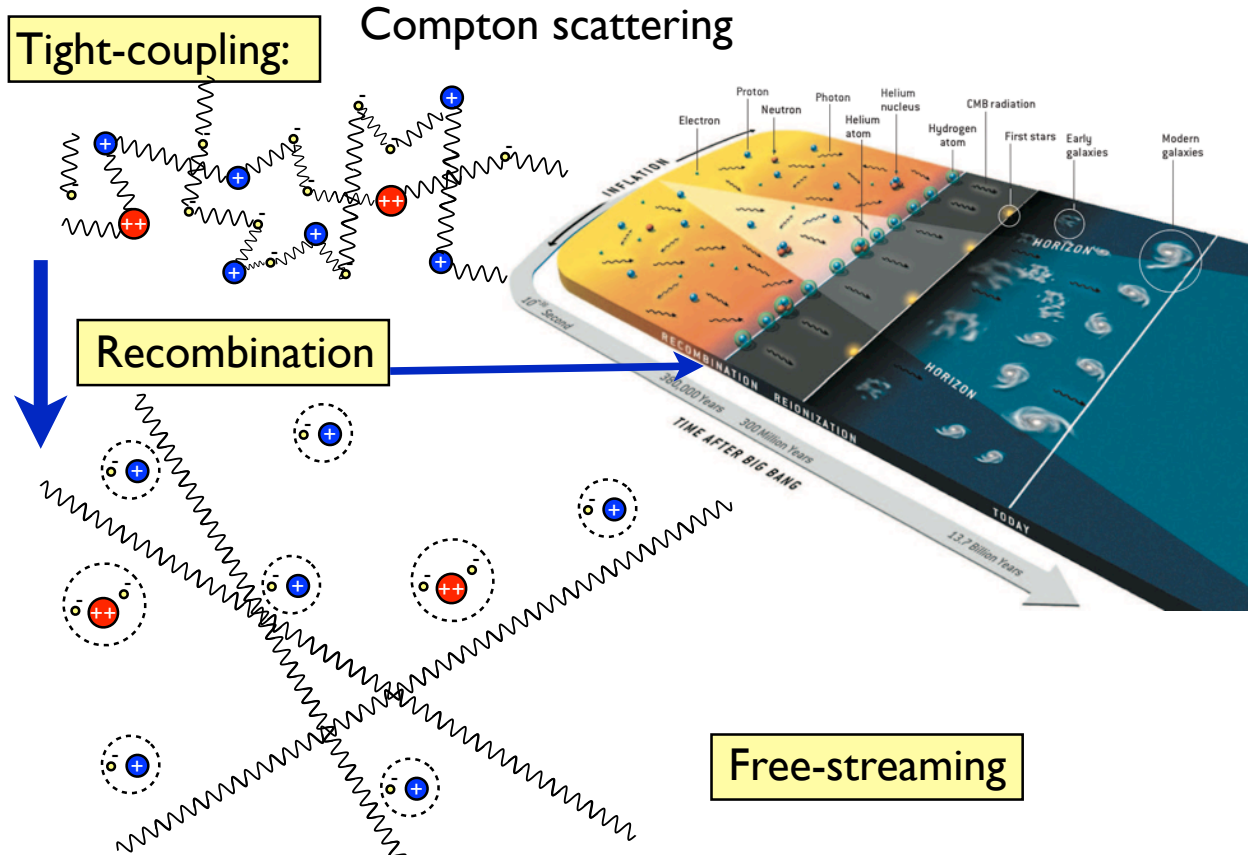


Figure 4.7: Pictorial representation of Compton scattering of photons off charged particles, leading to tight-coupling. After recombination, neutral atoms form and the photons free-stream through the Universe without interacting until the time of re-ionisation. These historic events are depicted on the diagram on the right of the figure.

4.3.2 Recombination in detail

Let us now describe the process of recombination in more detail. We shall ignore helium recombination although this will be displayed in the figures. During tight-coupling the reaction $e^- + p \leftrightarrow H + \gamma$ is in equilibrium so that any neutral hydrogen formed is quickly broken apart into free electrons and free protons. The number densities of electrons n_e , free protons n_p and hydrogen n_H during this time are given by

$$\frac{n_e n_p}{n_H} = \frac{n_e^{(0)} n_p^{(0)}}{n_H^{(0)}} \quad (4.65)$$

where $n_i^{(0)}$ is defined to be the species-dependent equilibrium number density given by (2.18) as all species involved are non-relativistic during this time. For brevity let us rewrite (2.18) which for species "i" is

$$n_i^{(0)} = g_i \left(\frac{m_i T}{2\pi} \right)^{3/2} e^{-m_i/T} \quad (4.66)$$

The condition (4.65) comes from the Boltzmann equation which tells us how we move out of equilibrium and which we state here without derivation (see Dodelson's book for a derivation in chapter 3)

$$a^{-3} \frac{d(n_e a^3)}{dt} = n_e^{(0)} n_p^{(0)} \langle \sigma v \rangle \left(\frac{n_H}{n_H^{(0)}} - \frac{n_e n_p}{n_e^{(0)} n_p^{(0)}} \right) \quad (4.67)$$

where $\langle \sigma v \rangle$ is the thermally average cross-section. Note this is similar to the Boltzmann equation (2.45) for the case of freeze-out discussed in section 2.5. It is important to realise that all of these processes involving the freeze out of particles, the fixed ratio of neutrons to protons or recombination and decoupling, *all involve the same basic physics*, namely solving Boltzmann equations for an out of equilibrium phenomenon.

It follows that equilibrium is maintained when the terms inside the brackets vanish while neutrality of the Universe ensures $n_e = n_p$. The quantity of interest is the *electron ionisation fraction* (also called the free electron fraction) which is the ratio

$$X_e \equiv \frac{n_e}{n_e + n_H} = \frac{n_p}{n_p + n_H}. \quad (4.68)$$

We see that the denominator is the total number of hydrogen nuclei. When the Universe is completely neutral $X_e \rightarrow 0$ while when it is completely ionized then according to (4.68) $X_e \rightarrow 1$. However, we are ignoring helium recombination which when included results to $X_e \approx 1.15$ when the Universe is fully ionized.

Now we use (4.66) in the RHS of (4.65) and ignore the small mass difference between p and H in the ratio that occurs in the prefactor to the exponential while we use $n_p = n_e$ in the LHS to get

$$\frac{n_e^2}{n_H} = \left(\frac{m_e T}{2\pi} \right)^{3/2} e^{-(m_e + m_p - m_H)/T} \quad (4.69)$$

We then use (4.68) to get $\frac{n_e}{n_H} = X_e/(1 - X_e)$ and so the above equation leads to the Saha equation

$$\frac{X_e^2}{1 - X_e} = \frac{1}{n_e + n_H} \left(\frac{m_e T}{2\pi} \right)^{3/2} e^{-\epsilon_H/T} \quad (4.70)$$

where $\epsilon_H = m_e + m_p - m_H \approx 13.6\text{eV}$ is the binding energy of hydrogen. To get to the Saha equation we have used only equilibrium physics and so we still haven't exploited the full potential of the Boltzmann equation. In fact for $T > \epsilon_H$ the Boltzmann equation (4.67) is very stiff and cannot be numerically solved to obtain X_e . For those temperatures *we have to use the Saha equation*.

Let us see what the Saha equation tells us. Neglecting the small numbers of helium atoms, then the denominator $n_e + n_H = n_p + n_H$ is just the baryon density which is given by $n_b = \eta n_\gamma \sim 10^{-9} T^3$. Hence for temperatures $T > \epsilon_H$ i.e. greater than 13.6eV the exponential is of order 1 and the RHS of the Saha equation (4.70) is of order $\frac{1}{n_b} \left(\frac{m_e T}{2\pi}\right)^{3/2} = 10^9 (m_e/T)^{3/2} \sim 10^{15}$. In other words it is huge which can only be accommodated if the denominator of the LHS nearly vanishes or $X_e \simeq 1$, implying all the hydrogen is ionised.

It is only when the temperature has dropped well below ϵ_H that significant recombination can take place. In fact as X_e falls it becomes more difficult to maintain equilibrium as the rate for recombination also falls. In order to solve for the free electron fraction X_e accurately, the Boltzmann equation (4.67) needs to be solved. But before doing that we can rewrite it in a more manageable form. Remembering that $n_e = n_p$, the Boltzmann equation (4.67) becomes

$$a^{-3} \frac{d(n_e a^3)}{dt} = n_b \langle \sigma v \rangle \left[(1 - X_e) \left(\frac{m_e T}{2\pi} \right)^{3/2} e^{-\epsilon_H/T} - X_e^2 n_b \right] \quad (4.71)$$

where we have used the fact that the ratio $n_e^{(0)} n_p^{(0)} / n_H^{(0)}$ is equal to the RHS of (4.69). Now since $n_b a^3$ is a constant and using $n_e = X_e n_b$, we can replace n_e on the LHS of Eqn. (4.71), pull the n_b through the derivative to get

$$\frac{dX_e}{dt} = \left[(1 - X_e) \beta - X_e^2 n_b \alpha^{(2)} \right] \quad (4.72)$$

where the ionisation rate is denoted by

$$\beta \equiv \langle \sigma v \rangle \left(\frac{m_e T}{2\pi} \right)^{3/2} \exp(-\epsilon_H/T) \quad (4.73)$$

and the recombination rate is

$$\alpha^{(2)} \equiv \langle \sigma v \rangle \quad (4.74)$$

The superscript (2) in $\alpha^{(2)}$ is because recombination to the ground ($n=1$) state is not relevant as it leads to the production of an ionising photon which immediately ionises the neutral atom, so that the net effect of such a recombination is zero. In fact to proceed capture must be to an excited state of hydrogen, and the temperature dependent part of the rate for this is (without proof)

$$\alpha^{(2)} \propto \left(\frac{\epsilon_H}{T} \right)^{1/2} \ln \left(\frac{\epsilon_0}{T} \right) \quad (4.75)$$

As soon as the Saha equation stops being valid we solve (4.72) numerically. The solution which includes helium recombination is shown in figure 4.8. We see that recombination occurs suddenly at $z \sim 1000$.

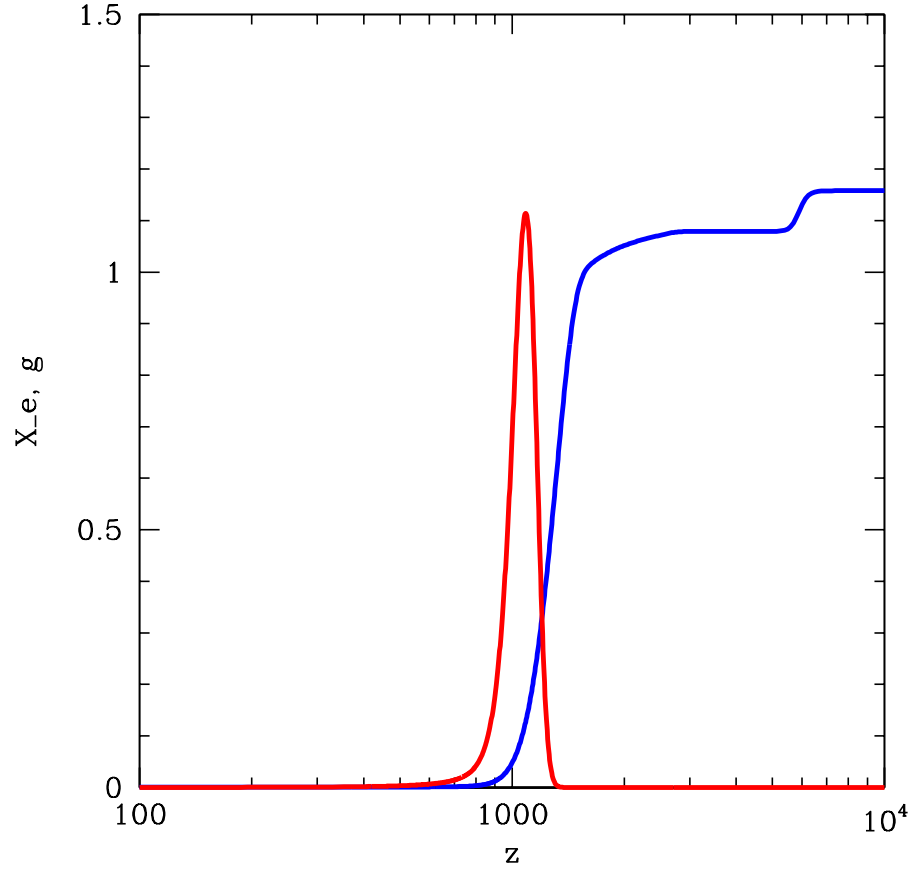


Figure 4.8: Blue curve: The evolution of the electron ionisation fraction X_e as a function of redshift z . Note how it drops abruptly around $z \sim 1000$ as the system moves out of equilibrium. Decoupling occurs during that period before recombination comes to an end.

Red curve: The visibility function $g(z)$, representing the probability that a photon we observe today last scattered of an electron between redshift z and $z+dz$. The peak of the visibility function defines the Last Scattering Surface (which has a finite thickness).

4.3.3 Decoupling in detail

The process most relevant to the CMB anisotropies is not so much recombination but rather it is decoupling. Decoupling occurs roughly when the rate for photons to Compton scatter off electrons becomes smaller than the expansion rate, i.e. when $\frac{n_e \sigma_T}{H} < 1$. Let's work out when that occurs and show that it occurs during recombination. The scattering rate $n_e \sigma_T$ can be written as $X_e n_b \sigma_T$, where $\sigma_T = 0.665 \times 10^{-24} \text{ cm}^2$ is the Thomson cross-section. Now $n_b = \bar{\rho}_b / m_b = 3H_0^2 \Omega_{0b} / (m_b a^3)$, hence,

$$n_e \sigma_T = \sigma_0 \frac{\Omega_{0b} h^2 X_e}{a^3} \quad (4.76)$$

where $\sigma_0 = 2.306 \times 10^{-5} \text{ Mpc}^{-1}$ (to account for helium, multiply the above expression by $1 - Y_{He}$ where Y_{He} is the helium fraction). We now divide by H which we get from the Friedmann equation as

$$H = H_0 \sqrt{\Omega_{0m}} a^{-3/2} \sqrt{1 + \frac{a_{eq}}{a}} \quad (4.77)$$

and $H_0 = 3.336 h \times 10^{-4} \text{ Mpc}^{-1}$. The final answer is (assuming that both matter and radiation to be important)

$$\frac{n_e \sigma_T}{H} = 0.069 \frac{\Omega_{0b} h X_e}{\sqrt{\Omega_{0m}} a^{3/2}} \left[1 + \frac{a_{eq}}{a}\right]^{-1/2} \quad (4.78)$$

Assuming typical values for the baryon density $\Omega_{0b} \sim 0.04$, total matter density $\Omega_{0m} \sim 0.27$ and $h \sim 0.7$ we find that $a_{eq} \sim 3 \times 10^{-4}$ and so

$$\frac{n_e \sigma_T}{H} \sim 117 X_e \quad (4.79)$$

i.e. when X_e drops below $1/117 \sim 0.01$, photons decouple. From the figure it is clear that X_e drops very quickly from unity to 10^{-3} , therefore decoupling takes place during recombination. That is the formation of the CMB primary anisotropies!

4.3.4 The optical depth, the visibility function and the Last Scattering Surface

We have seen that before decoupling, photons are tightly coupled to baryons. During that time the Universe was opaque and scattering of photons was very frequent. After decoupling photons free stream and don't scatter any more. So we ask the question: when did a CMB photon that we observe today last scattered? This is a question about probability. Rephrasing the question, what was the probability that a photon we observe today last scattered off an electron between time z and $z + dz$? The answer is provided by the visibility function $g(z)$.

To find the visibility function we first need to define a new quantity called the *optical depth* $\tau(z)$. The optical depth to redshift z

$$\tau(z) = \int_0^z \frac{n_e \sigma_T}{H} \frac{dz'}{1 + z'} \quad (4.80)$$

The meaning of the optical depth is now clear. It's rate of change measures decoupling. Since $n_e \sigma_T / H$ is practically zero until decoupling the optical depth only starts rising above zero when the Universe becomes opaque. Thus the function $e^{-\tau}$ has the opposite behaviour. It is equal to 1 for $z < z_{dec}$ and then quickly drops to zero for $z > z_{dec}$. If $e^{-\tau} = 1$ then photons free-stream while if $e^{-\tau} = 0$ then photons are tightly-coupled.

The visibility function is then defined as

$$g(z) = \frac{n_e \sigma_T}{1+z} e^{-\tau} \quad (4.81)$$

For $z \ll z_{dec}$ the factor $n_e \sigma_T$ drops to zero, while for $z \ll z_{dec}$ the exponential $e^{-\tau}$ drops to zero thus $g(z)$ peaks only around z_{dec} . The visibility function $g(z)$ represents the probability that a photon we observe today last scattered of an electron between redshift z and $z + dz$. The peak of the visibility function defines the Last Scattering Surface. It is plotted in figure 4.8. The finite thickness of the Last Scattering Surface will lead to an important effect on the CMB anisotropies: diffusion damping. We shall describe this effect further below.

4.3.5 The Boltzmann equation for photons

We have already seen that the background FRW description of photons is governed by a distribution function $\bar{f}(t, p)$. From the distribution function we extract things like the energy density and pressure.

To describe the CMB anisotropies we have to go beyond the background FRW Universe and include small fluctuations. The tools we used when we studied structure formation are directly applicable here as well. In particular, the photons acquire an energy density fluctuation $\delta\rho_\gamma(\eta, \vec{k})$ from which we obtain a density contrast δ_γ . They also have a velocity fluctuation v_γ , pressure $\delta P_\gamma = \frac{1}{3}\delta\rho_\gamma$ and shear σ_γ . These quantities are part of the photon energy-momentum tensor. In general, the energy-momentum tensor is obtained from a distribution function and the same holds here. But now, in order to obtain the fluctuations we just described, we need a more general distribution function than $\bar{f}$. For an inhomogeneous Universe the photon distribution function will depend on time, spatial position or equivalently Fourier mode $\vec{k}$, and photon momentum vector: $f(\eta, \vec{x}, \vec{p})$. We separate the momentum vector into magnitude and direction as $\vec{p} = p\hat{p}$ so that $f = f(\eta, \vec{x}, p, \hat{p})$. The vector $\hat{p}$ denotes the direction of photon propagation.

We now do a trick: Taylor expansion. We will assume that $f = f(\eta, \vec{x}, p, \hat{p})$ is a small fluctuation away from the average distribution function $\bar{f}$:

$$f = \bar{f}(\eta, p) + \delta f(\eta, \vec{x}, p, \hat{p}) \quad (4.82)$$

Now remember that $\bar{f}$ is the Bose-Einstein distribution function with temperature $T(\eta)$. Without loss of generality, we may assume that f is also given by the Bose-Einstein distribution function, only now the temperature depends on spatial position as well as the photon momentum, i.e. $T = T(\eta, \vec{x}, \hat{p})$. Therefore

$$f(\eta, \vec{x}, p, \hat{p}) = \frac{2}{(2\pi)^3} \frac{1}{\exp[p/T(\eta, \vec{x}, \hat{p})] - 1} \quad (4.83)$$

Notice that we have omitted p from the temperature $T(\eta, \vec{x}, \hat{p})$, i.e. the temperature $T(\eta, \vec{x}, \hat{p})$ depends only on the direction of momentum but not its magnitude. The reason is that the magnitude of the photon momentum is virtually unchanged during a Compton scatter.

We further expand the temperature as the average background temperature $\bar{T}(\eta)$ and the temperature fluctuation $\Delta T(\eta, \vec{x}, \hat{p})$. But since we have expressed the observed CMB spectrum as a temperature contrast $\Theta(\hat{n})$ let us do the same here and define the *local* temperature contrast $\Delta(\eta, \vec{x}, \hat{p})$ as

$$\Delta(\eta, \vec{x}, \hat{p}) = \frac{T(\eta, \vec{x}, \hat{p}) - \bar{T}(\eta)}{\bar{T}(\eta)} \quad (4.84)$$

Replacing $T(\eta, \vec{x}, \hat{p})$ with $\Delta(\eta, \vec{x}, \hat{p})$ in the distribution function and expanding as a Taylor series for small $\Delta(\eta, \vec{x}, \hat{p})$ we get that

$$\delta f = -p \frac{\partial \bar{f}}{\partial p} \Delta(\eta, \vec{x}, \hat{p}) \quad (4.85)$$

and so the fluctuation of the distribution function away from the background $\bar{f}$ is given in terms of the temperature fluctuation. Now just as the background distribution $\bar{f}$ obeyed the Boltzmann equation, so will the full distribution function. The Boltzmann equation is slightly more complicated in this case. It is

$$\frac{\partial f}{\partial t} + \frac{dx^i}{dt} \vec{\nabla}_i f + \frac{dp}{dt} \frac{\partial f}{\partial p} + \frac{d\hat{p}^i}{dt} \frac{\partial f}{\partial \hat{p}^i} = C[f] \quad (4.86)$$

where the term $C[f]$ on the RHS is the term due to collisions of photons with free electrons, i.e. it is the terms which describes Compton scattering. The velocity term $\frac{dx^i}{dt}$ is related to $\frac{d\hat{p}^i}{dt}$ as

$$\frac{dx^i}{dt} = (1 + \Psi + \Phi) \frac{\hat{p}^i}{a} \quad (4.87)$$

where to remind you Φ and Ψ are the two gravitational potentials which are part of the metric perturbation, while the terms $\frac{dp}{dt}$ and $\frac{d\hat{p}^i}{dt}$ are evaluated using the geodesic equation for photons. After some calculations (see Dodelson chapter 4, pp.89-92) the Boltzmann equation becomes

$$\frac{\partial f}{\partial t} + \frac{\hat{p}^i}{a} \vec{\nabla}_i f - p \frac{\partial f}{\partial p} \left[H - \dot{\Phi} + \frac{\hat{p}^i}{a} \vec{\nabla}_i \Psi \right] = C[f] \quad (4.88)$$

We then isolate the fluctuation part using (4.82) and replace δf with Δ using (4.85). The collision term can also be evaluated (see Dodelson chapter 4) and finally after long calculations, and also switching to Fourier space, we get the Boltzmann equation for the temperature fluctuation as

$$\Delta' + ik\mu\Delta - \Phi' + ik\mu\Psi = an_e\sigma_T [\Delta_0 - \Delta + ik\mu u_b] \quad (4.89)$$

where we have switched from cosmic time t to conformal time η and where $\mu = \hat{k} \cdot \hat{p}$ is the cosine of the angle between the Fourier vector and the momentum of the photon. The first two terms describe the free-streaming of photons in empty space. The third and fourth term describe the effect of gravity. The term on the RHS describes Compton scattering and depends on the product of the electron number density n_e times Thompson scattering cross-section σ_T , the temperature anisotropy Δ , the monopole of the temperature anisotropy Δ_0 (described below) and the baryon velocity $\vec{u}_b = \vec{\nabla}_i u_b$. In describing the Compton scattering term we have neglected the angular dependence of the Compton scattering amplitude which leads to terms related to CMB polarisation and are not important for the temperature anisotropies. We will deal with CMB polarisation later. The term $n_e\sigma_T$ was the term we have encountered already when we discussed decoupling. So here it appears explicitly in the perturbed Boltzmann equation. In fact $an_e\sigma_T$ is related to the optical depth.

By inspection, we see that (4.89) depends on the Fourier direction $\hat{k}$ and momentum direction $\hat{p}$ only through their dot product μ . Therefore the temperature anisotropy in momentum space should only depend on μ and not on $\hat{k}$ and $\hat{p}$ separately, i.e. $\Delta = \Delta(\eta, k, \mu)$. Note that μ is a cosine and so takes values from -1 to $+1$ only. Therefore we can make use of the Legendre polynomials which allows us to expand any function of a variable μ taking values in the interval $[-1, 1]$ as a series of Legendre polynomials and appropriate coefficients. From (4.43) and (4.44) we have

$$\Delta(\eta, k, \mu) = \sum_{\ell} i^{\ell} (2\ell + 1) \Delta_{\ell}(\eta, k) P_{\ell}(\mu) \quad (4.90)$$

with inverse given as

$$\Delta_\ell(\eta, k) = \frac{(-i)^\ell}{2} \int_{-1}^1 d\mu P_\ell(\mu) \Delta(\eta, k, \mu) \quad (4.91)$$

You may be tempted to guess the the ℓ here is the same as the ℓ in C_ℓ and that C_ℓ will somehow be related to Δ_ℓ . If you have, then you are right, but the formula would have to wait.

What about Δ_0 ? The term Δ_0 is the local monopole term of the local temperature anisotropy and is found from the $\ell = 0$ moment of the above expansion (4.91) as

$$\Delta_0(\eta, k) = \frac{1}{2} \int_{-1}^1 d\mu \Delta(\eta, k, \mu) \quad (4.92)$$

where to remind you $P_0(\mu) = 1$.

We now proceed to convert the μ -dependence of (4.89) to an ℓ -dependence. This will have the effect of splitting (4.89) into a hierarchy of equations, one for each ℓ . We do that by multiplying both sides of (4.89) with $P_L(\mu)$ and then integrating over μ . The integrals over μ are performed using the Legendre polynomial identities (4.45) and (4.49) and the final result is a differential equation for the monopole

$$\Delta'_0 = k\Delta_1 + \Phi' \quad (4.93)$$

a differential equation for the dipole

$$\Delta'_1 = \frac{k}{3} (2\Delta_2 - \Delta_0) - \frac{k}{3} \Psi - an_e \sigma_T \left(\Delta_1 + \frac{k}{3} u_b \right) \quad (4.94)$$

and a set of differential equations for each higher moment Δ_ℓ , for $\ell \geq 2$

$$\Delta'_\ell = -\frac{k}{2\ell+1} [\ell\Delta_{\ell-1} - (\ell+1)\Delta_{\ell+1}] - an_e \sigma_T \Delta_\ell \quad (4.95)$$

To compute the CMB anisotropies what we now have to do is to solve the Boltzmann equation along with the Einstein equations (3.147) (3.146) and the fluid equations for baryons (and dark matter or any other species that we include into our model). But the Einstein equations require that we know the perturbed density, velocity, pressure and shear of photons. We get these from the perturbed distribution function using the perturbed version of (4.12). The answer is

$$\delta_\gamma = 4\Delta_0 \quad (4.96)$$

$$u_\gamma = -\frac{3}{k} \Delta_1 \quad (4.97)$$

$$\sigma_\gamma = \frac{3}{k^2} \Delta_2 \quad (4.98)$$

and of course the pressure contrast is always related as $\Pi_\gamma = \frac{1}{3}\delta_\gamma$ since $w = 1/3$ for radiation. Thus the energy-momentum variables are each given by lowest three multipole moments Δ_ℓ of the local temperature anisotropy.

4.3.6 The photon-baryon fluid during tight-coupling

During tight-coupling the term $an_e \sigma_T$ becomes very large. Looking at the multipole moment equations (4.95) we see that if $an_e \sigma_T$ becomes very large then all Δ_ℓ moments for ($\ell \geq 2$) must

vanish (more precisely they become extremely small with each higher ℓ receiving an additional power of $1/(an_e\sigma_T)$). Furthermore the same condition applied in (4.94) forces $\Delta_1 = -\frac{k}{3}u_b$.

This is called the tight-coupling approximation. Applying this approximation to our equations means that the only surviving equations are (4.93) and (4.94) and the later becomes

$$\Delta'_1 = -\frac{k}{3}\Delta_0 - \frac{k}{3}\Psi \quad (4.99)$$

Combining (4.93) and (4.99) we can eliminate Δ_1 to get a single 2nd order differential equation for Δ_0 as

$$\Delta''_0 + \frac{k^2}{3}\Delta_0 = \Phi'' - \frac{k^2}{3}\Psi \quad (4.100)$$

We recognise the above equation as a forced harmonic oscillator with speed of sound $c_s = \frac{1}{\sqrt{3}}$. It is sourced by gravity through the RHS. It should be now clear why we get a nice smooth and oscillating angular power spectrum C_ℓ . It is because of oscillations in the local photon temperature anisotropy. But we have been too rough with our approximation and the equation above neglects important effects coming from the baryons. The fact that the baryon velocity appears in (4.94) in the Compton scattering term, should have given us a warning signal. The Compton scattering term basically violates conservation of the energy-momentum tensor of the photons, $\nabla_\mu T_{(\gamma)\nu}^\mu \neq 0$, but the total energy momentum tensor of the photons plus baryons is still conserved: $\nabla_\mu (T_{(\gamma)\nu}^\mu + T_{(b)\nu}^\mu) = 0$. This means that there is momentum transfer from the photons to the baryons and vice versa. To make sure that the total energy momentum tensor is conserved, there should be a similar Compton scattering term appearing in the baryon velocity equation (3.169). That term will furthermore be multiplied by a ratio of the background baryon and photon energy densities $\bar{\rho}_b$ and $\bar{\rho}_\gamma$. More precisely the baryon-to-photon ratio is

$$R = \frac{3\bar{\rho}_b}{4\bar{\rho}_\gamma} \quad (4.101)$$

To find the correct equations which include the effects from the baryons we need to include the next order in an expansion in powers of $1/(an_e\sigma_T)$. This leads to baryons contributing an effective mass (the term $k^2c_s^2$ can be thought of as an effective mass for the oscillator) and a damping term and

$$\Delta''_0 + \frac{R}{1+R}\mathcal{H}\Delta'_0 + k^2c_s^2\Delta_0 = -\frac{k^2}{3}\Psi + \frac{R}{1+R}\mathcal{H}\Phi' + \Phi'' \quad (4.102)$$

where now the speed of sound is changed to

$$c_s = \frac{1}{\sqrt{3(1+R)}} \quad (4.103)$$

This contribution from the baryons will turn out to have an important effect as we shall see later on. The good news is that little has changed in (4.102). The only difference is that now we have a damped harmonic oscillator (sourced by gravity) and furthermore the sound speed is time dependent.

For simplicity, let's assume that the potentials are approximately constant. Furthermore, let's assume that the speed of sound is slowly varying so that $c'_s \approx 0$. However, $c'_s/c_s = -\frac{R\mathcal{H}}{2(1+R)}$ which is one half the term appearing in front of the Δ' . Hence, if the speed of sound is slowly varying we can ignore the damping term. Under these approximations our equation becomes

$$\Delta''_0 + k^2c_s^2\Delta_0 = -\frac{k^2}{3}\Psi \quad (4.104)$$

Let's first try to understand what's going on before solving the equation. We are dealing with a simple harmonic oscillator with a constant forcing provided by gravity. Basically the term $k^2 c_s^2$ looks like a pressure term. Indeed this is the pressure provided by the photon-baryon fluid which is trying to resist being squashed by gravity. It's Jeans analysis again only now the "density" is zero. The relevant scale in this case is not the Jeans length (which is the horizon) but a new scale called the *sound horizon*:

$$r_s(\eta) = \int_0^\eta c_s(\eta') d\eta' \quad (4.105)$$

So for modes outside the horizon (Jeans length) Δ stays constant ($\Delta_0 = \frac{1}{4}\delta_\gamma \approx \text{const}$). After the mode crosses the horizon it will have to decay, but only slightly, for it then enters the sound horizon (which is almost as large as the horizon) and starts oscillating.

The solution to (4.104) is

$$\Delta_0(\eta, k) = -(1 + R)\Psi + A \cos[kr_s(\eta)] + B \sin[kr_s(\eta)] \quad (4.106)$$

which is approximately the solution to (4.102) as long as the potentials and the speed of sound are slowly varying. For adiabatic initial conditions the constant $B = 0$ leaving only A . As $\eta \rightarrow 0$, we also have $R \rightarrow 0$ and the constant A is found to be $\frac{1}{2}\Phi_{(sup)}$ (remember that $\Phi = \Psi$ in the absence of shear). Hence, for adiabatic initial conditions our solution becomes

$$\Delta_0(\eta, k) = -(1 + R)\Psi + \frac{\Phi_{(sup)}}{2} \cos[kr_s(\eta)] \quad (4.107)$$

which means that the solution for the dipole is

$$\Delta_1(\eta, k) = -\frac{c_s \Phi_{(sup)}}{2} \sin[kr_s(\eta)] \quad (4.108)$$

Now these are the solutions for $\eta < \eta_{dec} = \eta_*$. So at decoupling, the intrinsic temperature monopole and dipole are given by

$$\Delta_0(\eta_*, k) = -(1 + R)\Psi + \frac{\Phi_{(sup)}}{2} \cos[kr_s(\eta_*)] \quad (4.109)$$

$$\Delta_1(\eta_*, k) = -\frac{c_s \Phi_{(sup)}}{2} \sin[kr_s(\eta_*)] \quad (4.110)$$

We shall return to these solutions later in order to understand the peak structure of the temperature anisotropies but for the moment let us turn to the time after decoupling.

4.3.7 Photons after decoupling: free-streaming

After decoupling we get that $an_e\sigma_T \rightarrow 0$. This leads to the opposite approximation to tight-coupling: the *free-streaming* approximation. Rather than dealing with the moment expansion directly, let us return to (4.89). The reason is that unlike the tight-coupling case, the higher multipole moments Δ_ℓ do not vanish. What happens is that the power stored in the monopole and dipole at decoupling, starts to spread to the other multipoles which will in turn give us the angular power spectrum C_ℓ . But one thing at a time.

Setting $an_e\sigma_T = 0$ in (4.89) gives us the free-streaming equation

$$\Delta' + ik\mu\Delta = \Phi' - ik\mu\Psi \quad (4.111)$$

which is valid for $\eta > \eta_*$. On the LHS we have photon free-streaming and on the RHS we have gravity. This equation is easy to solve as there is no explicit time η appearing anywhere. It is thus an inhomogeneous first order linear ordinary differential equation. The LHS can be rewritten as $e^{-ik\mu\eta} \frac{d}{d\eta} [e^{ik\mu\eta} \Delta]$ so that the full solution that we may take to time η_0 , i.e. today, is

$$\Delta(\eta_0, k, \mu) = e^{ik\mu(\eta_* - \eta_0)} \Delta(\eta_*, k, \mu) + \int_{\eta_*}^{\eta_0} d\eta e^{ik\mu(\eta - \eta_0)} [\Phi'(\eta, k) - ik\mu\Psi(\eta, k)] \quad (4.112)$$

The solution above depends on an initial condition $\Delta(\eta_*, k, \mu)$ which gives the anisotropies at η_* and an integral which gives the anisotropies after η_* . Notice how the μ -dependence is completely accounted for either by the exponential or the $ik\mu$ terms: there is no μ -dependence in the potentials, while the μ -dependence in the initial condition $\Delta(\eta_*, k, \mu)$ is easily calculated: $\Delta(\eta_*, k, \mu) = \Delta_0(\eta_*, k) + 3i\mu\Delta_1(\eta_*, k)$ where $\Delta_0(\eta_*, k)$ and $\Delta_1(\eta_*, k)$ are the monopole and dipole at last scattering, which have been calculated using the tight-coupling approximation. Thus we have succeeded in calculating (very approximately) the intrinsic photon temperature anisotropy $\Delta(\eta_0, k, \mu)$ today. I emphasise the words very approximately, as we have ignored a few important effects. The first is the time variation of the potentials during tight-coupling and the second is the fact that the last scattering surface has a finite thickness. We shall return to these later.

We can manipulate (4.112) by integrating the Ψ term by parts and by replacing $\Delta(\eta_*, k, \mu)$ with the monopole and with the baryon velocity. The integration by parts is as follows. Consider on the Ψ -term. We have

$$\begin{aligned} \int_{\eta_*}^{\eta_0} d\eta e^{ik\mu(\eta - \eta_0)} (-ik\mu) \Psi(\eta, k) &= - \int_{\eta_*}^{\eta_0} d\eta \frac{d}{d\eta} (e^{ik\mu(\eta - \eta_0)}) \Psi(\eta, k) \\ &= -\Psi(\eta_0, k) + e^{ik\mu(\eta_* - \eta_0)} \Psi(\eta_*, k) + \int_{\eta_*}^{\eta_0} d\eta e^{ik\mu(\eta - \eta_0)} \Psi'(\eta, k) \end{aligned}$$

Since during tight-coupling $\Delta_1 = -\frac{k}{3}u_b$, the initial condition becomes

$$\Delta(\eta_*, k, \mu) = \Delta_0(\eta_*, k) - ik\mu u_b(\eta_*, k) \quad (4.113)$$

so that back to (4.112) we get

$$\Delta(\eta_0, k, \mu) = e^{ik\mu(\eta_* - \eta_0)} [\Delta_0 + \Psi - ik\mu u_b](\eta_*, k) + \int_{\eta_*}^{\eta_0} d\eta e^{ik\mu(\eta - \eta_0)} [\Phi'(\eta, k) + \Psi'(\eta, k)] \quad (4.114)$$

Notice that we have also ignored the $-\Psi(\eta_0, k)$ term. The reason is that as this term has no μ -dependence it contributes only to the monopole $\Delta_0(\eta_0, k)$ and is therefore unobservable (the monopole will contribute only to C_0 which is by definition zero).

The first term in (4.114) is what we call the *Primary Anisotropies*. The primary CMB anisotropies are the ones formed at decoupling and consist of the effective temperature anisotropy $\Delta_0 + \Psi$ and a local Doppler effect anisotropy $-ik\mu u_b$. The second term with the integral is a kind of a secondary anisotropy as it depends on all the time after decoupling. In particular it leads to an effect called the *Integrated Sachs-Wolfe* (ISW) effect.

4.3.8 The formal solution to the Boltzmann equation: the line-of-sight integral

We have expressed the temperature anisotropy today $\Delta(\eta_0, k, \mu)$ in terms of the primary anisotropies at decoupling $\Delta_0 + \Psi - ik\mu u_b$ and one type of secondary anisotropy due to the decay of the gravitational potentials, called the Integrated Sachs-Wolfe effect. This was done under the assumption of

instantaneous decoupling. Here we shall find the full solution to the Boltzmann equation without any approximation.

We start from the Boltzmann equation (4.89) and re-arrange it as follows.

$$\Delta' + ik\mu\Delta + an_e\sigma_T\Delta = \Phi' - ik\mu\Psi + an_e\sigma_T[\Delta_0 + ik\mu u_b] \quad (4.115)$$

Remember the optical depth? Well, the term $an_e\sigma_T$ is related to the optical depth. In fact as you should be able to check easily, if you take the defining equation for the optical depth (??) and differentiate wrt to conformal time η you will find that

$$\tau' = -an_e\sigma_T \quad (4.116)$$

so that the Boltzmann equation becomes

$$\Delta' + ik\mu\Delta - \tau'\Delta = \Phi' - ik\mu\Psi - \tau'[\Delta_0 + ik\mu u_b] \quad (4.117)$$

So now on the LHS we have only terms which depend on Δ and on the RHS we have a source term. The LHS may be integrated in the same way we did it for free-streaming. In fact the only difference is the $\tau'\Delta$ term. We find that the LHS is given by

$$LHS = e^{-ik\mu\eta+\tau(\eta)} \frac{d}{d\eta} \left[e^{ik\mu\eta-\tau(\eta)} \Delta \right] \quad (4.118)$$

so that the complete solution to the Boltzmann equation is given by

$$\Delta(\eta, k, \mu) = \int_0^\eta d\eta' e^{ik\mu(\eta'-\eta)+\tau(\eta)-\tau(\eta')} [\Phi' - ik\mu\Psi - \tau'(\Delta_0 + ik\mu u_b)] \quad (4.119)$$

so that taking $\eta = \eta_0$ to get the solution today we get

$$\Delta(\eta_0, k, \mu) = \int_0^{\eta_0} d\eta e^{ik\mu(\eta-\eta_0)-\tau(\eta)} [\Phi' - ik\mu\Psi - \tau'(\Delta_0 + ik\mu u_b)] \quad (4.120)$$

where one of the τ terms has disappeared as $\tau(\eta_0) = 0$. We proceed further to express the $e^{ik\mu(\eta-\eta_0)}ik\mu e^{-\tau}\Psi$ term as $e^{-\tau}\Psi \frac{d}{d\eta} e^{ik\mu(\eta-\eta_0)}$ and then integrate it by parts to get

$$\Delta(\eta_0, k, \mu) = \int_0^{\eta_0} d\eta e^{ik\mu(\eta-\eta_0)} [g(\eta)(\Delta_0 + \Psi + ik\mu u_b) + e^{-\tau}(\Phi' + \Psi')] \quad (4.121)$$

where $g(\tau) = -\tau'e^{-\tau} = an_e\sigma_T e^{-\tau}$ is the visibility function. Equation (4.121) is the full solution of the Boltzmann equation in terms of the gravitational potentials, the intrinsic temperature monopole $\Delta_0(\eta, k)$ and the baryon velocity $u_b(\eta, k)$. You may have already guessed how the various terms compare to the approximate solution we found earlier. Let's find out explicitly. We need two facts. Firstly, the visibility function peaks at decoupling (see figure 4.8), so the instantaneous decoupling approximation amounts to setting

$$g(\eta) = \delta(\eta - \eta_*) \quad (4.122)$$

Therefore we may integrate the term proportional to $g(\eta)$ to get

$$e^{ik\mu(\eta_*-\eta_0)}(\Delta_0 + \Psi + ik\mu u_b)(\eta_*, k) \quad (4.123)$$

This is nothing but the primary anisotropy term we found in (4.114)! Secondly, the term $e^{-\tau}$ is like a step function which equals 1 for $\eta > \eta_*$ and 0 for $\eta < \eta_*$ (see discussion in section ??). Thus the term proportional to $e^{-\tau}$ can be written as

$$\int_{\eta_*}^{\eta_0} d\eta e^{ik\mu(\eta-\eta_0)} (\Phi' + \Psi') \quad (4.124)$$

which is the ISW term in (4.114)! This is how good our approximation was. When we discuss the features of the CMB anisotropy spectrum further below we shall therefore use the instantaneous decoupling approximation of (4.114) and include the effect of the finite thickness of the visibility function in a different way.

Let us now find a different form of the solution (4.121) that is called the *line-of-sight* integral. This form will be more useful to make contact with the angular power spectrum C_ℓ .

What we do is to relate $\Delta(\eta_0, k, \mu)$ to the multipole moments $\Delta_\ell(\eta_0, k)$ today so that we don't have to worry about μ . We do that using (4.91). Before performing the μ -integral let us do a further integration by parts, this time on the $ik\mu u_b$ term. The procedure is the same as for the $ik\mu\Psi$ term and we get an alternative form of (4.121) which now involves the derivative of the visibility function g' :

$$\Delta(\eta_0, k, \mu) = \int_0^{\eta_0} d\eta e^{ik\mu(\eta-\eta_0)} [g(\Delta_0 + \Psi - u'_b) - g'u_b + e^{-\tau}(\Phi' + \Psi')] \quad (4.125)$$

so that the only place that μ appears is in the exponential. Therefore we have only one integral over μ to perform in (4.91), namely

$$\int_{-1}^1 d\mu e^{ik\mu(\eta-\eta_0)} P_\ell(\mu) \quad (4.126)$$

To do that we use an important relation called the Rayleigh relation. The Rayleigh relation is an expansion of the $e^{ik\mu(\eta-\eta_0)}$ as a series in Legendre polynomials and is

$$e^{ix\mu} = \sum_{\ell} (2\ell + 1) i^\ell j_\ell(x) P_\ell(\mu) \quad (4.127)$$

where the expansion coefficients $j_\ell(x)$ are functions you may have encountered before. They are the spherical Bessel functions and are solutions to the spherical Bessel equation. You may have encountered them in quantum mechanics and in particular regarding the hydrogen atom. They are nothing but the radial eigenfunctions of the wave-function of the hydrogen atom. In fact the wave-function of the hydrogen atom splits into spherical Bessel functions and spherical harmonics: $\psi(r, \theta, \varphi) = j_\ell(r) Y_{\ell m}(\theta, \varphi)$ where ℓ in this case is the angular momentum quantum number and m the magnetic quantum number. The spherical Bessel functions are described in more detailed in the next subsection.

Using the Rayleigh relation we can perform the μ integral noting also that since the argument of the Bessel function must be positive, we need to use the complex-conjugate Rayleigh relation:

$$\begin{aligned} \frac{(-i)^\ell}{2} \int_{-1}^1 d\mu e^{ik(\eta-\eta_0)\mu} P_\ell(\mu) &= \frac{(-i)^\ell}{2} \int_{-1}^1 d\mu e^{-ik(\eta_0-\eta)\mu} P_\ell(\mu) \\ &= \frac{(-i)^\ell}{2} \sum_{\ell'} (2\ell' + 1) (-i)^{\ell'} j_{\ell'}[k(\eta_0 - \eta)] \int_{-1}^1 d\mu P_{\ell'}(\mu) P_\ell(\mu) \\ &= (-1)^\ell j_\ell[k(\eta_0 - \eta)] \end{aligned} \quad (4.128)$$

Defining

$$\Delta_\ell(\eta_0, k) = (-1)^\ell \tilde{\Delta}_\ell(\eta_0, k) \quad (4.129)$$

the local temperature multipoles today are given by

$$\tilde{\Delta}_\ell(\eta_0, k) = \int_0^{\eta_0} d\eta j_\ell[k(\eta_0 - \eta)] [g(\Delta_0 + \Psi - u'_b) - g' u_b + e^{-\tau} (\Phi' + \Psi')] \quad (4.130)$$

The above equation is called the line-of-sight integral. It gives us directly the local temperature multipoles in terms of a set of known functions, i.e. the spherical Bessel functions, and a set of sources: Δ_0 , Ψ , Φ and u_b (and their derivatives). This provides us with tremendous simplification when it comes to calculate the C_ℓ 's. The spherical Bessel functions can be calculated once, tabulated, stored and used everytime we want to get the C_ℓ 's for a given model; the Spherical Bessel functions are the same for all models. What changes is only Δ_0 , Ψ , Φ and u_b . The line-of-sight integral is the heart of any good C_ℓ calculator like CMBfast, CAMB, CMBeasy and DASH.

We shall return to the line-of-sight integral when we discuss projection effects. Now let us find the final relation we need so that we can calculate the C_ℓ 's. We need to relate the C_ℓ 's to $\Delta_\ell(\eta_0, k)$. As you may suspect the ℓ is the same but what is the exact relation?

4.3.9 Special functions: Spherical Bessel functions

The spherical Bessel functions $j_\ell(x)$, indexed by an integer $\ell = 0, 1, 2, \dots$ are solutions to the differential equation

$$\left[\frac{1}{x^2} \frac{\partial}{\partial x} \left(x^2 \frac{\partial}{\partial x} \right) + 1 - \frac{\ell(\ell+1)}{x^2} \right] j_\ell(x) = 0 \quad (4.131)$$

The first few spherical Bessel functions are

$$j_0(x) = \frac{\sin x}{x} \quad (4.132)$$

$$j_1(x) = \frac{\sin x}{x^2} - \frac{\cos x}{x} \quad (4.133)$$

$$j_2(x) = \left(\frac{3}{x^3} - \frac{1}{x} \right) \sin x - \frac{3}{x^2} \cos x \quad (4.134)$$

The spherical Bessel functions obey the orthogonality relation

$$\int_0^\infty dr r^2 j_\ell(kr) j_\ell(qr) = \frac{\pi}{2k^2} \delta^{(r)}(k - q) \quad (4.135)$$

(notice the r^2 appearing under the integral). Since both r and k are spherical variables which take values in $[0, \infty)$ the Dirac δ function has support in the same range, i.e. $\int_0^\infty \delta^{(r)}(x - y) = 1$.

We find it sometimes useful to consider the asymptotic forms of $j_\ell(x)$ as $x \rightarrow 0$ or $x \rightarrow \infty$. These are

$$\text{As } x \rightarrow 0 \quad \text{then } j_\ell(x) \rightarrow \frac{1}{1 \cdot 3 \cdot 5 \cdots (2\ell + 1)} x^\ell \quad (4.136)$$

$$\text{As } x \rightarrow \infty \quad \text{then } j_\ell(x) \rightarrow \frac{1}{x} \sin \left(x - \frac{\ell\pi}{2} \right) \quad (4.137)$$

The Bessel functions obey a number of recurrence relations that relate different orders ℓ and/or their first derivatives. These are

$$j_\ell = \frac{x}{2\ell+1} (j_{\ell-1} + j_{\ell+1}) \quad (4.138)$$

and

$$\frac{dj_\ell}{dx} = \frac{1}{2\ell+1} [\ell j_{\ell-1} - (\ell+1)j_{\ell+1}] \quad (4.139)$$

$$= \frac{\ell}{x} j_\ell - j_{\ell+1} \quad (4.140)$$

$$= j_{\ell-1} - \frac{\ell+1}{x} j_\ell \quad (4.141)$$

Applying the recurrence relations we also find

$$\frac{d}{dx} [x^{\ell+1} j_\ell] = x^{\ell+1} j_{\ell-1}, \quad \frac{d}{dx} [x^{-\ell} j_\ell] = x^{-\ell} j_{\ell+1} \quad (4.142)$$

and repeated application leads to

$$j_\ell = (-x)^\ell \left(\frac{1}{x} \frac{d}{dx} \right)^\ell j_0 \quad (4.143)$$

Finally, an important integral which we will use regarding the Sachs-Wolfe effect is

$$\int_0^\infty dx x^{n-2} j_\ell^2(x) = 2^{n-4} \pi \frac{\Gamma(\ell + \frac{n-1}{2}) \Gamma(3-n)}{\Gamma(\ell + \frac{5-n}{2}) \Gamma^2(2 - \frac{n}{2})} \quad (4.144)$$

4.3.10 Relating the local temperature anisotropy to the angular power spectrum

We have seen how we can obtain the local temperature anisotropy $\Delta(\eta_0, k, \mu)$ today. First let's make the μ dependence more explicit: $\Delta(\eta_0, k, \mu) = \Delta(\eta_0, k, \hat{k}, \hat{p})$ (since $\mu = \hat{k} \cdot \hat{p}$). To obtain the power spectrum, the first step is to relate the observed temperature anisotropy from direction $\hat{n}$, i.e. $\Theta(\hat{n})$ to the local temperature anisotropy. The observed temperature anisotropy is observed today at η_0 and is observed here at $\vec{r}=0$. Thus

$$\Theta(\hat{n}) = \Theta(\eta_0, \vec{r}, \hat{n}) \Big|_{\vec{r}=0} \quad (4.145)$$

Now that we have introduced explicitly the functional dependence on position we can consider taking the Fourier transform. The Fourier transform of $\Theta(\eta_0, \vec{r}, \hat{n})$ is simply $\Theta(\eta_0, \vec{k}, \hat{p})$ where we identify the direction $\hat{n}$ with the direction of photon momentum $\hat{p}$, i.e.

$$\Theta(\eta_0, \vec{r}, \hat{n}) = \int \frac{d^3 k}{(2\pi)^3} e^{i\vec{k} \cdot \vec{r}} \Theta(\eta_0, k, \hat{k}, \hat{n}) \quad (4.146)$$

and so setting $\vec{r}=0$ we get

$$\Theta(\hat{n}) = \int \frac{d^3 k}{(2\pi)^3} \Theta(\eta_0, k, \hat{k}, \hat{n}) \quad (4.147)$$

How is $\Theta(\eta_0, k, \hat{k}, \hat{n})$ related to $\Delta(\eta_0, k, \hat{k}, \hat{n})$? Since $\Theta(\hat{n})$ is a random variable, then so is $\Theta(\eta_0, k, \hat{k}, \hat{n})$. We can express $\Theta(\eta_0, k, \hat{k}, \hat{n})$ in terms of a random variable which encapsulates the initial conditions

as set by inflation, namely $\xi(\vec{k})$ and a transfer function which propagates $\xi(\vec{k})$ from inflation to today to give us $\Theta(\eta_0, k, \hat{k}, \hat{n})$. The transfer function is none other than $\Delta(\eta_0, k, \hat{k}, \hat{n})$ so that the relation is

$$\Theta(\eta_0, k, \hat{k}, \hat{n}) = \Delta(\eta_0, k, \hat{k}, \hat{n}) \xi(\vec{k}) \quad (4.148)$$

To get the C_ℓ 's we need the correlation $\langle \Theta(\hat{n})\Theta(\hat{n}') \rangle$ as well as the correlation of the initial random variable $\xi(\vec{k})$:

$$\langle \xi(\vec{k})\xi(\vec{k}') \rangle = (2\pi)^3 P_0(k) \delta^{(3)}(\vec{k} - \vec{k}') \quad (4.149)$$

which (as we have done in the case of the matter power spectrum) depends on the initial power spectrum $P_0(k)$ as is given by inflation. With this in hand we proceed to relate the two correlations:

$$\begin{aligned} \langle \Theta(\hat{n})\Theta(\hat{n}') \rangle &= \int \frac{d^3k}{(2\pi)^3} \int \frac{d^3k'}{(2\pi)^3} \Delta(\eta_0, k, \hat{k}, \hat{n}) \Delta(\eta_0, k', \hat{k}', \hat{n}') \langle \xi(\vec{k})\xi(\vec{k}') \rangle \\ &= \int \frac{d^3k}{(2\pi)^3} \int d^3k' P_0(k) \delta^{(3)}(\vec{k} - \vec{k}') \Delta(\eta_0, k, \hat{k}, \hat{n}) \Delta(\eta_0, k', \hat{k}', \hat{n}') \\ &= \int \frac{d^3k}{(2\pi)^3} P_0(k) \Delta(\eta_0, k, \hat{k}, \hat{n}) \Delta(\eta_0, k, \hat{k}, \hat{n}') \end{aligned} \quad (4.150)$$

But $\langle \Theta(\hat{n})\Theta(\hat{n}') \rangle$ is related to C_ℓ via (4.62) so that

$$\frac{1}{4\pi} \sum_\ell (2\ell + 1) C_\ell P_\ell(\nu) = \int \frac{d^3k}{(2\pi)^3} P_0(k) \Delta(\eta_0, k, \hat{k}, \hat{n}) \Delta(\eta_0, k, \hat{k}, \hat{n}') \quad (4.151)$$

where $\nu = \hat{n} \cdot \hat{n}'$. We hit both sides by $P_L(\nu)$ and integrate over ν to get

$$C_\ell = \frac{1}{4\pi^2} \int d^3k P_0(k) \int d\nu P_\ell(\nu) \Delta(\eta_0, k, \hat{k}, \hat{n}) \Delta(\eta_0, k, \hat{k}, \hat{n}') \quad (4.152)$$

and following the calculation through we find

$$C_\ell = \frac{2}{\pi} \int dk k^2 P_0(k) |\tilde{\Delta}_\ell(\eta_0, k)|^2 \quad (4.153)$$

This is our final formula which relates the initial power spectrum $P_0(k)$ to the photon transfer functions $\Delta_\ell(k)$ which encapsulate the cosmological evolution after inflation. Notice that it has a form similar to the matter power spectrum, e.g. $\sim P_0|T(k)|^2$.

The procedure to calculate the C_ℓ 's can then be summarized as follows:

- Compute the sources Δ_0 , Ψ , Φ and u_b and their derivatives for a series of η values and k -values.
- Compute the line-of-sight integral (4.130) to obtain $\Delta_\ell(\eta_0, k)$. This basically converts η to ℓ .
- Compute the C_ℓ by integrating in k over $P_0(k)$ and $|\Delta_\ell|^2$ using (4.153).

4.3.11 The effective temperature and the ordinary Sachs-Wolfe effect

Now that we have found the solutions to the temperature anisotropies let's see how they look like on the C_ℓ plot.

First consider the anisotropies at last scattering surface: $\Delta_0 + \Psi - ik\mu u_b$. Furthermore let's concentrate on superhorizon scales. There the baryon velocity scales as η (see section ??) so for

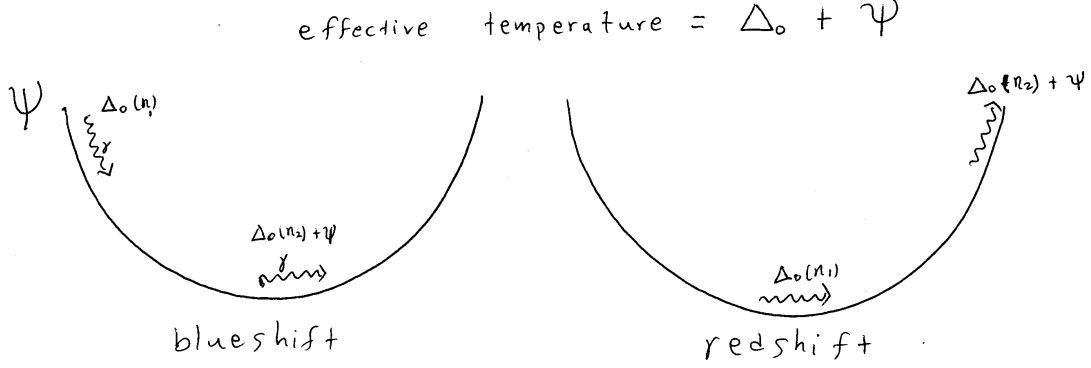


Figure 4.9: The effective temperature monopole. Photons may acquire or lose energy as they pass through potential wells via gravitational redshifting. This changes their temperature monopole from Δ_0 to $\Delta_0 + \Psi$. That is why the primary anisotropies contain $\Delta_0 + \Psi$ rather than Δ_0 in equation (4.114). The effective temperature at decoupling on superhorizon scales is what leads to the ordinary Sachs-Wolfe effect.

small η and superhorizon scales we may set it to zero. The other two terms are constant so that on super-horizon scales the primary anisotropies at decoupling are $(\Delta_0 + \Psi)(\eta_*, k) = \text{const}$ in both k and time. The term $\Delta_0 + \Psi$ has a well defined physical meaning. It is the effective photon temperature monopole. What is happening is shown in figure 4.9. A photon with initial monopole Δ_0 passes through a potential well and will acquire a redshift (lose energy) or blueshift (gain energy) depending on the potential difference. This redshift or blueshift changes the photon's energy and thus its temperature. The effective temperature after the gravitational redshift/blueshift is thus $\Delta_0 + \Psi$.

To proceed further we use the line-of-sight integral to find the anisotropies today and further assume that the potentials Φ and Ψ stay constant. This is a perfect assumption if the Universe is matter dominated but we shall return to it in the case of dark energy. Since Φ and Ψ are constant then the ISW term is zero and so the line-of-sight integral (4.130) gives

$$\Delta_\ell(\eta_0, k) = \int_0^{\eta_0} d\eta j_\ell[k(\eta_0 - \eta)]g(\eta)(\Delta_0 + \Psi) = j_\ell[k(\eta_0 - \eta_*)](\Delta_0 + \Psi)(\eta_*) \quad (4.154)$$

where in the 2nd equality we have assumed instantaneous decoupling. We then use this in the C_ℓ formula (4.153) to get

$$C_\ell^{SW} = \int dk k^2 P_0(k) \{j_\ell[k(\eta_0 - \eta_*)]\}^2 |\Delta_0 + \Psi|^2 \quad (4.155)$$

We now assume a primordial initial power spectrum $P_0 = Ak^{n-4}$ where n is the spectral index, and further set $x = k(\eta_0 - \eta_*)$ to get

$$C_\ell^{SW} = B \int dx x^{n-2} j_\ell^2(x) \quad (4.156)$$

But this integral can be performed in terms of Γ functions as in (??). Using (??) and further setting $n = 1$ for a scale-invariant spectrum, we find

$$\ell(\ell + 1)C_\ell^{SW} = \text{const} \quad (4.157)$$

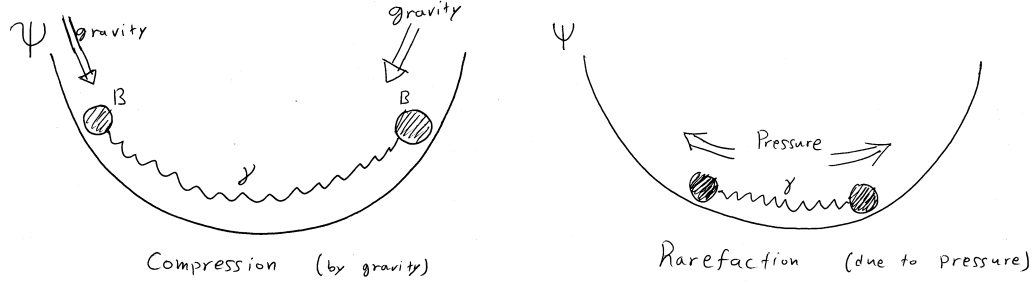


Figure 4.10: Acoustic oscillations: Left is an underdense region where pressure is minimal and so the gravitational force dominates and causes photons and baryons to start to compress. Right is an overdense region where pressure is maximal and so dominates over gravity leading to rarefaction.

This is the ordinary Sachs-Wolfe effect (Sachs and Wolfe 1967)⁷. This is the 2nd reason that we plot $\ell(\ell + 1)C_\ell$ rather than C_ℓ . Physically, the Sachs-Wolfe effect arises from the redshift (or blueshift) of photons as they pass through a gravitational potential well. It says that even if the initial photon temperature anisotropy was zero we would still see temperature anisotropy in the sky because of the redshifting (or blueshifting) of photons.

4.3.12 Acoustic oscillations

Now consider the tight-coupling solution to Δ_0 and furthermore let's set $R = 0$ so that

$$\Delta_0 + \Psi = \frac{\Phi_{sup}}{2} \cos\left(\frac{k\eta}{\sqrt{3}}\right) \quad (4.158)$$

This equation represents an oscillation but where did this oscillation come from? The answer comes from two places. The first comes from the approximations used to derive it: that photons are tightly coupled to the baryons due to Compton scattering. This tight coupling term introduces pressure in this photon-baryon fluid which resists compression. The second comes from the fact the Ψ appears in the above equation meaning that gravity plays a role. We have seen how Ψ is there due to the fact that photons acquire an effective temperature. But here its role is even bigger. It provides force opposite to pressure. As photons and baryons move through potential wells the density of photons changes leading to a change in temperature. At the bottom of the potential wells the compression due to gravity is maximal and photons are in an overdense region causing a temperature extreme. This is where pressure due to Compton scattering becomes maximum to force the photons to rarefact back into an underdense region. This leads to a series of acoustic oscillations which are oscillations in the density of photons (and baryons since they are coupled) and which in turn leads to oscillations in temperature. The two extreme cases are shown in figure 4.10.

This series of oscillations continues indefinitely until photons and baryons decouple at which point the pattern becomes frozen at the surface of last scattering. At that point oscillations in time will turn into oscillations in k . This is displayed in figure 4.11. The argument of the cosine is $k\eta/\sqrt{3}$ so the frequency of oscillation is set by k . Larger k means larger frequency, hence, more

⁷R.K. Sachs and A.M. Wolfe, "Perturbations of a cosmological model and angular variations of the Microwave Background", *Astrophys. J.* **147**, 73 (1967).

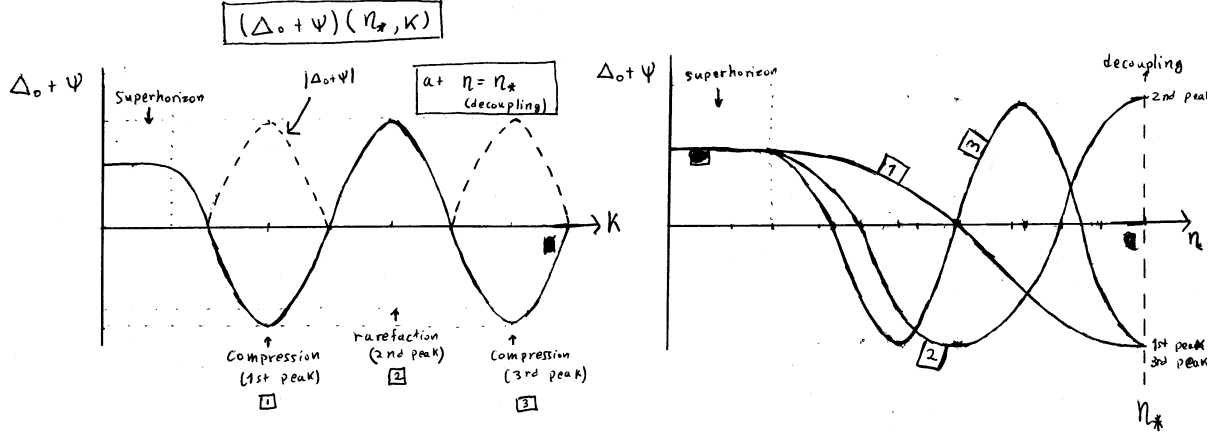


Figure 4.11: Left: Acoustic oscillations in $(\Delta_0 + \Psi)(k)$ at decoupling versus k showing the series of peaks seen in the power spectrum. Right: The oscillations versus time η showing how many oscillations elapse before the pattern freezes at decoupling. The labels "1"-"3" in the two panels correspond to each other.

oscillations. A wave comes from outside the horizon where $\Delta_0 + \Psi$ is constant and upon entering the horizon it has to compress under gravity. If the wavenumber is exactly right, the wave will undergo exactly half of an oscillation by decoupling and that would correspond to maximal compression, i.e. an over-density, at decoupling (wave "1"). Increasing the frequency further by choosing large k leads to a wave which has gone through a full oscillation (wave "2") at which point photons find themselves in an underdense region which corresponds to maximal rarefaction. This is another extreme point in the temperature. A third wave (wave "3") undergoes 1.5 oscillations and the final state at decoupling is again an over-density. However, to get the final C_ℓ we must square $\Delta_0 + \Psi$, hence, all of these extrema correspond to peaks in the C_ℓ power spectrum. Odd peaks correspond to over-densities and even peaks to under-densities.

4.3.13 The baryon drag

Let us now consider the full sub-horizon solution at decoupling given by (??) so let's rewrite it:

$$(\Delta_0 + \Psi)(\eta_*, k) = -R\Psi + \frac{\Phi_{sup}}{2} \cos(kr_s) \quad (4.159)$$

In this case we have switched on the baryons to contribute a non-zero density, i.e. $R \neq 0$. The first effect of non-zero R is first to change the sound speed to $\frac{1}{\sqrt{3(1+R)}}$ which is always smaller than the $R = 0$ case. This is because the baryons are heavy and by contributing an effective mass to the photon-baryon fluid, they reduce the effective speed of propagation in the plasma. Reducing the sound speed also reduces the sound-horizon as the photons now travel a shorter distance which in turn reduces the wavelength of the oscillation pattern seen in figure 4.11. The second effect is to introduce the term $-R\Psi$. Since Ψ is constant under our approximation, this has the effect of displacing the zero-point of the oscillator. To get the final pattern all you have to do is to lift the x -axis in both panels of figure 4.11. The result is shown in figure 4.12. The effect of shifting the zero-point of the oscillator is to make the odd extrema larger and the even extrema smaller. This

is translated in the angular power spectrum C_ℓ as making the odd peaks higher and the even peaks lower.

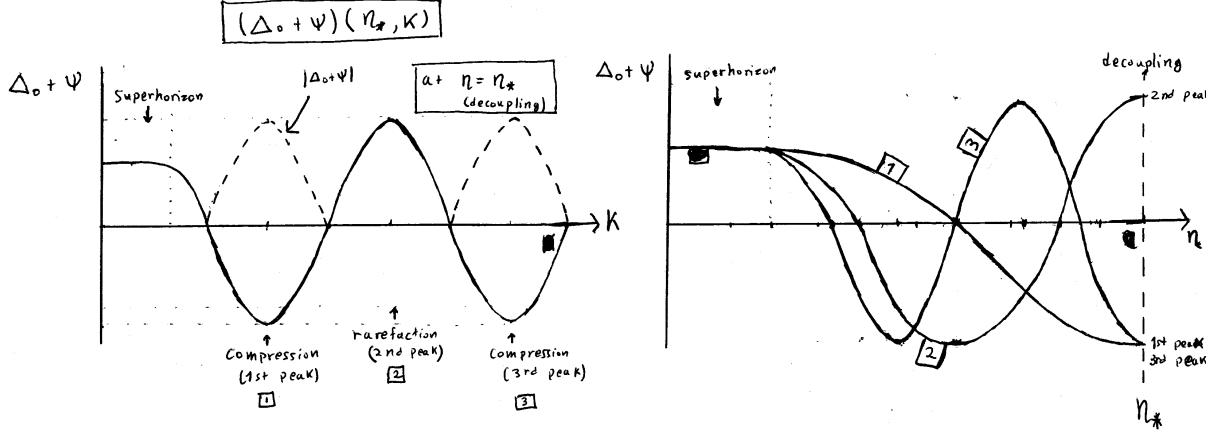


Figure 4.12: The baryon drag. Shifting the zero point of the oscillation due to the $-R\Psi$ term makes odd peaks larger and even peaks smaller. The effect is increased either by increasing the baryon density or by making Ψ bigger (e.g. via the addition of dark matter).

4.3.14 Photon-diffusion and Silk damping

To derive our tight-coupling equation and corresponding solution we assumed that $n_e\sigma_T$ is infinite. The effect of this is that photons and baryons move together. This is of course an approximation as in reality $n_e\sigma_T$ is finite (but large). Since $n_e\sigma_T$ is larger in the past then the approximation holds well for $\eta \ll \eta_*$, however, as we approach decoupling the approximation is bound to break down right before decoupling as $n_e\sigma_T$ approaches the Hubble rate H . The effect of finite $n_e\sigma_T$ is to introduce a drift of photons with respect to the baryons, i.e. the photons diffuse through an electron gas (remember photons most effectively scatter off electrons) in a random walk. This is shown in figure 4.13. Basically, the mean free path of the photons (i.e. the mean distance travelled between consecutive scatters) is $\lambda_{MFP} = \frac{1}{n_e\sigma_T}$. So in the limit $n_e\sigma_T \rightarrow \infty$ (complete tight-coupling) the mean free path goes to zero while in the free-streaming limit $n_e\sigma_T \rightarrow 0$ the mean free path goes to infinity (photons never scatter).

Now, during Hubble time H^{-1} the photons scatter $N = n_e\sigma_T H^{-1}$ times, i.e. the scattering rate $n_e\sigma_T$ times the time available to scatter H^{-1} . During a random walk the total distance travelled is equal to the mean free path times the square root of the number of steps taken (number of scatters). Hence, the total distance travelled is

$$\lambda_D \sim \lambda_{MFP} \sqrt{N} \sim \frac{1}{\sqrt{n_e\sigma_T H}} \quad (4.160)$$

and this length is called the diffusion length. The effect of this photon diffusion is to wash out anisotropies with wavelengths smaller than λ_D . This results into diffusion damping, also called Silk damping (J. Silk 1968)⁸. In k -space, Silk damping can be quantified by introducing a damping

⁸J. Silk, "Cosmic-Black-Body radiation and galaxy formation." *Astrophys. J.* **151**, 459 (1968)

Photon Diffusion

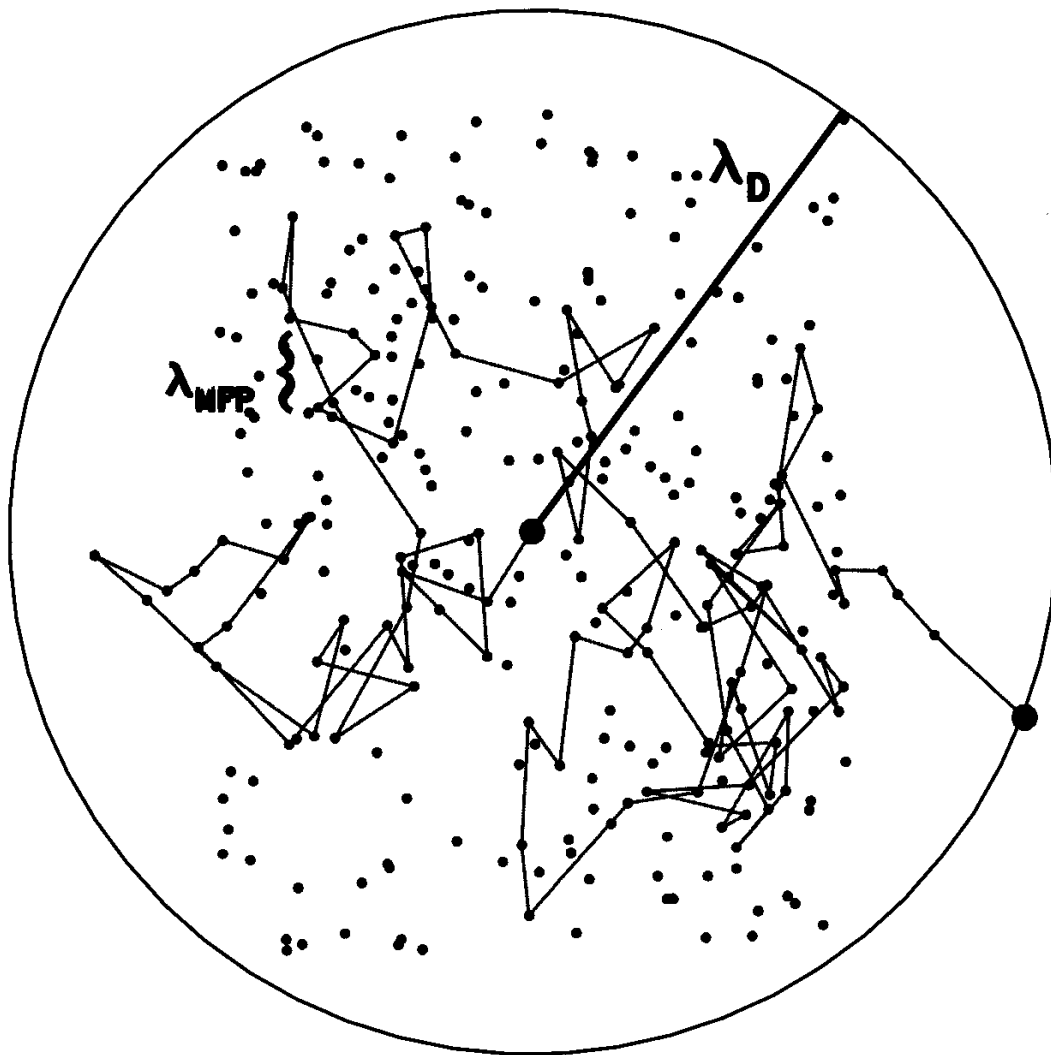


Figure 4.13: Photon diffusion through an electron gas. Points denote electrons. The broken scattered line represents a photon as it scatters off electrons in a random walk. The mean free path λ_{MFP} is the typical distance that a photon traverses between two consecutive scatterings. After a Hubble time the photon has scattered many times in different random directions so that the total distance travelled is of order the diffusion length λ_D . (taken from Dodelson, "Modern Cosmology").

coefficient k_D . Silk damping is not part of our tight-coupling approximation equations (4.102). To see the Silk damping we have to include the next order in the expansion in $\frac{1}{an_e\sigma_T}$, i.e. to go to 2nd order in the expansion. The calculation is long so here we only quote the answer. One finds that the damping coefficient is given by

$$\frac{1}{k_D^2(\eta)} = \int_0^\eta \frac{d\eta'}{6(1+R)an_e\sigma_T} \left[\frac{R^2}{1+R} + \frac{8}{9} \right] \quad (4.161)$$

where as you may notice the damping coefficient depends on the time at which is evaluated: $k_D = k_D(\eta)$.

Fortunately we can understand the damping effect without having to solve the equations to 2nd order. We simply take the undamped oscillator equations, i.e $-R\Psi + A\cos(kr_s)$ and multiply them by $\exp[-k/k_D]$ so that the solution which includes the damping is

$$(\Delta_0 + \Psi)(\eta, k) = [-R\Psi + A\cos(kr_s)] e^{-k/k_D} \quad (4.162)$$

Anisotropies with $k > k_D$ are thus exponentially damped. The result is shown in figure 4.15.

4.3.15 Acoustic driving

So far in our discussion we ignored the effect of time-varying gravitational potentials. If the potentials are time-varying then there is an additional forcing term appearing in the oscillator equation (4.102). This has the effect of increasing the oscillation amplitude for the modes relevant to the time where the potentials are time varying. Remember that potentials are oscillating (hence time varying) in the radiation era and are constant in the matter era. Thus if the last scattering surface was in complete matter domination, this effect would have been effectively absent as it would be due to much earlier times and therefore smaller scales (which are Silk damped). However, the last scattering surface is not far from the radiation to matter equality (only an order of magnitude in redshift) and so there are modes for which the potentials are still time varying to some extent. These modes correspond to modes around the 3rd peak (see figure 4.15).

This concludes our discussion of the primary anisotropies.

4.3.16 The Integrated Sachs-Wolfe effect

We now turn to the free-streaming regime. The free-streaming regime introduces a further term which appears in the line-of-sight integral. It is the Integrated Sachs-Wolfe term:

$$\Delta_\ell^{ISW}(\eta_0, k) \approx \int_{\eta_*}^{\eta_0} d\eta' j_\ell[k(\eta - \eta_0)] [\Phi' + \Psi'] \quad (4.163)$$

As for the driving effect, the ISW effect is non-zero only if the potentials are time-varying. The difference between the driving effect and the integrated Sachs-Wolfe effect is that the former is in the tight-coupling regime while the later is in the free-streaming regime. Physically the ISW effect is very similar to the ordinary Sachs-Wolfe effect, i.e. it is due to the redshifting or blueshifting of photons as they go through gravitational potential wells. The difference is that if the potential wells are time-varying, then the height of the potential well will change between the photon entering and the photon exiting the well. As the photon enters, it will acquire a blueshift as it travels to the bottom and a redshift as it travels back out of the potential well. If the potential well is constant

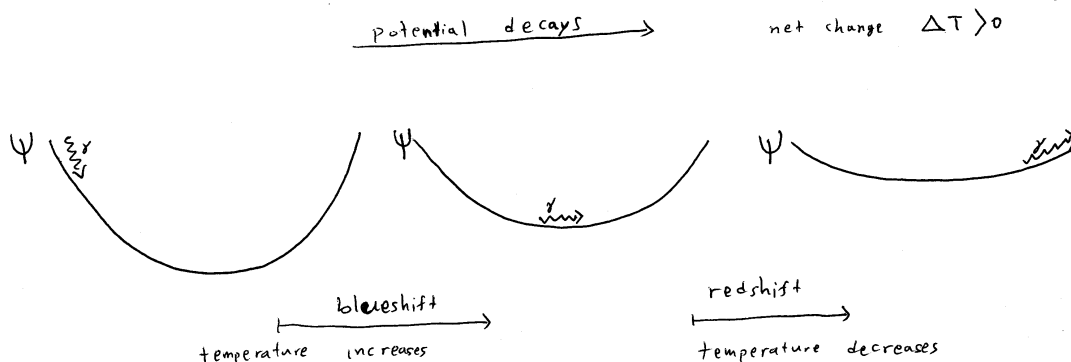


Figure 4.14: If the gravitational potential decays during the time that a photon enters and then leaves the potential, then net result is an increase in the temperature of the photon. This is the Integrated Sachs-Wolfe effect.

then the two effects cancel each other leaving no net effect. If however the potential wells are time-varying then there is a net change in energy of the photon which in turn results as a net change in the temperature. This is depicted in figure 4.14. The ISW effect is an integrated effect and so is projected on a wide range of scales, depending on the time it takes place. In Λ CDM cosmology we identify two cases of ISW. An early ISW effect occurs right after decoupling and is stronger if decoupling takes place closer to the radiation era as the potentials are still decaying and adjusting to their constant values during matter domination. The early ISW effect affects scales around the 1st and possibly 2nd peak. This has the effect of raising the 1st peak substantially higher. A late ISW effect can occur if the Universe departs from matter domination at late times. This may happen, for instance, if Λ comes to dominate at which point the potentials start to decay. This late ISW effect happens at low redshift (typically less than $z = 1 - 2$) and so is projected to large angular scales. It typically affects $\ell = 2 - 20$.

4.3.17 Projecting the primary anisotropies today

Having discussed the generation of the CMB anisotropies which are incorporated into the primary part $|\Delta_0 + \Psi|$ (plus the dipole of course, but we haven't discussed that) and the ISW effect we have to see how these are projected from k -space into ℓ space. Mathematically this is achieved via the spherical Bessel function in the line-of-sight integral. However, for small enough scales, (typically $\ell > 20$) we can resort to a very good approximation that does not require us to use the full Bessel function.

The spherical Bessel function for large ℓ has a peak which is approximately when its argument is equal to ℓ , i.e. when

$$k(\eta_0 - \eta_*) \sim \ell \quad (4.164)$$

This peak is substantially higher than the rest of the Bessel function, and so the Bessel function acts like a sort of delta-function. We can use this to convert k to ℓ so that we can perform the integral over k . Alternatively, for fixed k we can find which ℓ it mostly contributes to. Now let's look at the primary anisotropies. They have a cosine (monopole) or sine (dipole) solution with

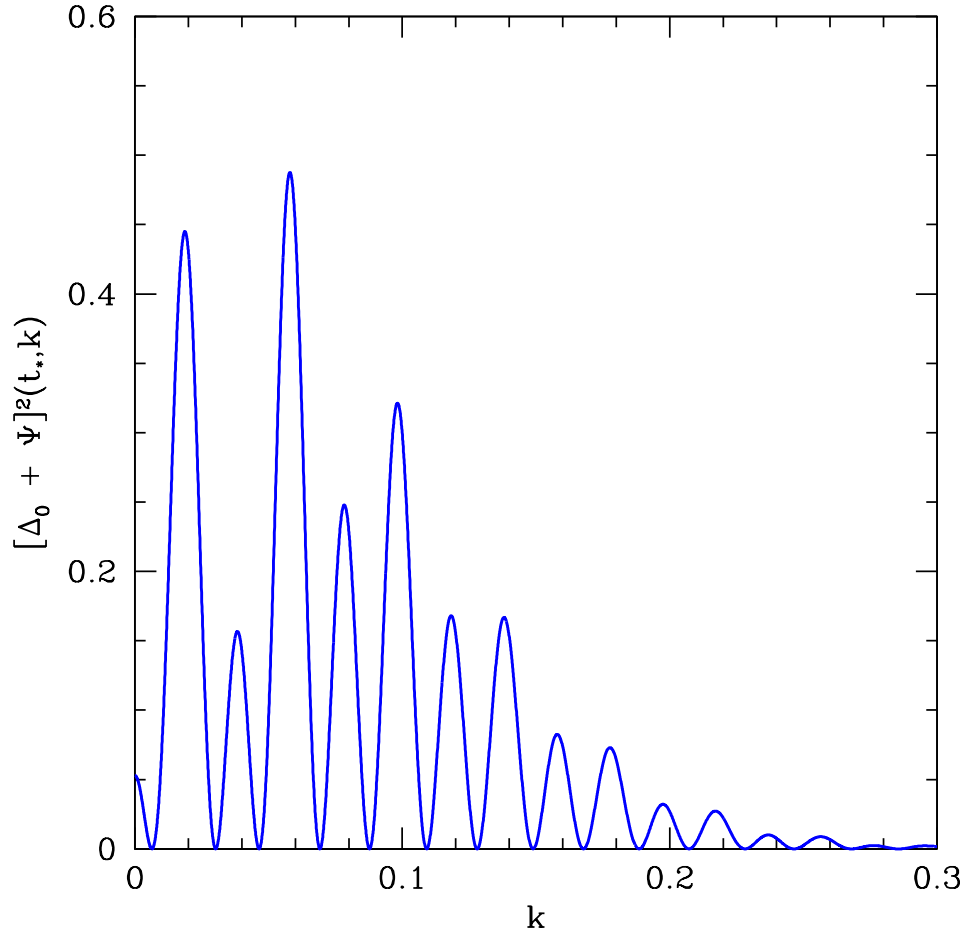


Figure 4.15: The anisotropies at decoupling, calculated numerically (so no approximation). You can see the alternating heights of the odd and even peaks due to the baryon drag and the Silk damping at large k . Acoustic driving increases the height of the 3rd peak over the rest.

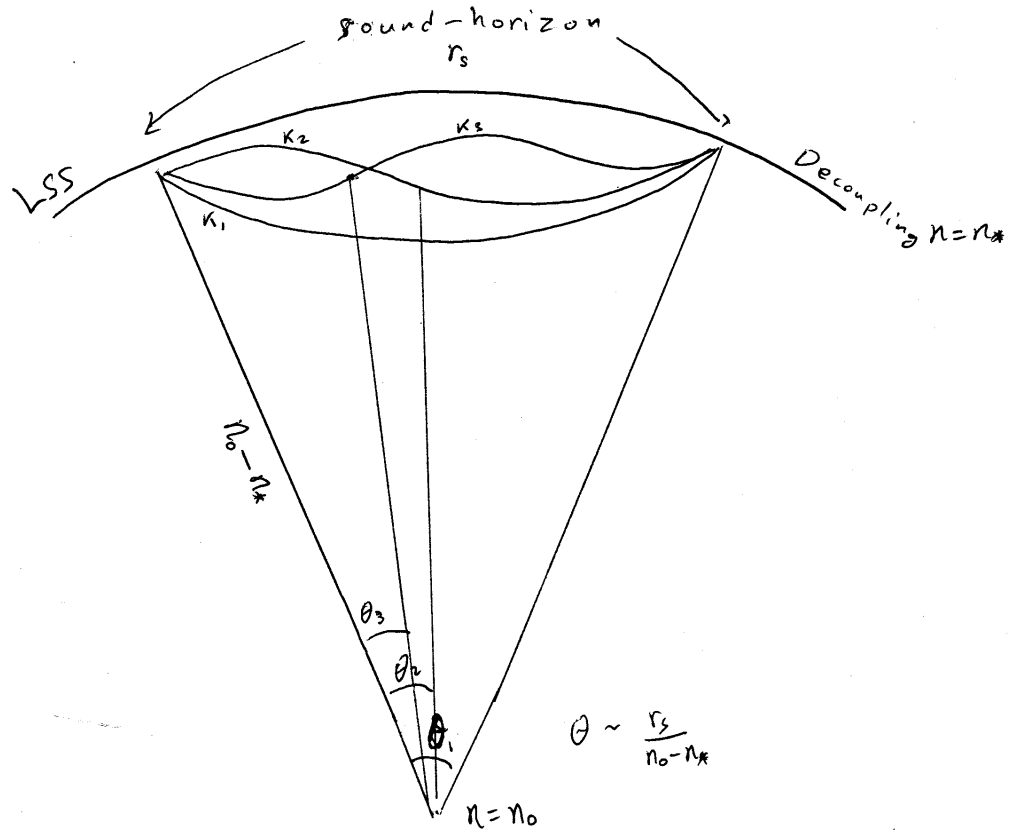


Figure 4.16: Projecting the sound-horizon to today. Different wavenumbers k project to different angles, hence to different ℓ .

frequency given by the sound horizon kr_s . Furthermore, to get the final C_ℓ we have to square Δ_ℓ so that a negative turning point becomes positive. The dipole is smaller than the monopole so that the major contribution comes from the monopole so that we expect to have peaks when $\cos(kr_s) = 1$ and troughs when $\cos(kr_s) = 0$. Thus the k value which contributes to a peak is

$$k_{peak} = \frac{n\pi}{r_s} \quad (4.165)$$

where n is an integer denoting which peak we are considering. For a trough we have a similar relation

$$k_{trough} = \frac{(2n+1)\pi}{2r_s} \quad (4.166)$$

Using this in (4.164) we get that

$$\ell_{peak} = \frac{n\pi(\eta_0 - \eta_*)}{r_s} \quad (4.167)$$

and

$$\ell_{trough} = \frac{(2n+1)\pi(\eta_0 - \eta_*)}{2r_s} \quad (4.168)$$

and furthermore the difference between two consecutive peaks or two consecutive troughs is

$$\Delta\ell = \frac{\pi(\eta_0 - \eta_*)}{r_s} \quad (4.169)$$

This is a very important result. It says that the peak structure of the CMB anisotropy measure the ratio of the angular diameter distance to last scattering surface, i.e. $\eta_0 - \eta_*$ to the sound horizon at decoupling. Both of these are background numbers that do not require to solve the perturbation equations to determine. We shall return to this when we discuss dark energy.

We can derive (4.169) in a more intuitive manner. The angle subtended by the sound horizon to day is (this angle is very small as $\eta_0 - \eta_*$ is large)

$$\theta \sim \frac{r_s}{\eta_0 - \eta_*} \quad (4.170)$$

However since $\ell \sim \frac{\pi}{\theta}$ we get (4.169) exactly!

4.3.18 Further effects from secondary anisotropies

There are a few more secondary effects on the CMB which we won't have time to go through. They are:

- **Reionization.** Reionization occurs when the first stars start to form, ionizing the medium around them. Eventually the whole Universe gets reionized around $z \sim 6$. This has the effect of making photon scatter off free electrons once more. On small scales this has a similar effect to diffusion and suppresses the C_ℓ 's by a factor $e^{-2\tau_0}$ where τ_0 is the optical depth at reionization. On large scales it may generate small anisotropies. These are not very observable in the temperature spectrum but are very important for polarization.
- **CMB weak lensing.** As photons travel through potential wells generated by galaxies they experience a further effect not included in the discussion so far. The potentials act like gravitational lenses, thus deflecting the photons from their original trajectory. It accounts for about 1–2% effect on the temperature spectrum and around 10% on the polarization spectrum. It makes the peaks shallower without changing their position.

- **Sunyaev-Zel'dovich effect.** The thermal SZ effect is due to the presence of ionized electrons in clusters which have a temperature different than the passing CMB photons. As the photons scatter with them, they re-thermalize at a different temperature, causing a distortion to the CMB spectrum. It is projected on small angular scales, around $\ell = 2000 - 3000$ where it is expected to be the dominant effect. The kinetic SZ effect (also called Ostriker-Vishniac) is due to the peculiar motions of the electrons relative to the photons and also leads to a spectral distortion of the CMB spectrum.

4.4 CMB polarization

4.4.1 Polarization from Compton scattering

The temperature anisotropies are not the only kind of anisotropies of the CMB spectrum. It turns out that the CMB is also polarized. The reason has to do with Compton scattering. The Compton scattering differential cross-section has an angular dependence.

4.4.2 Stokes parameters

To describe polarization it is convenient to introduce a set of parameters called Stokes parameters. They are I , Q , U and V . Consider an electromagnetic wave propagating in the z -direction. The wave has electric field along the x axis E_x and along the y axis E_y . We won't need the magnetic field for this construction. Since we are dealing with a wave, the E_x and E_y components of the electric field will have plane wave solutions

$$E_x = E e^{i(\omega_x t - k_x z)} \hat{x} \quad (4.171)$$

$$E_y = E e^{i(\omega_y t - k_y z)} \hat{y} \quad (4.172)$$

where E is a constant. The parameter I describes the total intensity of the wave, i.e.

$$I = |E_x|^2 + |E_y|^2 \quad (4.173)$$

This in turn is proportional to the temperature for a Planck spectrum. This is the part of the CMB that we have been dealing with so far.

The other three parameters describe pure polarization. They are defined as

$$Q = |E_x|^2 - |E_y|^2 \quad (4.174)$$

$$U = 2\text{Re}[E_x E_y] \quad (4.175)$$

$$V = 2\text{Im}[E_x E_y] \quad (4.176)$$

It turns out that V describes pure circular polarization while Q and U describe linear polarization. In particular Q describes polarization along the x and y axes while U describes polarization along axes at 45 degrees to x and y . These are shown in figure 4.18.

The Stokes parameters are very handy because we can measure them directly. Unfortunately they introduce an ambiguity in describing polarization: they depend on the orientation of the plane of polarization, i.e if the photon is coming along z , they depend on the orientation of the x and y axis. To be more precise, I and V are rotationally invariant while Q and U transform into each other as

$$\begin{pmatrix} Q' \\ U' \end{pmatrix} = \begin{pmatrix} \cos 2\theta & \sin 2\theta \\ -\sin 2\theta & \cos 2\theta \end{pmatrix} \begin{pmatrix} Q \\ U \end{pmatrix} \quad (4.177)$$

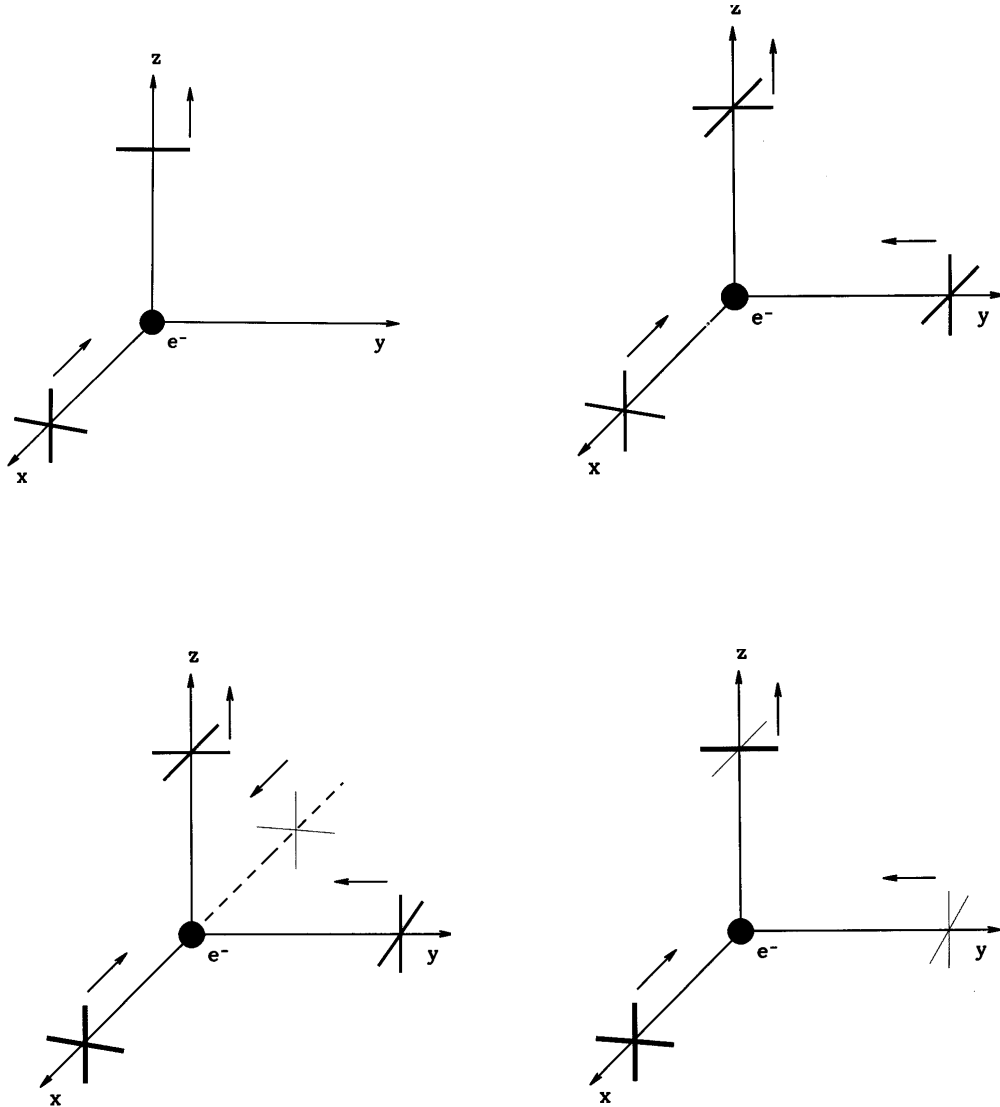


Figure 4.17: Top Left: an initially unpolarized photon moving along the x axis collides with an electron and subsequently moves along the z axis polarized in the y direction.

Top right: Monopole produces no polarization. Incident unpolarized radiation coming from both x and y directions produces NO polarization coming out of the z direction.

Bottom left: Dipole produces no polarization. Incident unpolarized hotter than average radiation (heavy line) coming from $+x$ axis meets unpolarized colder than average radiation coming from $-x$ direction (thus a dipole) meets average unpolarized radiation coming from the y -direction. The net result after scattering is unpolarized radiation in the z -direction.

Bottom right: Quadrupole produces polarization! Incident unpolarized radiation, hotter than average, coming from the x direction meets unpolarized radiation, colder than average, coming from the y -direction. The result after scattering is polarized radiation propagating in the z direction. It is hotter than average along the y -axis and colder than average along the x -axis.

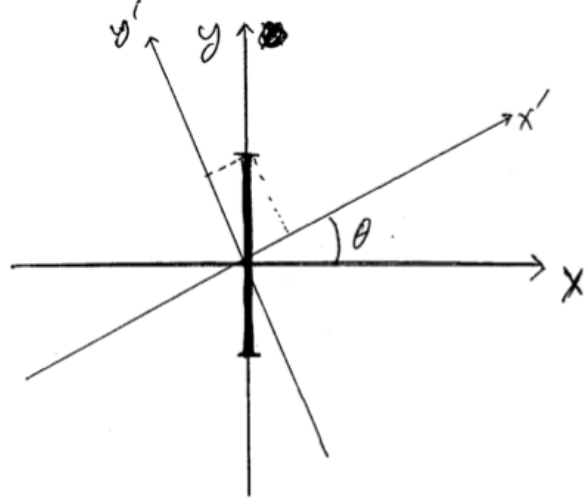
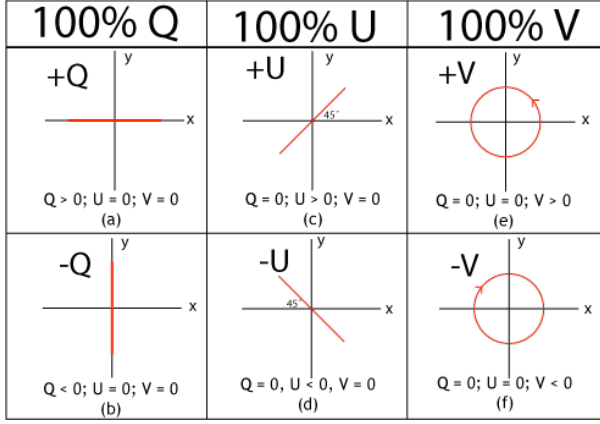


Figure 4.18: Left: The Stokes parameters. Right: Rotating x - y by an angle θ create new Q and U from an initial Q polarization.

where θ is the angle between the old and the new coordinate system. See fig. 4.18. Mathematically this means that Q and U form a spin-2 field.

In terms of the CMB, it turns out that Compton scattering cannot produce V -type polarization so we thus ignore this type.

4.4.3 E and B modes and polarization spectra

To quantify polarization without having to rely on a particular axis we can define two new polarization variables called E and B modes. The E and B modes are rotationally invariant and don't depend on the axis. The definition of E and B in terms of Q and U requires the use of spin-weighted spherical harmonics and a set of spin raising and spin lowering operators so we shall not consider their construction. Rather we shall describe how they look like. The E mode is the analog of a "grad" or "electric field" while the B mode is the analogue of a "curl" or "magnetic field". The E and B modes are shown in figure 4.19.

The E mode is even under parity while B is odd. You can see this by reflecting the four cases in figure 4.19 through an axis passing through the centre. The E modes transform to themselves while a positive B mode transforms into negative and vice versa. Since the intensity I is also even under parity, given the CMB spectrum we can decompose it into the following non-zero combinations of correlations: C_ℓ^{TT} , C_ℓ^{TE} , C_ℓ^{EE} and C_ℓ^{BB} .

The first detection of the TE cross correlation was done by the DASI experiment and subsequently detected by WMAP at higher accuracy and larger angular scales. The first detection of the EE auto-correlation was done by the CBI experiment and subsequently by WMAP. The only non-detected type of polarization is the pure BB mode. The types of spectra are shown in figure.

Before finishing, here are a few facts about CMB polarization

- Polarization does not have an ISW effect. Hence, the TE cross-correlation has half of the ISW

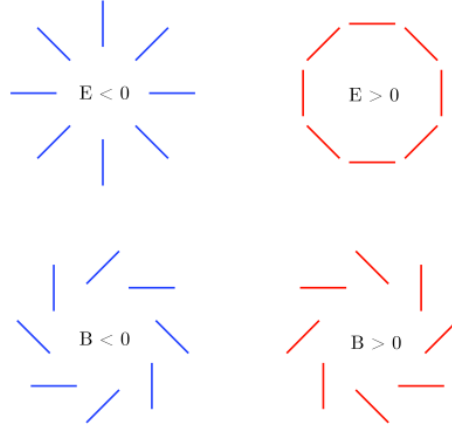


Figure 4.19: The E and B modes.

part coming from T while both EE and BB have no ISW part and should go to zero on large scales.

- Actually, due to reionization both EE and BB are expected to have significant amplitude on large scales since during reionization, Compton scattering regenerates the anisotropies on large scales.
- The BB -type of spectrum cannot be generated by scalar modes, except on small scales due to weak lensing of the E mode. On large scales the only signal in the BB spectrum comes from gravitational waves. Detecting the BB spectrum will give us direct information on the energy scale of inflation.

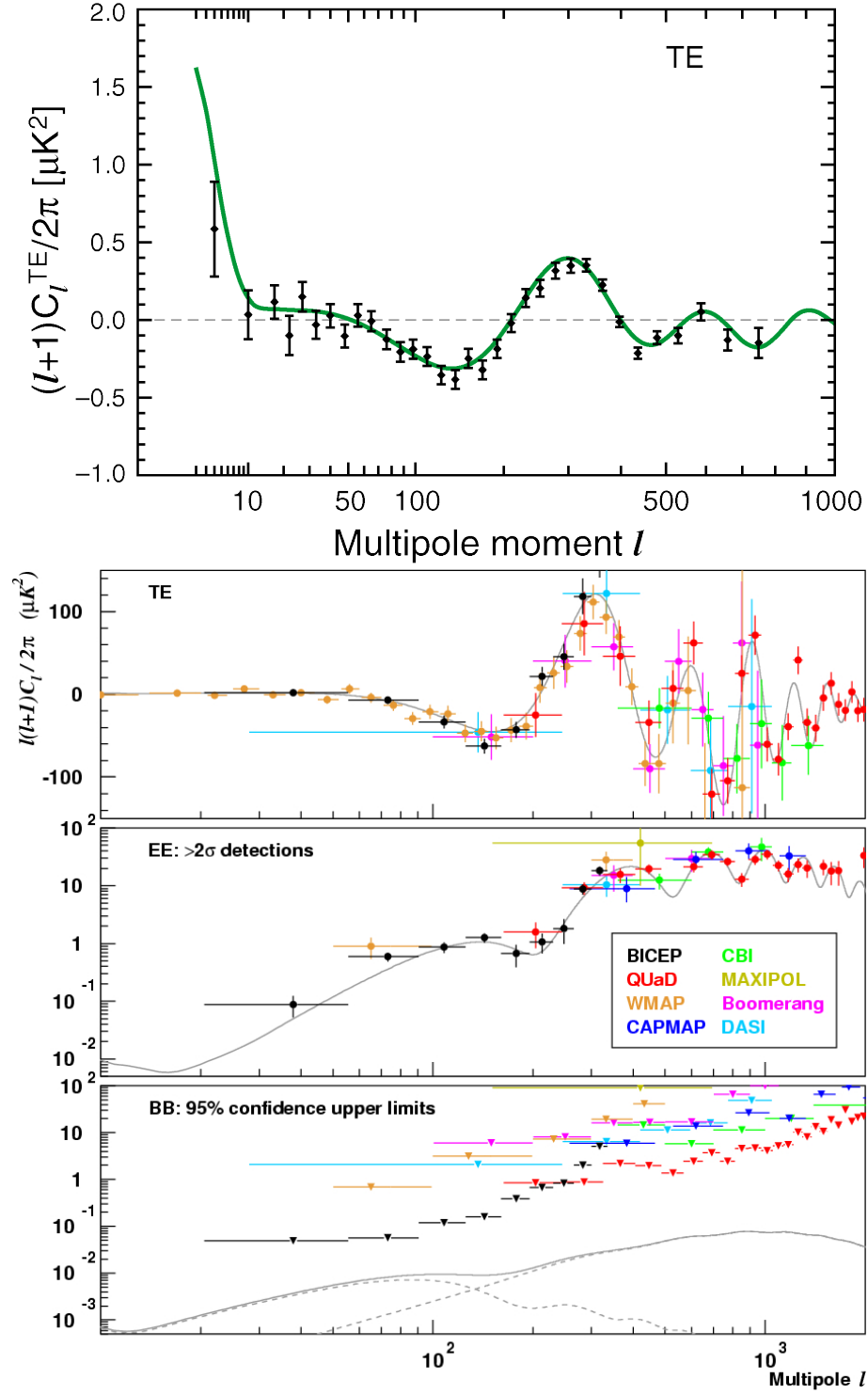


Figure 4.20: Top: A plot of $(\ell+1)C_\ell^{TE}$ versus ℓ as measured by WMAP after 7 years of data plotted with the best fit Λ CDM model. You can see the acoustic oscillations. The rise of the spectrum on large scales is due to reionization. Bottom: Polarization as measure by all experiments: TE , EE and upper limits on BB .

5 The Inflationary Universe

In many ways the Inflationary Universe can be considered as an add on to the Hot Big Bang. It was introduced in 1981 by Alan Guth (MIT) [*The inflationary universe: a possible solution to the horizon and flatness problems. Phys. Rev. D* **68** (2003) 103503] as a way to solve what was considered by many as the problems associated with the particular initial conditions associated with the HBB (homogeneity, isotropy and no defects). However, probably its biggest success was that it produced another incredibly important feature. Whilst the question of whether inflation solves the initial condition problem may be open to debate, very few people argue about the fact it provides an impressive way to generate primordial density perturbations. These perturbations have been observed in the anisotropies of the cosmic microwave background as measured by COBE and WMAP as well as other wonderful experiments. We will begin by discussing the problems with the HBB and how inflation tries to address them, and finish with examples of inflation and an introduction to how they generate structure in the universe.

5.1 Problems with the Hot Big Bang

There are a number of issues that the HBB simply can not address and have to adopt as initial conditions. These provided the original motivation for the inflationary cosmology, and we now turn our attention to these issues.

The flatness problem

From the Friedmann equation we can obtain the density parameter as a function of redshift (or scale factor). Starting with Eqn. (1.64) (and assuming $w = -1$ for the case of the vacuum type energy)

$$H^2(z) = H_0^2 (\Omega_{m0}(1+z)^3 + \Omega_{r0}(1+z)^4 + \Omega_{v0} - (\Omega_0 - 1)(1+z)^2) \quad (5.1)$$

then it follows using $\Omega = \frac{8\pi G\rho}{H^2}$ that

$$\Omega(z) = \frac{\Omega_{m0}(1+z)^3 + \Omega_{r0}(1+z)^4 + \Omega_{v0}}{\Omega_{m0}(1+z)^3 + \Omega_{r0}(1+z)^4 + \Omega_{v0} - (\Omega_0 - 1)(1+z)^2} \quad (5.2)$$

This tells us that if the total $\Omega_0 = 1$ today then it has always been unity. This is just a statement about the geometry of the universe, it can not change from a flat $k = 0$ situation to an open ($k < 0$) or closed ($k > 0$) case. However what about the case $\Omega_0 \neq 1$? It proves more convenient to go to scale factor representation, otherwise we have to start talking about $z \rightarrow -1$ as the scale factor gets very large compared to today's value. Setting $a_0 = 1$ for convenience here we have

$$\Omega(a) = \frac{\Omega_{m0}a^{-3} + \Omega_{r0}a^{-4} + \Omega_{v0}}{\Omega_{m0}a^{-3} + \Omega_{r0}a^{-4} + \Omega_{v0} - (\Omega_0 - 1)a^{-2}} \quad (5.3)$$

Now it is clear that $\Omega \rightarrow 1$ for large and small a as long as $\Omega_{v0} \neq 0$ in the former case and either of $\Omega_{r0} \neq 0$ or $\Omega_{m0} \neq 0$ in the latter. Infact without vacuum energy being present, $\Omega = 1$ is unstable as can be seen by dropping it from Eqn. (5.3) in the large a limit. So given that $\Omega \rightarrow 1$ for both large and small a , we can say it is an attractor no matter which way we go in time. Unfortunately this is a problem as far as the initial conditions are concerned. As we expect the early Universe to be

radiation dominated then (dropping the vacuum and matter components), and Taylor expanding Eqn. (5.3) we obtain

$$\Omega(a_{\text{init}}) \simeq 1 + \frac{(\Omega_0 - 1)}{\Omega_{r0}} a_{\text{init}}^2 \quad (5.4)$$

At the Planck scale we would have had $a_{\text{init}} \sim 10^{-32}$, implying that the universe must have already been flat to sixty powers of 10 ! Even at nucleosynthesis, an epoch we really think we understand where $a \sim 10^{-10}$, we would still have to fine tune to be within unity to one part in 20. Why so fine tuned when we might have expected to find $\Omega(a_{\text{init}}) - 1 \simeq O(1)$? A mechanism is required to explain why it had the value it appears to have had so early on.

The horizon problem

The Nobel prize winning observation by COBE that all cosmic microwave photons appear to be in thermal equilibrium at almost the same temperature is a puzzle? Why is it so isotropic? It is not difficult to see that in the HBB the Universe has not had enough time for different regions to reach a state of thermal equilibrium by today. The regions could not have interacted before the photons were emitted because of the finite horizon size,

$$\int_{t_*}^{t_{\text{dec}}} \frac{cdt}{a(t)} \ll \int_{t_{\text{dec}}}^{t_0} \frac{cdt}{a(t)}. \quad (5.5)$$

In other words, the distance light could travel before the microwave background was released is much smaller than the present horizon distance. In fact, any regions separated by more than about 2 degrees would be causally separated at decoupling in the hot big bang theory. This can be clearly seen in Figure. (5.1). In the big bang theory there is therefore no explanation of why the Universe appears so homogeneous.

The same argument that prevents the smoothing of the Universe also prevents the creation of irregularities. The COBE satellite detected irregularities in the CMB on all large angular scales (and Smoot was awarded the Nobel prize for that remarkable work in 2006), too large to be accounted for as emerging in the period between the big bang and the time of decoupling, because the horizon size at decoupling subtends only a degree or so. Hence these perturbations must have been part of the initial conditions.

The monopole problem

Modern particle theories predict a variety of ‘unwanted relics’, which can not be present today as they would have dramatically altered the evolution of the Universe. These include magnetic monopoles, domain walls, gravitinos and moduli fields associated with the extra dimensions arising in superstring theories. They are all massive particles created in the very early Universe but are diluted less rapidly than radiation as the Universe expands. hence they would rapidly come to dominate the dynamics, and lead to rapid closure of the Universe. We must eliminate them, while preserving the rest of the matter which we like.

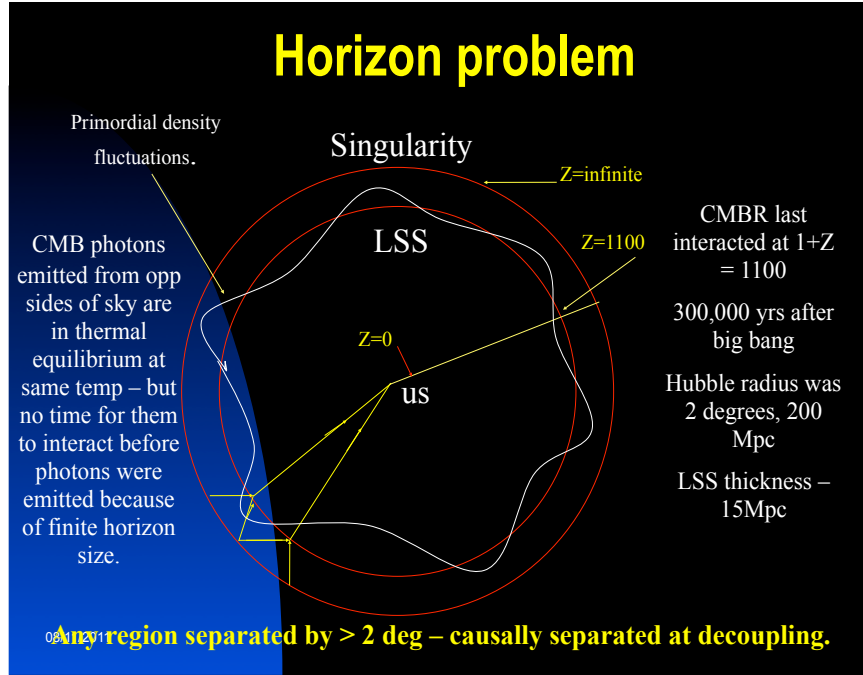


Figure 5.1: The horizon problem. CMB photons emitted from opposite sides of the sky are today in thermal equilibrium at the same temperature but they have not had enough time for them to interact since they were emitted because of the finite horizon size.

5.2 Enter Inflation – A.Guth 1981, A. Linde 1982

Inflation is defined to be any epoch where $\ddot{a} > 0$, an accelerated expansion. From the acceleration equation

$$\frac{\ddot{a}}{a} = -\frac{4\pi G}{3} \left(\rho + \frac{3p}{c^2} \right),$$

this corresponds to a negative pressure ($p < -\frac{\rho c^2}{3}$) and from the definition $H = \frac{\dot{a}}{a}$, we see that it also corresponds to $\frac{d(H^{-1}/a)}{dt} < 0$, i.e. the Hubble length as measured in comoving coordinates, *decreases* during inflation. At any other time, the comoving Hubble length increases. This is the key property of inflation; although typically the expansion of the Universe is very rapid, the crucial characteristic scale of the Universe is actually becoming smaller, when measured relative to that expansion.

We have already seen an example of an inflationary solution, the vacuum dominated regime $p = -\rho c^2$, has a solution given in Eqn. (1.61)

$$a(t) \propto \exp(Ht). \quad (5.6)$$

This is the famous **de Sitter solution** and as can be seen it means there is no singularity at $t = 0$. If this were the true case, in some sense there would be no HBB, no questions about what came before the bang, as the universe would only have zero size in the infinite past. There are many many more, a number of mixed radiation and vacuum as well as matter-vacuum solutions are

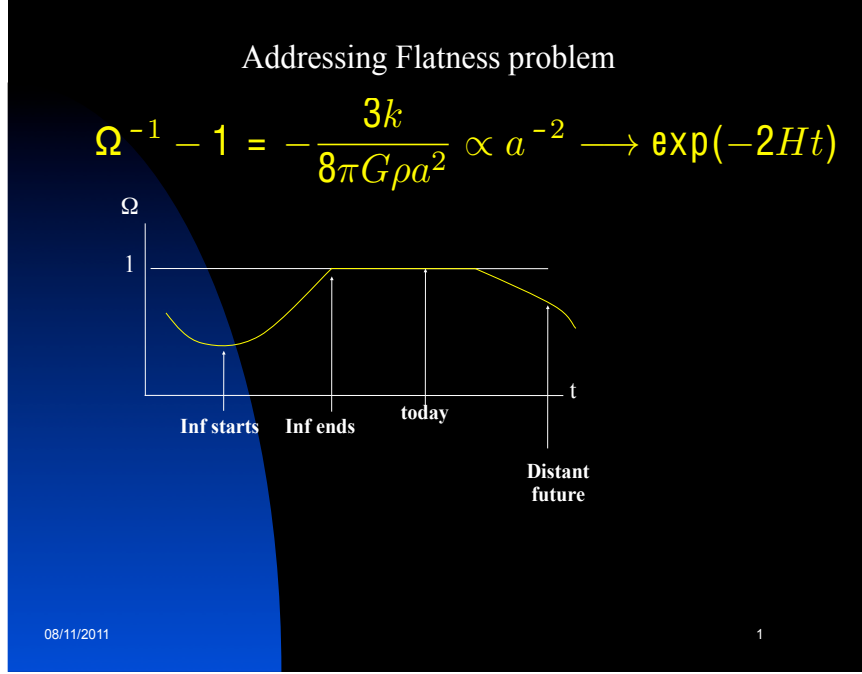


Figure 5.2: The flatness problem solved

derived in section 1.7! Of course, we know the HBB has many successes, and it is none inflating, so inflation can not last for ever, it must terminate and enter the HBB regime smoothly at some epoch. As it does so, the energy in the cosmological constant is converted into conventional matter through a process known as **reheating**. If inflation occurs early enough then none of the successes of the HBB are lost. Typical models of inflation have the epoch when inflation occurs being around $t_{\text{inf}} \sim 10^{-34}\text{sec}$ after the initial singularity, a time that is appropriate to the Grand Unified Theory (GUT) energy scale of $\sim 10^{16}\text{GeV}$ (recall $1\text{GeV} \equiv 10^9\text{eV}$).

The flatness problem

Inflation solves the flatness problem by rapidly forcing Ω towards unity rather than away from it. This is clear from the fact that the comoving Hubble length H^{-1}/a is decreasing. We require enough inflation to force Ω extremely close to unity to ensure that it will remain close to it today. Remember, as soon as we enter the HBB phase, $\Omega = 1$ is an unstable point. In particular we see that the Friedmann equation becomes

$$|\Omega(a) - 1| = \frac{|k|}{a^2 H^2} \propto \exp(-2Ht),, \quad (5.7)$$

and so it is $\Omega(a)$ which is forced to one, implying we are driven towards a universe that looks as if it is spatially flat ($k \simeq 0$).

Relic abundances

The rapid expansion of the inflationary stage rapidly dilutes the unwanted relic particles, because

the energy density during inflation falls off more slowly than the relic particle density. Very quickly their density becomes negligible. Of course they do not disappear totally and will one day re-enter the horizon – the ultimate in sweeping something under the carpet. We need to ensure that after inflation, the energy density of the Universe can be turned into conventional matter without recreating the unwanted relics. This *reheating* period must have a temperature that never gets hot enough to allow their thermal recreation. It will then allow for the particles we want to create and lead naturally into the standard HBB period, vital for the success of nucleosynthesis and the CMB.

The horizon problem and homogeneity

Inflation rapidly increases the size of any region of the Universe, but it keeps its characteristic scale, the Hubble scale fixed. So, a small patch of the Universe, small enough for thermalisation before inflation, can expand to a patch much larger than the size of our presently observable Universe. this ensures that all the cosmic microwave radiation are in thermal equilibrium. Moreover, it also allows for irregularities to be generated in the CMB, irregularities which would then evolve to form structures. We can rephrase the horizon solution by saying that because of inflation, light can travel much further before decoupling than it can afterwards. However, a word of warning. A number of leading cosmologists, led by the renowned Sir Roger Penrose don't buy this argument about the horizon problem. Another way of thinking about our universe is through entropy, that somewhat mystical thermodynamic property which tells us about the number of ways of realising an outcome. Basically the argument goes that if our universe underwent inflation, its entropy during the inflationary phase was substantially lower than it is today. Because a low-entropy state is less likely to be chosen randomly than a high-entropy one, inflation is unlikely to arise through randomly-chosen initial conditions – it is less likely than say the conditions for a standard HBB which has a high entropy state (see a recent article on this by Carroll and Chen *Gen.Rel.Grav.* 37 (2005) 1671-1674 or the extensive writing of Penrose in his book 'The Road to Reality : A Complete Guide to the Laws of the Universe' (Knopf 2005: ISBN: 0679454438). We won't go into this further, but point out that we are entering areas that are currently being hotly debated – which is fun !

The amount of inflation

The amount of inflation is normally specified by the *the number of e-foldings* N between some initial time and end time (which may or may not correspond to the beginning and end of inflation), given by

$$N \equiv \ln \frac{a(t_{\text{end}})}{a(t_{\text{initial}})} = \int_{t_i}^{t_e} H dt, \quad (5.8)$$

(To see that recall $H = \frac{\dot{a}}{a}$). We can estimate the amount of inflation required to solve the various cosmological problems. Consider the **flatness problem**. First we make a few plausible assumptions to ease the situation: inflation is of the exponential form (i.e. $a(t) \propto \exp(Ht)$) ending at $t = 10^{-34}$ sec, with the Universe immediately entering a radiation era which persists until today some 3×10^{17} sec later. Imagine also that today $|\Omega - 1| \leq 0.01$, a reasonable constraint on the value of Ω . Now during the radiation era, we have seen that, $|\Omega - 1| \propto t$, hence $|\Omega(10^{-34} \text{ sec}) - 1| \leq 3 \times 10^{-54}$. During inflation H is approximately constant, so $|\Omega - 1| \propto \frac{1}{a^2}$. From this it follows that in order to satisfy the constraint by the end of inflation, the scale factor has to grow during inflation by an

amount

$$\frac{a_{\text{tend}}}{a_{\text{tbegin}}} \sim 10^{27} \sim \exp(62), \quad (5.9)$$

corresponding to around 62 e-foldings. Although this looks large, inflation is typically so rapid that most inflation models give much more.

The cosmological constant

Unfortunately, a period of inflation says nothing about why the present value of the cosmological constant should be so small. In fact it should now be clear that inflation effectively relies on such a constant if only for a finite period of time. There is a very important point to be made here that may not be obvious at first.

5.3 Inflation and scalar fields

The most common framework in which inflation is obtained is based on the existence of scalar fields, in particular scalar field potentials. Some of you have already come across scalar fields such as the Higgs field in your lectures on the standard model. Needless to say, as far as particle physics is concerned they have remained elusive, although Dec 2011 brought some tantalising hints of the existence of a Higgs – time will tell but we really need them! They represent spin zero particles, transforming as a scalar (that is, it is unchanged) under coordinate transformations. It therefore only has one-degree of freedom, the scalar field is a number that varies with position but does not have any direction, unlike a vector like the electromagnetic field.

The traditional starting point for particle physics models is the action, which is an integral of the Lagrange density over space and time and from which the equations of motion can be obtained. A scalar field Lagrangian is like one for a particle, the difference between the kinetic energy and the potential energy of the field

$$L = -\frac{1}{2}(\partial_\mu \phi)(\partial^\mu \phi) - V(\phi), \quad (5.10)$$

where $\partial_\mu \phi \equiv \frac{\partial \phi}{\partial x^\mu}$ etc... The stress energy tensor is defined in terms of the matter Lagrangian

$$T_{\mu\nu} = -2 \frac{\partial L}{\partial g^{\mu\nu}} + g_{\mu\nu} L. \quad (5.11)$$

In this case the matter Lagrangian is that of the scalar field and we obtain

$$T_{\mu\nu} = (\partial_\mu \phi)(\partial_\nu \phi) + L g_{\mu\nu}, \quad (5.12)$$

where $g_{\mu\nu}$ is the metric tensor. If ϕ represents an isotropic fluid then we can write down the pressure and energy density from the definition

$$T^\mu_\nu = \text{diag}(-\rho, p, p, p), \quad (5.13)$$

from which we obtain for a homogeneous field (i.e. only dependent on time)

$$\rho_\phi = \frac{1}{2} \dot{\phi}^2 + V(\phi) \quad (5.14)$$

$$p_\phi = \frac{1}{2} \dot{\phi}^2 - V(\phi). \quad (5.15)$$

The potential energy $V(\phi)$ measures how much internal energy is associated with a particular field value. Normally, like all systems, scalar fields try to minimize this energy; however, a crucial ingredient which allows inflation is that scalar fields are not always very efficient at reaching this minimum energy state. In a given theory, there would be a specific form for the potential $V(\phi)$. However, we are not presently in a position where there is a well established fundamental theory that one can use, so, in the absence of such a theory, inflation workers tend to regard $V(\phi)$ as a function to be chosen arbitrarily, with different choices corresponding to different models of inflation. We will return to this in more detail shortly, but for now some example potentials are

$$V(\phi) = \lambda (\phi^2 - M^2)^2 \quad \text{Higgs potential} \quad (5.16)$$

$$V(\phi) = \frac{1}{2} m^2 \phi^2 \quad \text{Massive scalar field} \quad (5.17)$$

$$V(\phi) = \lambda \phi^4 \quad \text{Self-interacting scalar field} \quad (5.18)$$

5.4 Inflation dynamics

The equations for an expanding Universe containing *just* a homogeneous scalar field are easily obtained by substituting Eqs. (5.14) and (5.15) into the Friedmann equation (1.49) and fluid equations (1.56), giving

$$H^2 = \frac{8\pi G}{3} \left[V(\phi) + \frac{1}{2} \dot{\phi}^2 \right], \quad (5.19)$$

$$\ddot{\phi} + 3H\dot{\phi} + V'(\phi) = 0, \quad (5.20)$$

where prime indicates $d/d\phi$. Here we have ignored the curvature term k , since we know that by definition it will quickly become negligible once inflation starts. We could of course have included it and we would have seen that because it comes as k/a^2 in the Friedmann equation it would rapidly become negligible during inflation. Since

$$\ddot{a} > 0 \iff p < -\frac{\rho}{3} \iff \dot{\phi}^2 < V(\phi) \quad (5.21)$$

we will have inflation whenever the potential energy dominates. This should be possible provided the potential is flat enough, as the scalar field would then be expected to roll slowly. The potential should also have a minimum or some other feature which would allow inflation to end.

To solve these equations we usually use the **slow-roll approximation** (SRA), which assumes that a term can be neglected in each of the equations of motion to leave the simpler set

$$H^2 \simeq \frac{8\pi G}{3} V \quad (5.22)$$

$$3H\dot{\phi} \simeq -V' \quad (5.23)$$

The **slow-roll parameters**

$$\epsilon(\phi) \equiv \frac{1}{16\pi G} \left(\frac{V'}{V} \right)^2 \quad ; \quad \eta(\phi) \equiv \frac{1}{8\pi G} \frac{V''}{V}, \quad (5.24)$$

measures the slope of the potential (ϵ), and the curvature (η), and the necessary conditions for the slow-roll approximation to hold are

$$\epsilon \ll 1 \quad ; \quad |\eta| \ll 1. \quad (5.25)$$

As $\epsilon \rightarrow 1$ it is a mark of the end of inflation. To see this recall the requirement for inflation (or equivalently for acceleration) given in Eqn. (5.21) is $\dot{\phi}^2 < V(\phi)$. Looking at Eqn. (5.24) and using the slow roll equations (5.22) and (5.23) we see that

$$\epsilon(\phi) = 4\pi G \frac{\dot{\phi}^2}{H^2} = 3 \left(\frac{\dot{\phi}^2}{\dot{\phi}^2 + 2V(\phi)} \right) \quad (5.26)$$

where the final term has come from the Friedmann equation $H^2 = \frac{8\pi G}{3} \rho_\phi$ ⁹

Multiplying out we see that

$$\frac{\dot{\phi}^2}{V(\phi)} = \frac{2\epsilon(\phi)}{3 - \epsilon(\phi)} \quad (5.27)$$

from which it follows that for $\epsilon(\phi) \geq 1$ we have $\dot{\phi}^2 \geq V(\phi)$ hence inflation ends once that condition is reached.

5.5 Number of efolds

We showed earlier that any successful model of inflation requires of order 60 e-foldings of inflation in order to solve the flatness problem. The criteria of having slow roll inflation allows us to express this requirement directly in terms of the underlying particle physics motivated potential $V(\phi)$ and its derivative $dV/d\phi$, which is remarkable. To see this recall the definition of N-efolds given in Eqn. (5.8):

$$N \equiv \ln \frac{a(t_{\text{end}})}{a(t_{\text{initial}})} = \int_{t_i}^{t_e} H dt = \int_{\phi_i}^{\phi_e} \left(\frac{H^2}{H\dot{\phi}} \right) d\phi \quad (5.28)$$

where ϕ_i and ϕ_e are the values the inflaton field has at the start of the N-folds of inflation and end of inflation respectively. Assuming we are in the slow-roll regime, we can now use Eqns. (5.22) and (5.23) to obtain

$$N = -8\pi G \int_{\phi_i}^{\phi_e} \left(\frac{V(\phi)}{V'\phi} \right) d\phi. \quad (5.29)$$

This is important, what we have obtained is a relation between the number of e-folds of inflation and the underlying potential. Recall, we have said we require of order 60 e-foldings of inflation in order to address the issues associated with the initial conditions of the HBB such as the flatness and horizon problem. What Eqn. (5.29) allows us to do, is for any given underlying potential, to determine the initial value of ϕ_i required in order to obtain 60 efoldings. As will be seen from Eqn. (5.34) below, for chaotic large field inflation models setting $N = 60$ implies $\phi_i > m_{\text{pl}}$ (where $G = m_{\text{pl}}^{-2}$), in other words the initial value of ϕ must be greater than the Planck mass. Note the final value ϕ_f is obtained by setting $\epsilon = 1$ in Eqn. (5.24).

5.6 Some examples of Inflation: polynomial chaotic inflation

A particularly nice example of an inflaton potential is a simple polynomial potential. It could be a massive non-interacting field, $V(\phi) = m^2 \phi^2/2$ where m is the mass of the scalar field, or it could

⁹You may well question why we have used the full Friedmann equation here but the slow roll equation in the previous equation. This is because we are looking for the moment when the slow-roll term breaks down, in other words when the two terms become comparable in the energy density for the scalar field.

be a massless self-interacting field, $V(\phi) = \lambda\phi^4$, where λ is the self coupling of the field. Consider the first case. The slow-roll equations are

$$3H\dot{\phi} + m^2\phi = 0 \quad ; \quad H^2 = \frac{4\pi G m^2 \phi^2}{3}, \quad (5.30)$$

and the slow-roll parameters are

$$\epsilon = \eta = \frac{1}{4\pi G \phi^2}, \quad (5.31)$$

implying that inflation can proceed provided $|\phi| > 1/\sqrt{4\pi G}$, i.e. away from the minimum.

The solutions to the equations give

$$\phi(t) = \phi_i - \frac{m}{\sqrt{12\pi G}} t, \quad (5.32)$$

$$a(t) = a_i \exp \left[\sqrt{\frac{4\pi G}{3}} m \left(\phi_i t - \frac{m}{\sqrt{48\pi G}} t^2 \right) \right], \quad (5.33)$$

(where $\phi = \phi_i$ and $a = a_i$ at $t = 0$) and the total amount of inflation is

$$N_{\text{tot}} = 2\pi G \phi_i^2 - \frac{1}{2}. \quad (5.34)$$

An important thing to bear in mind is that we need to ensure that we are in a position where classical physics remains a valid approximation. This is simply the requirement $V \ll G^{-2}$, but it is still easy to get enough inflation provided m is small enough. In fact, m is required to be small from observational limits on the size of density perturbations produced.

As an exercise the reader may want to try and repeat the exercise for potential $V(\phi) = \lambda\phi^4$, assuming the field starts at $t = 0$ from rest rolling towards $\phi = 0$ from the positive side of the potential. Show that the slow roll equations give

$$\phi(t) = \phi_i \exp \left(-\sqrt{\frac{2\lambda}{3\pi G}} t \right), \quad (5.35)$$

$$a(t) = a_i \exp \left\{ \phi_i^2 \pi G \left[1 - \exp \left(-\sqrt{\frac{4\lambda}{3\pi G}} t \right) \right] \right\}, \quad (5.36)$$

(where $\phi = \phi_i$ and $a = a_i$ at $t = 0$) and the total amount of inflation is

$$N_{\text{tot}} = \pi G \phi_i^2 - 1. \quad (5.37)$$

From inflation to the HBB – reheating

The original inflation model of Guth was a small field model in which the potential had a false minimum at $\phi = 0$. The field would undergo a first order phase transition in order to tunnel to its true vacuum. However, although the model inflated, it predicted its own demise, too large inhomogeneities at the end of inflation. Why was this? First order transitions proceed via the nucleation of bubbles of true vacuum in a sea of false vacuum. The bubbles evolve satisfying causal physics, expanding at the speed of light. However, the intervening false vacuum regions of space continue growing exponentially fast, so unless the bubbles nucleate extremely close to each other

in space and time, they can never catch each other up and percolate thereby eliminating the false vacuum. In other words inflation can never end in this picture. The filling factor of the bubbles of true vacuum remains small at any time. This became known as the **graceful exit problem** and led to the development of alternative models which proceeded either via a second order transition or with out any transition at all, but in both cases the field evolving slowly and smoothly down its potential with no bubble nucleation resulting.

During inflation, all matter except the inflaton scalar field is redshifted to extremely low densities. **Reheating** is the process whereby the inflaton's energy density is converted back into conventional matter after inflation, re-entering the standard big bang theory, but in doing so having addressed the traditional problems associated with the HBB.

As the slow-roll conditions break down, ϕ evolves from being overdamped to being underdamped, moving rapidly on the Hubble timescale and oscillating at the bottom of the potential, where it decays into conventional matter. This is an active and technically demanding area of research and there has recently been something of a revolution in the way we think reheating takes place. Traditional treatments (e.g. as given in Kolb & Turner, *The Early Universe*, Addison-Wesley, , Redwood City, 1994) added a phenomenological decay term; this was constrained to be very small with reheating being inefficient. In particular there was a long time delay (redshifting) between the end of inflation and the Universe returning to thermal equilibrium; hence a low reheat temperature compared to the energy density at the end of inflation.

In **preheating**, this picture is turned on its head. Kofman et al *Phys.Rev.Lett.* **73** (1994) 3195-3198 showed that the decay can initially proceed through broad parametric resonance, with extremely efficient transfer of energy from the coherent oscillations of the inflaton field. The result is a very short reheating period, with most of the inflaton energy density at the end of inflation available for conversion into thermalized form. A higher reheat temperature is possible with some amazing possibilities, such as non-thermal phase transitions and baryogenesis occurring at the electroweak scale.

We should point out that these closing phases of inflation, the final sixty efoldings or so are the only ones of observational interest for us (although a number of authors are looking for affects from earlier efoldings). The models we have looked at can easily accommodate an infinite number of efoldings, but the scales that left our de Sitter horizon at those early times are now much larger than our own observable horizon which is of order cH_0^{-1} . From a particle physics stand point this observation that we are only really probing the physics associated with the final moments of the Inflaton's evolution is sobering. It means any information we glean from observations from features in the CMB and LSS will inform us about the evolution of the field over a few Planck lengths, hence only a small part of the underlying potential. However, all is not lost. As we will see, we can look for specific features in the data, evidence of deviations away from simple scaling laws in the power spectrum of density perturbations for instance, production of gravitational waves during inflation, deviations from Gaussian perturbations in the CMB, all of which can be determined in terms of the two slow roll parameters we have defined earlier. This will be a major part of the section on perturbations from inflation.

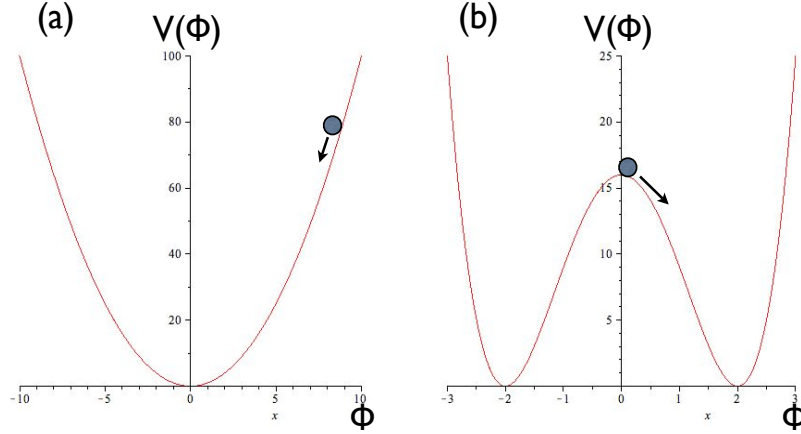


Figure 5.3: Two of the two types of single-field inflation models: (a) large-field inflation; (b) small-field inflation. Large field (or chaotic) models emerge from considering either mass-like or self interacting potentials, whereas small field (or new inflation) models are more like the Higgs potential.

5.7 Inflation models

There are a number of models on offer, some better motivated than others. A favourite way of differentiating them, at least for models driven by a single scalar field is by determining whether inflation occurs for **large** values of the scalar field as in type (a) of Figure. (5.3) or for small values as in type (b) of the same figure.

Chaotic inflation models (large field) were first proposed by Andrei Linde [*Chaotic Inflation, Phys. Lett. B* **129** (1983) 177] and are generically written in the form $V \propto \phi^{2\beta}$ where β is an integer. They are found in a number of situations and satisfy the slow roll conditions in the regime $\phi \gg \frac{1}{\sqrt{8\pi G}}$. The initial conditions place the field well up the potential, and these could be due to large fluctuations at the Planck era. The fact the initial values of the field are large in Planck units means that these models can lead to many e-folds of inflation, indeed they can lead to eternal inflation. Chaotic inflation ends when $\phi \sim \frac{1}{\sqrt{8\pi G}}$, and is followed by the inflaton field moving towards its vev of $\langle \phi \rangle = 0$, oscillating before it settles down. For the case of $V = \frac{1}{2}m^2\phi^2$ then the field equation (5.20) becomes

$$\ddot{\phi} + 3H\dot{\phi} + m^2\phi = 0 \quad (5.38)$$

If $m \gg H$ the Hubble drag is small and we have a solution where ϕ oscillates with angular frequency $\omega \sim m$ and with an amplitude which is damped by the Hubble drag. A very important feature emerges when we consider the energy density and pressure averaged over say one oscillation of the

scalar field about its vev. Denoting these averaged quantities with a bar we see that

$$\bar{\rho} = \frac{1}{T} \int_0^T dt \left(\frac{1}{2} \dot{\phi}^2 + V(\phi) \right) \quad (5.39)$$

and

$$\bar{p} = \frac{1}{T} \int_0^T dt \left(\frac{1}{2} \dot{\phi}^2 - V(\phi) \right) \quad (5.40)$$

where $T = \frac{2\pi}{m}$ is the period of oscillation. Now the oscillation solution for ϕ can be written as $\phi = \phi_0 \sin(mt)$ where ϕ_0 is the amplitude of oscillation, hence we find that the average energy density and pressure become $\bar{\rho} = \frac{1}{2} m^2 \phi_0^2$ and $\bar{p} = 0$. In other words the oscillating field corresponds to a nearly pressureless fluid, i.e. like dust ! There is a subtlety involved here which we need to check. Consistency with pressure-less fluid requires $\bar{\rho} \propto a^{-3}$ which implies that the amplitude of the oscillation must drop off as $a^{-3/2}$. This can be shown to be true by directly substituting into Eqn. (5.38) and using the WKB approximation which takes ϕ_0 to be a slowly varying function on the timescale of the oscillation.

New inflation models (small field) categorise a class of models based on inflation occurring for small values of the inflaton field. Examples include the original New inflation model based on Higgs field type potentials and also natural inflation models both of which have the gradients vanishing at the origin. In these models it is in principle possible for the field to remain at $\phi = 0$ for ever if it is placed there. This is equivalent to the field being trapped in a false vacuum as opposed to the true vacuum which is usually assumed to be at $V = 0$ (although that is convention and it need not be the case).

Alan Guth's original model in 1981 was of the small field type. Based on a first order potential, $\phi = 0$ corresponded to a false vacuum. Whilst the field was stuck there, the energy was dominated by the potential energy of the field and drove inflation. To end inflation the field would tunnel to the true vacuum but that brought with it a number of problems including how to gracefully end inflation. The problem was that bubbles of true vacuum would be produced, but these would be separated by regions of false vacuum which were still inflating. Therefore the bubbles could never percolate to reheat the universe. We were left with a very inhomogeneous universe, and this proved the downfall of the original Guth model. Another issue concerning small field inflation is the initial conditions required on the scalar field. Given that inflation occurs when the universe is hot say at the GUT scale, then we would normally expect the field to experience thermal fluctuations of order T_{GUT} , meaning that the potential should differ from its minima by an amount $V \sim T_{\text{GUT}}^4$. In other words making sure the potential is such that initially $\phi \simeq 0$ is a fine tuning issue, something inflation is meant to alleviate. For completeness below we provide some of the more popular examples of single field inflation models.

Polynomial chaotic inflation	$V(\phi) = \frac{1}{2} m^2 \phi^2$ $V(\phi) = \lambda \phi^4$
Power-law inflation	$V(\phi) = V_0 \exp\left(\sqrt{\frac{16\pi G}{3}} \phi\right)$
'Natural' inflation	$V(\phi) = V_0 [1 + \cos \frac{\phi}{f}]$
Intermediate inflation	$V(\phi) \propto \phi^{-\beta}$

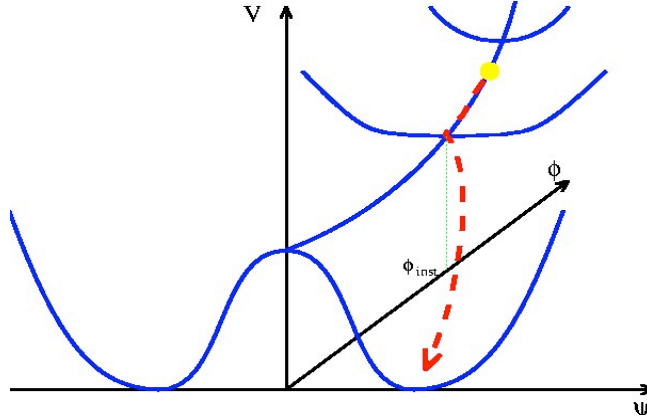


Figure 5.4: A typical Hybrid Inflation type potential with inflaton field ϕ and waterfall or defect field ψ

Note for the Power-law inflation case there is an exact solution of the form $a(t) \propto t^p$ which is not slow roll for $p > 1$. Of course in that case inflation never ends.

Hybrid inflation models are a very interesting class as they have more than one scalar field and appear to offer the possibility of occurring in particle physics contexts. An example shown schematically in Figure. (5.4) is one with a potential

$$V(\phi, \psi) = \frac{\lambda}{4} (\psi^2 - M^2)^2 + \frac{1}{2} m^2 \phi^2 + \frac{1}{2} \lambda' \phi^2 \psi^2, \quad (5.41)$$

where ϕ is the inflaton field and ψ the waterfall field whose destabilisation from $\psi = 0$ marks the end of inflation. When ϕ^2 is large, the minimum of the potential in the ψ -direction is at $\psi = 0$. The ϕ field slowly rolls down this ‘valley’ until it reaches $\phi_{\text{inst}}^2 = \lambda M^2 / \lambda'$, where the ‘waterfall field’ $\psi = 0$ becomes unstable and the ψ field rapidly rolls into one of the true minima at $\phi = 0$ and $\psi = \pm M$ ending inflation. Note the exact global symmetry $\psi \rightarrow -\psi$ which is broken spontaneously broken in the vacuum, but is restored for $\phi > \phi_{\text{inst}}$. The breaking of the symmetry implies that for suitable choices of the potential, topological defects could form at the end of a period of inflation, the end of hybrid inflation can then be regarded as a phase transition.

The couplings are supposed to satisfy

$$0 < \lambda \leq 1 \quad 0 < \lambda' \leq 1 \quad (5.42)$$

and we expect $M^2 G \ll 1$ meaning we expect the waterfall field vev to be less than the Planck scale. The inflaton field mass is constrained from the requirement that we have inflation, in particular by

demanding $\eta \ll 1$ where η is given by Eqn. (5.24)

$$\eta(\phi) \equiv \frac{1}{8\pi G} \frac{V''}{V} = \frac{1}{8\pi G} \frac{4m^2}{\lambda M^4} \ll 1 \quad (5.43)$$

or equivalently

$$\frac{m^2}{\lambda G M^4} \ll 1. \quad (5.44)$$

While in the ‘valley’, it is like a single field model with an effective potential for ϕ of the form

$$V_{\text{eff}}(\phi) = \frac{\lambda}{4} M^4 + \frac{1}{2} m^2 \phi^2. \quad (5.45)$$

The constant term would not normally be allowed as it would give a present-day cosmological constant. When it dominates, it allows both for the energy density during inflation to be much lower than normal while still giving suitably large density perturbations, and for ϕ to roll very slowly.

6 Primordial density perturbations produced during inflation

Having investigated a variety of models for inflation and seen how they address a number of the cosmological issues associated with the Hot Big Bang we now turn to derive probably the most outstanding success of the Inflationary paradigm, the initial scalar field quantum fluctuations which provide the seeds for structure formation. This is a demanding chapter as it relies on bringing together a number of features of classical and quantum theory, but it is worth the effort as we will end up linking the physics of the very large (cosmology) with the physics of the very small (particle physics) through the observations of fluctuations in the cosmic microwave background radiation.

We begin with a few preliminaries. Quantum Field Theory is usually based on the Heisenberg Picture although the Schrodinger picture is also widely used. We require a Hilbert Space, an infinite dimensional vector space such that at any given time each physical state corresponds to a state vector $|X\rangle$, normalised such that $\langle X|X\rangle = 1$ where $|X\rangle$ can include an arbitrary phase.

- Now each observable corresponds to a Hermitian operator $\hat{A}$ (recall these self-adjoint operators are defined by $\hat{A}^\dagger = \hat{A}$ and have the properties that all their eigenvalues are real, $\hat{A}|a_m\rangle = a_m|a_m\rangle$ with $a_m^* = a_m$, and their eigenvectors are orthonormal $\langle a_n|a_m\rangle = \delta_{mn}$).
- If an observable is measured when the state vector is $|X\rangle$, the probability of finding a particular value a_n is obtained from $\hat{A}|a_n\rangle = a_n|a_n\rangle$ to be $P = |\langle a_n|X\rangle|^2$.
- The expectation value of the observable A in the state $|X\rangle$ is given by $\langle X|\hat{A}|X\rangle$.
- Immediately after a value a_n has been found, the state vector is $|a_n\rangle$.

The time dependence of the system is determined from the Hamiltonian operator $\hat{H}(t, \hat{q}_1, \hat{p}_1, \hat{q}_2, \hat{p}_2, \dots)$ where the operators $\hat{q}_n$ correspond to the degrees of freedom and the operators $\hat{p}_n$ are their canonical conjugates. In the **Schrodinger picture** it is the degrees of freedom that are time-independent, while the state vectors satisfy the Schrodinger equation

$$\frac{d}{dt}|B\rangle = -i\hat{H}|B\rangle$$

For a time independent $\hat{H}$ the observable which is the energy is conserved. In general, the observable is conserved if its associated operator is time independent and commutes with $\hat{H}$ – this is Noether's theorem. The Schrodinger equation is equivalent to replacing $|B\rangle$ with the vacuum $|0\rangle$ given by

$$|B\rangle = \hat{U}(t)|0\rangle, \quad \dot{\hat{U}} = -i\hat{H}\hat{U}$$

where $\hat{U}$ is unitary (i.e. $U^\dagger = U^{-1}$). It follows that by replacing $|B\rangle \rightarrow |0\rangle = \hat{U}^{-1}(t)|B\rangle$, $\hat{A} \rightarrow \hat{U}^{-1}\hat{A}\hat{U}$, we have reassigned state vectors and operators to the physical states and observables in a way that now makes the state vectors time independent and imposes the time dependence on the observables. This is the **Heisenberg picture** and is used more often when quantising field theories. In this picture the operator $\hat{A}(q_n, p_n, t)$ satisfies

$$\frac{d\hat{A}}{dt} = i[\hat{H}, \hat{A}] + \frac{\partial \hat{A}}{\partial t} \quad (6.1)$$

where $[\hat{H}, \hat{A}]$ is the commutator. Setting $\hat{H} = \hat{A}$ it follows that $\hat{H}$ is time independent in the Heisenberg picture if and only if it is time independent in the Schrodinger picture (which is when $\frac{\partial \hat{A}}{\partial t} = 0$). In the Heisenberg picture we start with the Lagrangian operator $L(t, \hat{q}_1, \dot{\hat{q}}_1, \hat{q}_2, \dot{\hat{q}}_2, \dots)$ and from it derive the Hamiltonian operator, in a manner similar to the case of classical mechanics. The degrees of freedom therefore satisfy

$$\dot{\hat{q}}_n = \frac{\partial \hat{H}}{\partial \hat{p}_n}, \quad \dot{\hat{p}}_n = -\frac{\partial \hat{H}}{\partial \hat{q}_n}. \quad (6.2)$$

Now there is a consistency relation that needs to be satisfied by $\hat{q}_n$ and $\hat{p}_n$ and this follows from noting that since we have $\hat{A}(q_n, p_n, t)$, then

$$\frac{d\hat{A}}{dt} = \sum_n \left(\frac{\partial \hat{A}}{\partial \hat{q}_n} \frac{\partial \hat{q}_n}{\partial t} + \frac{\partial \hat{A}}{\partial \hat{p}_n} \frac{\partial \hat{p}_n}{\partial t} + \frac{\partial \hat{A}}{\partial t} \right) \quad (6.3)$$

Comparing Eqns. (6.3) and (6.1) and using Eqn. (6.2) we see that compatibility of the two requires

$$\frac{\partial \hat{H}}{\partial \hat{p}_n} = i[\hat{H}, \hat{q}_n], \quad \frac{\partial \hat{H}}{\partial \hat{q}_n} = -i[\hat{H}, \hat{p}_n] \quad (6.4)$$

The two equations follow from setting $\hat{A} = \hat{p}_n$ and then $\hat{A} = \hat{q}_n$ in Eqns. (6.1) and (6.3). Note that the $\hat{q}_n$ and $\hat{p}_n$ can not all commute, if they did the right hand side of Eqn. (6.4) would vanish. In particular it means given an operator $\hat{A}$ in terms of the $\hat{q}_n$ and $\hat{p}_n$, the order they appear is important because qp is not the same as pq !

In many situations the quantum theory is obtained from the classical theory simply by promoting the classical degrees of freedom q_n (and their conjugate momenta p_n) to operators. This is the case in classical mechanics and bosonic fields in a quantum field theory. Given that, it should be expected that the quantum theory has a classical limit. This is when the states $|A\rangle$ have q_n which are sharply defined values during the time period under consideration. We will concentrate on the case of scalar fields, the simplest of all the bosonic fields.

6.1 Harmonic oscillators

We will be interested in quantum theories which are equivalent to a set of harmonic oscillators, as it describes the Fourier components of a free scalar field. Given that lets first of all review the case of a single SHO, which in turn requires a brief review of Lagrangians. In classical mechanics a system is defined by its Lagrangian $L(q, \dot{q}, t)$. The dynamics of the system of particles are determined by solving the Euler-Lagrange equations

$$\frac{\partial L}{\partial q} - \frac{d}{dt} \frac{\partial L}{\partial \dot{q}} = 0$$

or equivalently from the associated Hamiltonian:

$$H(q, p, t) = p\dot{q}(q, p, t) - L(q, \dot{q}(q, p, t), t)$$

via

$$\dot{q} = \frac{\partial H}{\partial p} \quad \dot{p} = -\frac{\partial H}{\partial q}$$

where $p \equiv \frac{\partial L}{\partial \dot{q}}$ is the canonical momentum. As an example, for a particle of unit mass moving in one spatial dimension we have $L = \frac{1}{2}\dot{q}^2 - V(q)$, hence $H = \frac{1}{2}p^2 + V(q)$ and $p = \dot{q}$. The E-L equation of motion is nothing other than Newton's famous acceleration equation

$$\ddot{q} + \frac{dV}{dq} = 0.$$

For the case of a Simple Harmonic Oscillator oscillating with frequency ω , $V(q) = \frac{1}{2}\omega^2 q^2$ where, and the E-L equation becomes

$$\ddot{q} + \omega^2 q = 0. \quad (6.5)$$

The general solution can be written as

$$q = \frac{1}{\sqrt{2\omega}}(ae^{-i\omega t} + a^*e^{i\omega t}) \quad (6.6)$$

and the Hamiltonian becomes $H = \omega|a|^2$.

Following the procedure just described we now promote the classical position q to an operator. The Lagrangian, Hamiltonian and canonical momentum become

$$L = \frac{1}{2}\dot{\hat{q}}^2 - \frac{1}{2}\omega^2 \hat{q}^2, \quad \hat{H} = \frac{1}{2}\hat{p}^2 + \frac{1}{2}\omega^2 \hat{q}^2, \quad \hat{p} = \dot{\hat{q}}. \quad (6.7)$$

Direct substitution of the Hamiltonian $H = \frac{1}{2}p^2 + \frac{1}{2}\omega^2 q^2$ into the consistency condition Eqn. (6.4) leads to the canonical commutation relation

$$[\hat{q}, \hat{p}] = i \quad (6.8)$$

explicitly demonstrating that q and p don't commute. To see this consider $\frac{\partial \hat{H}}{\partial \hat{p}_n} = i[\hat{H}, \hat{q}_n]$. The LHS is just (dropping hats) p . The RHS is

$$\begin{aligned} i[H, q] &= i\frac{1}{2}[(p^2 + \omega^2 q^2), q] = i\frac{1}{2}[p^2, q] = i\frac{1}{2}[p^2 q - qp^2] \\ &= i\frac{1}{2}[p(pq - qp) + pqp - (qp - pq)p - pqp] = i\frac{1}{2}[p[p, q] + [p, q]p], \end{aligned} \quad (6.9)$$

hence for the RHS and LHS to be the same we require $[p, q] = -i$. Comparing with the classical solution Eqn. (6.6) we define an operator $\hat{a}$ such that we can write the operator solution as

$$\hat{q}(t) = \frac{1}{\sqrt{2\omega}}(\hat{a}e^{-i\omega t} + \hat{a}^\dagger e^{i\omega t}) \quad (6.10)$$

The condition Eqn. (6.8) then becomes

$$[\hat{a}, \hat{a}^\dagger] = 1 \quad (6.11)$$

The Hamiltonian follows (recall $\hat{p} = \dot{\hat{q}}$)

$$\hat{H} = \frac{1}{2}(\hat{a}^\dagger \hat{a} + \hat{a} \hat{a}^\dagger) = \left(\hat{a}^\dagger \hat{a} + \frac{1}{2}\right), \quad (6.12)$$

which becomes

$$\hat{H} = \left(\hat{n} + \frac{1}{2}\right), \quad \text{where } \hat{n} \equiv \hat{a}^\dagger \hat{a} \quad (6.13)$$

with $\hat{n}$ being the occupation number. The eigenvalues of $\hat{H}$ (or $\hat{n}$) give the energy levels of the harmonic oscillator. In the Schrodinger picture, we solve the Schrodinger equation $(\frac{d}{dt}|A\rangle = -i\hat{H}|A\rangle)$ where the degrees of freedom are time independent, to obtain the energy levels $(n + \frac{1}{2})$ ($n > 0$), as seen above :

$$\hat{H}|n\rangle = (n + \frac{1}{2})\omega|n\rangle, \quad \text{where } \hat{n}|n\rangle = n|n\rangle \quad (6.14)$$

Recall, $\hat{H}$ is Hermitian, therefore the eigenvalues are real and the eigenvectors which are orthogonal can be chosen to be orthonormal, $\langle n|m\rangle = \delta_{nm}$. In that sense they provide a basis for the required Hilbert space of quantum theory.

How does the same result arise in the Heisenberg picture? In that case the dynamics is in the degrees of freedom and the state vectors are time independent. We first define the ground state by

$$\hat{a}|0\rangle = 0|0\rangle \quad (6.15)$$

and then build up the states with particles present via $|1\rangle \equiv \hat{a}^\dagger|0\rangle$, leading to the entire basis $\hat{a}^\dagger|n\rangle = \sqrt{n+1}|n+1\rangle$. Given this set of oscillators the Hamiltonian is

$$\hat{H} = \sum_i \left(\hat{n}_i + \frac{1}{2}\right) \omega_i \quad (6.16)$$

with $\hat{n}_i \equiv \hat{a}_i^\dagger \hat{a}_i$ and the canonical commutation relations are

$$[\hat{a}_i, \hat{a}_j^\dagger] = \delta_{ij}, \quad [\hat{a}_i, \hat{a}_j] = 0 \quad (6.17)$$

as previously derived.

6.2 Quantised free scalar field

Having discussed the case of a regular quantum mechanical oscillator with a finite number of degrees of freedom, we now turn to discuss the quantisation of a real scalar field operator. The Lagrangian density is given by

$$\mathcal{L} = -\frac{1}{2}(\partial_\mu \hat{\phi})(\partial^\mu \hat{\phi}) - \frac{1}{2}m^2 \hat{\phi}^2, \quad (6.18)$$

The Hamiltonian density (that is that Hamiltonian per unit spatial volume, i.e. $H = \int d^3x \mathcal{H}$ is given by

$$\mathcal{H} \equiv \Pi \dot{\hat{\phi}} - \mathcal{L} \quad (6.19)$$

where $\Pi \equiv \frac{\partial \mathcal{L}}{\partial \dot{\hat{\phi}}}$.

Now we know that for the harmonic oscillator the solution of the field equation is a sum of plane waves. This implies that the Hamiltonian will be a sum of harmonic oscillator Hamiltonians. It will prove useful to calculate these in terms of Fourier components where we are considering the system to be contained in a box of side L :

$$\begin{aligned} \hat{\phi}(\mathbf{x}, t) &= L^{-3} \sum_{\mathbf{k}} [\phi_k(t) \hat{a}_{\mathbf{k}} e^{i\mathbf{k} \cdot \mathbf{x}} + \phi_k^*(t) \hat{a}_{\mathbf{k}}^\dagger e^{-i\mathbf{k} \cdot \mathbf{x}}] \\ &= L^{-3} \sum_{\mathbf{k}} [\phi_k(t) \hat{a}_{\mathbf{k}} + \phi_k^*(t) \hat{a}_{-\mathbf{k}}^\dagger] e^{i\mathbf{k} \cdot \mathbf{x}} \end{aligned} \quad (6.20)$$

Now the mode function $\phi_k(t)$ depends only on k . The solution satisfies the wave equation

$$\square \phi + \frac{dV}{d\phi} = 0 \quad (6.21)$$

$$\square \phi + m^2 \phi = 0 \quad (6.22)$$

where here the potential is given by $V(\phi) = \frac{1}{2}m^2\phi^2$ and in a general curved space $\square \phi = \nabla_\mu \nabla^\mu \phi = \frac{1}{\sqrt{-g}} \partial_\mu (\sqrt{-g} g^{\mu\nu} \partial_\nu \phi)$. In flat space $g_{\mu\nu} = \eta_{\mu\nu}$ and we obtain for the Fourier mode $\phi_k(t)$

$$\begin{aligned} \ddot{\phi}_k - \nabla^2 \phi_k + m^2 \phi_k &= 0 \\ \ddot{\phi}_k + (\mathbf{k}^2 + m^2) \phi_k &= 0 \end{aligned} \quad (6.23)$$

Introducing the energy of the k th mod by $E_k \equiv \sqrt{\mathbf{k}^2 + m^2}$ we choose the plane wave solution

$$\phi_k(t) = \frac{1}{\sqrt{2E_k}} e^{-iE_k t} \quad (6.24)$$

The Hamiltonian follows from inserting Eqn. (6.20) in Eqn. (6.19) and integrating over space to give

$$\hat{H} = \sum_k \left(\hat{n}_k + \frac{1}{2} \right) E_k, \quad \text{where} \quad \hat{n}_k \equiv L^{-3} \hat{a}_{\mathbf{k}}^\dagger \hat{a}_{\mathbf{k}}, \quad (6.25)$$

which is like the quantum mechanical result Eqn. (6.16) except $\hat{a}_i \rightarrow L^{-3/2} \hat{a}_{\mathbf{k}}$. The canonical commutation relations follow

$$L^{-3} [\hat{a}_{\mathbf{k}}, \hat{a}_{\mathbf{k}'}^\dagger] = \delta_{\mathbf{k}\mathbf{k}'}. \quad (6.26)$$

Note that if we increase n_{bfk} by one, then the Hamiltonian increases and the corresponding energy of the system increases by an amount E_k . The momentum of the system can be obtained from Eqns (5.12) and (5.10). In particular to momentum density is

$$T_i^0 = \dot{\phi} \partial_i \phi \quad (6.27)$$

Of course for the case of a homogeneous field as we were discussing in section (5.3) the field has zero momentum density, but we are considering the case where it is not homogeneous (as we ultimately want to consider fluctuations about the homogeneous background field). Promoting ϕ to an operator, inserting the Fourier expansion of ϕ and integrating over all space we arrive at the result

$$\mathbf{p} = \sum_{\mathbf{k}} \hat{n}_{\mathbf{k}} \mathbf{k} \quad (6.28)$$

where the momentum operator in say the z direction is given by $\hat{p}_z$ and satisfies $[\hat{p}_z, \hat{a}_k^\dagger] = k_z \hat{a}_k^\dagger$. It follows that as in the case of the energy, if we increase n_{bfk} by one, we increase the momentum by an amount $\mathbf{k}$.

The operators $\hat{n}_{\mathbf{k}}$ therefore commute for different values of $\mathbf{k}$, hence it is possible to find orthonormal states $|\hat{n}_{k_1}, \hat{n}_{k_2}, \dots\rangle$ that are eigenvectors of every $\hat{n}_{\mathbf{k}}$ with associated eigenvalue $n_{\mathbf{k}}$. It is possible then to build up a full Fock space by starting with the vacuum state $|0, 0, \dots\rangle$, then by acting on them with the operators $\hat{a}_{\mathbf{k}}$ to build up states with non-zero $n_{\mathbf{k}}$ just as in the case of the harmonic oscillator. The states $|\hat{n}_{k_1}, \hat{n}_{k_2}, \dots\rangle$ are the basis for the Hilbert space in quantum theory – this is the Fock space.

We have started with the classical Hamiltonian, promoted the fields to operators and found that the consistency of the quantum theory requires the commutation relation (6.26) be satisfied. Further we saw that $\hat{a}_{\mathbf{k}}^\dagger$ creates particles and $\hat{a}_{\mathbf{k}}$ annihilates them. This is the usual canonical quantisation procedure.

We have so far worked in a box, but we will want to eventually go to the continuum case in momentum which means considering an infinite box. In that situation we make use of the usual transformations between the Fourier sum and Fourier integral

$$\left(\frac{2\pi}{L}\right)^3 \sum_n \rightarrow \int d^3k \quad (6.29)$$

$$g_n \rightarrow g(\mathbf{k}) \quad (6.30)$$

$$\left(\frac{2\pi}{L}\right)^3 \delta \rightarrow \delta^3(\mathbf{k} - \mathbf{k}'). \quad (6.31)$$

The last relation leads to the following important (yet somewhat confusing) relation which is used when considering volume averages

$$[\delta^3(\mathbf{k} - \mathbf{k}')]^2 = \left(\frac{2\pi}{L}\right)^3 \delta^3(\mathbf{k} - \mathbf{k}'). \quad (6.32)$$

It should now be clear that in the continuum case we have instead of Eqn. (6.20) and Eqn. (6.25)

$$\hat{\phi}(\mathbf{x}, t) = \left(\frac{1}{2\pi}\right)^3 \int [\phi_k(t) \hat{a}_{\mathbf{k}} + \phi_k^*(t) \hat{a}_{-\mathbf{k}}^\dagger] e^{i\mathbf{k} \cdot \mathbf{x}} d^3\mathbf{k} \quad (6.33)$$

$$\hat{H} = \left(\frac{1}{2\pi}\right)^3 \int E_k \left(\hat{a}_{\mathbf{k}}^\dagger \hat{a}_{\mathbf{k}} + \frac{1}{2} L^3 \right) d^3\mathbf{k} \quad (6.34)$$

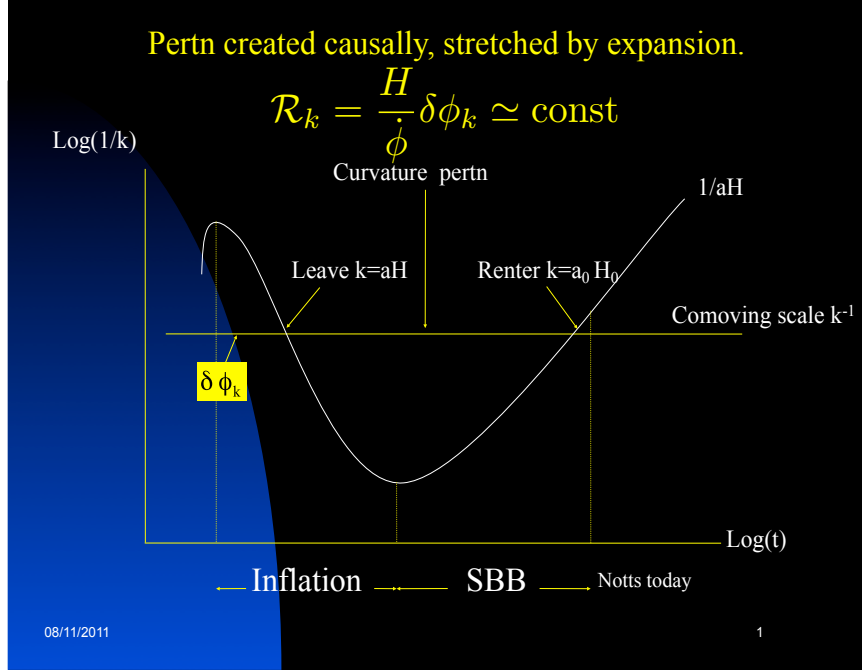


Figure 6.1: A mode of comoving wavenumber k leaves the horizon during inflation when $k = aH$ and freezes in as a classical curvature perturbation ζ (called $\mathcal{R}$ in the figure), before re-entering the horizon today on cosmological scales.

with the commutation relations

$$\left(\frac{1}{2\pi}\right)^3 [\hat{a}_{\mathbf{k}}, \hat{a}_{\mathbf{k}'}^\dagger] = \delta^3(\mathbf{k} - \mathbf{k}'), \quad [\hat{a}_{\mathbf{k}}, \hat{a}_{\mathbf{k}'}] = 0 \quad (6.35)$$

The reason for the $1/\sqrt{E_k}$ factor in ϕ_k now becomes apparent. It ensures the Lorentz invariance of the combination $d^3k/\sqrt{E_k}$ thereby guaranteeing that $\phi(\mathbf{x}, \mathbf{t})$ is a scalar.

6.3 Generating field perturbations as the modes exit the horizon during inflation

After all this build up recalling the case of the Harmonic oscillators and free scalar field quantisation we are in a position to return to the inflationary universe and consider the build up of evolution of the perturbations in the inflation field during inflation. We will focus on some particular comoving wavenumber k which will begin life well within the horizon and at some epoch become larger than the horizon and leave it. Now well before horizon exit, such a mode doesn't feel the curvature of the spacetime and considers its life to be effectively in flat spacetime, where notions such as particle number etc make sense. Given that we are expecting the field not to be excited, it is essentially in its ground state in this regime, i.e. in its vacuum state. Here is the important physics though. Vacuum fluctuations in a light scalar field will 'freeze in' at horizon exit to become classical perturbations. This was first shown by Bunch and Davies and occurs because the time scale a/k of the vacuum fluctuation becomes much bigger than the Hubble time H^{-1} , hence it can not fluctuate on a reasonable timescale. The basic idea is in Figure (6.1)

6.3.1 Massless scalar field during inflation – generation of quantum fluctuations

We are considering the fluctuations of a field about some background homogeneous field, so we first have to consider the first order perturbations of a light scalar field during inflation. We consider the case of almost exponential inflation, as it is the most straightforward, but also representative of the generic case, which will have to be almost exponential as we do not expect the Hubble parameter to vary much during standard slow roll inflation. We ignore metric perturbations, the field perturbation is considered to be evolving in an unperturbed spacetime. Now the classical field equations for a set of fields $\phi_n(\mathbf{x}, t)$ in an unperturbed spacetime is given by Eqn. (6.21) which for the metric Eqn. (1.7) becomes (setting $c = 1$)

$$\ddot{\phi}_n + 3H\dot{\phi}_n - a^{-2}\nabla^2\phi_n + V_n = 0 \quad (6.36)$$

where $V_n \equiv \frac{\partial V}{\partial \phi_n}$. Perturbing about a background homogeneous solution $\bar{\phi}_n(t)$, i.e. by considering a new solution $\phi(\mathbf{x}, t) = \bar{\phi}_n(t) + \delta\phi(\mathbf{x}, t)$ we obtain to linear order in the perturbation

$$\ddot{\delta\phi}_n + 3H\dot{\delta\phi}_n - a^{-2}\nabla^2\delta\phi_n + V_{nm}\delta\phi_m = 0 \quad (6.37)$$

where $V_{nm} \equiv \frac{\partial^2 V}{\partial \phi_n \partial \phi_m}$ and a summation of m is assumed. Moving to Fourier space where we write $\delta\phi_n(\mathbf{x}, t) = \frac{1}{(2\pi)^3} \int \delta\phi_{kn}(t) e^{i\mathbf{k}\cdot\mathbf{x}} d^3\mathbf{k}$ we have for each mode $\delta\phi_{kn}$

$$\ddot{\delta\phi}_{kn} + 3H\dot{\delta\phi}_{kn} + \left(\frac{k}{a}\right)^2 \delta\phi_{kn} + V_{nm}\delta\phi_{km} = 0. \quad (6.38)$$

Now we are concerned with a few Hubble times either side of Horizon exit. We will see that any heavy fields acquire very few perturbations in this regime hence will only keep the light fields. This in turn allows us to drop V_{mn} from Eqn. (6.38), because we are now in a regime where the mass satisfies $m^2 \ll H^2$ and $V_{mn} \propto m^2$. In comparison to the gradient term we also have $V_{mn} \ll (k/a)^2$ because around Horizon exit we have $k/a \sim H$. It then follows that for light fields we can drop the potential term leaving us with the equation strictly only true for a massless free field,

$$\ddot{\delta\phi}_k + 3H\dot{\delta\phi}_k + \left(\frac{k}{a}\right)^2 \delta\phi_k = 0. \quad (6.39)$$

Note we have dropped the subscript n because this equation is true for each field separately as there is no longer any direct coupling between the fields ϕ_n and ϕ_m as the derivative term V_{mn} has been dropped. As we have mentioned earlier, we are interested in a few Hubble times around Horizon exit, so we set $H = H_k = k/a$ its value at horizon exit. Usually there is only a slight scale dependence of H which can be ignored and we can set H_k equal to a constant that we denote by H_* .

The equation can be converted into that of a harmonic oscillator with a time dependent frequency by going to conformal time and redefining the field perturbation. In particular we work with η where $dt \equiv a(\eta)d\eta$, and $\varphi \equiv a\delta\phi$. Constant H in conformal time leads to a simple relation which proves very useful. From $H = \frac{1}{a} \frac{da}{dt} = \frac{1}{a^2} \frac{da}{d\eta}$. Hence it follows that

$$\eta = -\frac{1}{aH}, \quad (6.40)$$

where the constant of integration was chosen such that $\eta \rightarrow 0$ as a hence $t \rightarrow \infty$. It follows that $a = 0$ or $t = 0$ (the initial singularity) corresponds to $\eta \rightarrow -\infty$. Using $H = H_k$ we find after a bit of algebra that Eqn. (6.39) becomes

$$\frac{d^2 \varphi_k(\eta)}{d\eta^2} + \omega_k^2(\eta) \varphi_k(\eta) = 0, \quad (6.41)$$

where the time dependent frequency is

$$\omega_k^2(\eta) = k^2 - \frac{2}{\eta^2} \equiv k^2 - 2(aH_k)^2, \quad (6.42)$$

the last equality arising from Eqn. (6.40).

6.3.2 Quantisation of the massless fields

Looking at Eqn. (6.42) we see that well before Horizon exit (i.e. $\eta \rightarrow -\infty$), the harmonic oscillator has a constant angular frequency k . We can then work in a small spacetime region with size $\Delta\eta \sim \Delta x$ which satisfies $k^{-1} \ll \Delta\eta \ll (aH_k)^{-1}$, a region which is large enough to contain many oscillations but small enough not to feel the spacetime curvature. Using the Equivalence Principle we can then work in flat spacetime field theory in this region. In quantising we use the analogue of Eqn. (6.33) for $\hat{\phi}$ to define the mode function φ_k in terms of the Fourier components $\varphi_{\mathbf{k}}$

$$(2\pi)^3 \hat{\phi}_{\mathbf{k}}(\eta) = \varphi_k(\eta) \hat{a}(\mathbf{k}) + \varphi_k^*(\eta) \hat{a}^\dagger(-\mathbf{k}) \quad (6.43)$$

This satisfies Eqn. (6.41) and given that we require initial conditions ($\eta \rightarrow -\infty$)

$$\varphi_k(\eta) = \frac{1}{\sqrt{2k}} e^{-ik\eta} \quad (6.44)$$

it has a general solution (which can be shown by direct substitution back into the evolution equation)

$$\varphi_k(\eta) = \frac{1}{\sqrt{2k}} e^{-ik\eta} \frac{(k\eta - i)}{k\eta}. \quad (6.45)$$

Consistency with the initial condition requirement clearly follows for $\eta \rightarrow -\infty$. Well after horizon, $\eta \rightarrow 0$ the solution approaches

$$\varphi_k(\eta) = -\frac{i}{\sqrt{2k}} \frac{1}{k\eta}. \quad (6.46)$$

Now as stated above, we assume the state corresponds initially to the vacuum with no ϕ particles present, something that is expected if there has been some inflation occurring before the horizon exit (remember inflation dilutes the particle number dramatically). In other words it is in the ground state of an harmonic oscillator. In that case a measurement of the Fourier components ϕ_k at some particular instant has an outcome of the measurement which is a gaussian distribution for the real and imaginary part of each component, there is no correlation other than the reality condition. It means that we are dealing with a gaussian random field whose ensemble average may be identified with the vacuum expectation value. As a result the mean $\langle \hat{\phi}_k \rangle$ vanishes. The spectrum is defined by

$$\langle \hat{\phi}_k \hat{\phi}_{k'} \rangle = \frac{2\pi^2}{k^3} \mathcal{P}_\varphi \delta^3(\mathbf{k} + \mathbf{k}') \quad (6.47)$$

Now inserting Eqn. (6.43) and using Eqn. (6.35) for the CCR, then recalling that the expectation value refers to the vacuum state it is straightforward to show

$$\mathcal{P}_\varphi(k, \eta) = \frac{k^3}{2\pi^2} |\varphi_k(\eta)|^2 \quad (6.48)$$

We want $\mathcal{P}_{\delta\phi}$, so we recall $\varphi = a\delta\phi$, hence we just need to divide by a^2 to obtain what we require:

$$\mathcal{P}_{\delta\phi}(k, \eta) = \frac{k^3}{2\pi^2} \left| \frac{\varphi_k(\eta)}{a} \right|^2 \quad (6.49)$$

Evaluating the solution a few Hubble times after horizon exit we have from Eqn. (6.46),

$$\mathcal{P}_{\delta\phi}(k, \eta) = \frac{1}{4\pi^2} \frac{1}{(a\eta)^2} = \left(\frac{H_k}{2\pi} \right)^2 \quad (6.50)$$

a result first obtained by Bunch and Davis in 1978, and well before inflation had been proposed as a solution to anything in cosmology. It just relied on having de Sitter expansion. It is hard to overestimate the impact this result has had on cosmology. We will use it to determine a number of cosmological observables arising from inflation, observables such as the spectral index associated with the inflaton field.

6.3.3 Going from Quantum to Classical physics

Well before Horizon exit, φ_k is a vacuum fluctuation of a free field in flat spacetime, a true quantum object. However, after horizon exit we can see from Eqn. (6.46) that φ_k is purely imaginary meaning that the operator $\hat{\varphi}_{\mathbf{k}}(\eta)$ can be written as

$$(2\pi)^3 \hat{\varphi}_{\mathbf{k}}(\eta) = \varphi_k(\eta) [\hat{a}(\mathbf{k}) - \hat{a}^\dagger(-\mathbf{k})]. \quad (6.51)$$

This has two consequences, the time dependence of $\hat{\varphi}_{\mathbf{k}}(\eta)$ is now trivial and the state continues to be an eigenvector. It implies that once $\varphi_k(t)$ is measured at some instance well after horizon exit, it will continue to have a definite value, it can in essence be considered as a classical object.

6.3.4 Including linear corrections from the potential – going beyond slow roll

We have so far ignored the influence of the potential, dropping the V_{mn} term in Eqn. (6.38). What effect does it have if we keep it in? Concentrating on one light field $\phi \equiv \phi_n$, and using the effective mass of the perturbation $\delta\phi$, namely $\frac{\partial V(\phi)}{\partial \phi^2} \equiv m^2(\phi)$, then Eqn. (6.38) becomes

$$\ddot{\delta\phi}_k + 3H\dot{\delta\phi}_k + \left(\frac{k}{a} \right)^2 \delta\phi_k + m^2(\phi)\delta\phi_k = 0. \quad (6.52)$$

In general we don't expect $m^2(\phi)$ to vary very much, it will of course be the actual mass squared of the free field if $V = \frac{1}{2}m^2\phi^2$ and more generally we expect it to have an almost constant value of say m_k^2 during the few Hubble times at horizon exit. We work in that regime. Working again with $\varphi = a\delta\phi$ and in conformal time η , the only change to Eqn. (6.41) is the addition of the potential term which leads to

$$\frac{d^2\varphi_k(\eta)}{d\eta^2} + \Omega_k^2(\eta)\varphi_k(\eta) = 0, \quad (6.53)$$

where the new time dependent frequency is

$$\Omega_k^2(\eta) = (am_k)^2 + k^2 - \frac{2}{\eta^2}, \quad (6.54)$$

with $aH_k = -1/\eta$. This can be solved exactly. The full solution which gives the initial condition Eqn. (6.44), is

$$\varphi(k, \eta) = e^{i(\nu + \frac{1}{2})\frac{\pi}{2}} \sqrt{\frac{\pi}{4k}} \sqrt{k\eta} H_\nu^{(1)}(k\eta) \quad (6.55)$$

where $H_\nu^{(1)}$ is the Hankel function of the first kind and

$$\nu = \sqrt{\frac{9}{4} - \frac{m_k^2}{H_k^2}} \simeq \frac{3}{2} - \frac{m_k^2}{3H_k^2}. \quad (6.56)$$

where we are in the regime $m_k \ll H_k$. Provided that ν is real there is a quantum to classical transition as in the massless case. This corresponds to the weaker condition $m_k^2 < (9/4)H_k^2$. To obtain the Power spectrum we once again consider the solution well after horizon exit ($\eta \rightarrow 0$) we have [**Check soln Ed!**]

$$\varphi(k, \eta) = e^{i(\nu - \frac{1}{2})\frac{\pi}{2}} \frac{2^\nu \Gamma(\nu)}{2^{\frac{3}{2}} \Gamma(\frac{3}{2})} \frac{1}{\sqrt{2k}} (k\eta)^{\frac{1}{2} - \nu} \quad (6.57)$$

Using

$$\mathcal{P}_\varphi(k, \eta) = \frac{k^3}{2\pi^2} |\varphi_k(\eta)|^2 \quad (6.58)$$

and recalling that $\mathcal{P}_{\delta\phi} = \frac{1}{a^2} \mathcal{P}_\varphi$ we obtain (recall $1 - 2\nu = -2 + \frac{2m_k^2}{3H_k^2}$)

$$\mathcal{P}_{\delta\phi}(k, \eta) = \frac{k^3}{2\pi^2} \frac{1}{2k} \frac{(k\eta)^{1-2\nu}}{a^2} \quad (6.59)$$

$$= \frac{1}{4\pi^2} \frac{1}{a^2 \eta^2} (k\eta)^{\frac{2m_k^2}{3H_k^2}} \quad (6.60)$$

$$= \left(\frac{H_k}{2\pi}\right)^2 \left(\frac{k}{aH_k}\right)^{\frac{2m_k^2}{3H_k^2}}. \quad (6.61)$$

This is valid as long as $m^2(\phi)$ and H have very little variation. Note that the correction to the massless de Sitter result is expected to be small because we are evaluating $\mathcal{P}_{\delta\phi}$ just after horizon crossing when $k \simeq aH_k$, hence the new factor is of order unity. There is another effect on the power spectrum as well as the effect of the potential, but which we will not be discussing and that is the effect of the metric perturbations. Recall in deriving the fluctuation equation (6.37) we have ignored any fluctuation in the metric. Given the background equation is roughly of the form $\square\phi + V' = 0$, then we expect a variation in $(\delta\square)\phi$, hence we expect Eqn. (6.37) to be more generally

$$\delta\ddot{\phi}_n + 3H\delta\dot{\phi}_n - a^{-2}\nabla^2\delta\phi_n + V_{nm}\delta\phi_m = (\delta\square)\phi_n(t) \quad (6.62)$$

where the RHS is the effect of the metric perturbation at first order. Without proof we quote the result that the generalisation of Eqn. (6.41) is

$$\frac{d^2\varphi_k(\eta)}{d\eta^2} + \left(k^2 - \frac{1}{z} \frac{d^2 z}{d\eta^2}\right) \varphi_k(\eta) = 0, \quad (6.63)$$

where z is given in terms of the unperturbed field by $z \equiv \frac{a\dot{\phi}}{H}$. Equation (6.63) is known as the Mukhanov-Sasaki equation. In terms of the alternate slow roll parameters

$$\epsilon_H = -\frac{\dot{H}}{H}, \quad \eta_H = \epsilon_H - \frac{1}{2} \frac{\dot{\epsilon}_H}{H\epsilon_H} \quad (6.64)$$

it is possible to show

$$\frac{1}{z} \frac{d^2 z}{d\eta^2} = 2a^2 H^2 \left(1 + \epsilon_H - \frac{3}{2} \eta_H + \frac{1}{2} \eta_H^2 - \frac{1}{2} \epsilon_H \eta_H + \frac{1}{2H} \frac{d\epsilon_H}{dt} - \frac{1}{2H} \frac{d\eta_H}{dt} \right) \quad (6.65)$$

6.4 Calculating the curvature perturbation ζ at horizon exit

In the previous subsection we saw how the vacuum fluctuation of each light scalar field when treated in the slow roll regime is converted to a classical perturbation at the time of horizon exit. The idea is that at least one of these perturbations should now generate the curvature perturbation ζ (to be defined below) and which in turn is probed when cosmological scales begin to enter the horizon. Now the fluctuation $\delta\phi(\mathbf{x}, t)$ is generally defined on what is known as flat slicing. This is a bit technical and we have not gone into the details of what it means, but basically when we consider the fluctuations of the metric we allow locally for the scale factor to become position dependent, i.e. $a(\mathbf{x}, t)$. The flat slicing is when we restrict ourselves to the homogeneous case where even locally we have $a(\mathbf{x}, t) = a(t)$, i.e. it is time dependent only.

The Primordial curvature perturbation ζ is an incredibly important quantity in cosmology. It arises when we consider smoothing the energy-momentum tensor and the metric on the shortest cosmological scale. The curvature perturbation has a constant value $\zeta(\mathbf{x})$ while that scale approaches horizon entry, and this constant value determines the total energy density perturbation providing the main initial condition for the evolution of all perturbations within the horizon. It is therefore accessible to observation. The only demand placed on $\zeta(\mathbf{x})$ is that the energy-momentum tensor and the metric tensor are smoothed on some comoving scale, which is well outside. It is defined in the following way. Considering a perturbed metric we can write the spatial metric in the form

$$g_{ij} = a^2(\mathbf{x}, t) \gamma_{ij}(\mathbf{x}) \quad (6.66)$$

with

$$a(\mathbf{x}, t) \equiv a(t) e^{\zeta(\mathbf{x})}, \quad \gamma_{ij}(\mathbf{x}) = \delta_{ij} \quad (6.67)$$

the relation for γ_{ij} implying that the metric perturbation h_{ij} is time independent (although this is not clear from what we have written here.) If the tensor perturbation h_{ij} is negligible then ζ determines the curvature perturbation, and this is the context in which it is normally considered. It turns out (see Lyth and Liddle chapter 5 for full details) that on a flat slicing surface

$$\zeta = -H \frac{\delta\rho}{\dot{\rho}} \quad (6.68)$$

where the background energy density is $\rho(t)$ and the perturbation is $\delta\rho(\mathbf{x}, t)$. It can be shown from this that to first order if and only if the pressure is a unique function of the energy density $p(\rho(\mathbf{x}))$ then ζ is conserved. Now in the case of the scalar field the perturbation $\delta\phi(\mathbf{x}, t)$ is defined on the flat slicing, whereas ζ is defined on the uniform energy density slicing. However $\phi(\mathbf{x})$ is independent of position up to a time shift, hence its value at any instant gives the energy density

or the Hubble parameter (through $H^2 = \frac{8\pi G}{3} \left(\frac{1}{2}\dot{\phi}^2 + V(\phi) \right)$). This implies that ϕ is also uniform on a slice of uniform density. It follows that we can replace ρ with ϕ in ζ giving to first order

$$\zeta = -H \frac{\delta\phi}{\dot{\phi}} \quad (6.69)$$

where $\delta\phi$ is defined on the flat slicing. We are nearly there, in fact we can begin to see the link in Eqn. (6.69) with the late time curvature perturbation ζ and the early universe ϕ perturbation. We need to evaluate Eqn. (6.69) a few Hubble times after horizon exit, because ζ is time independent from that moment onwards. In particular the spectrum is given by (recall the spectrum tells us about $\langle \zeta_k \zeta'_k \rangle$ etc...

$$\mathcal{P}_\zeta(k, \eta) = \left(\frac{H}{\dot{\phi}} \right)^2 \mathcal{P}_{\delta\phi}(k, \eta) \Big|_{k=aH} \quad (6.70)$$

where we are calculating at the epoch of horizon exit. Now we have seen from Eqn. (6.50) that a few Hubble times after horizon exit $\delta\phi(\mathbf{k})$ has the spectrum $\mathcal{P}_{\delta\phi}(k, \eta) = \left(\frac{H_k}{2\pi} \right)^2$. Therefore it follows that

$$\mathcal{P}_\zeta(k, \eta) = \frac{1}{4\pi^2} \left(\frac{H^2}{\dot{\phi}} \right)^2 \Big|_{k=aH}. \quad (6.71)$$

Although strictly true a few Hubble times after horizon exit, yet the subscript $k = aH$ suggests we are evaluating at horizon crossing, this should be fine because of the fact that H and $\dot{\phi}$ are slowly varying on the Hubble timescale.

We can go further yet and determine the curvature perturbation $\mathcal{P}_\zeta$ in terms of the slow roll parameters. It requires a bit of patience but basically we make use of the slow roll relations Eqns. (5.22), (5.23) and (5.24) to finally obtain

$$\mathcal{P}_\zeta(k, \eta) = \frac{(8\pi G)^2}{24\pi^2} \frac{V(\phi)}{\epsilon(\phi)} \Big|_{k=aH}. \quad (6.72)$$

Now the spectral index associated with the primordial curvature perturbation is defined by

$$n - 1 \equiv \frac{d \ln \mathcal{P}_\zeta(k, \eta)}{d \ln k}, \quad (6.73)$$

or equivalently $\mathcal{P}_\zeta(k, \eta) = A k^{n-1}$, where A is a constant. The case $n = 1$ is possibly the most famous and corresponds to the Harrison-Zeldovich scale invariant spectrum. Of course there is no apriori reason why n should be a constant, it could depend on $\mathbf{k}$ through $n(\mathbf{k})$ implying there would be a running of the spectral index $\frac{dn}{d \ln k}$. We will not consider such a possibility here. Observations constraint $\mathcal{P}_\zeta(k)$ when evaluated at a particular ‘pivot point’ $k_0 \equiv 0.002 \text{Mpc}^{-1}$. Infact the WMAP7 results have led to the observed values

$$\mathcal{P}_\zeta^{\frac{1}{2}}(k_0) = (4.9 \pm 0.2) \times 10^{-5}; \quad n = 0.96 \pm 0.03 \quad (6.74)$$

We conclude our chapter on perturbations by showing how we can constrain inflationary models using these results. We need a few formula to help us along the way. Instead of working with t we

work with the number of efolds N defined by $dN = -Hdt$. A number of formula follow:

$$\frac{d(\ln H)}{dN} = \epsilon \quad (6.75)$$

$$-\frac{d(\ln \epsilon)}{dN} = 4\epsilon - 2\eta \quad (6.76)$$

$$d \ln(aH) = Hdt \quad \text{for } H \sim \text{const} \quad (6.77)$$

$$d \ln k = d \ln(aH) \quad \text{at horizon crossing} \quad (6.78)$$

Finally from Eqns (6.72) and (6.73) we obtain

$$n - 1 = \frac{d \ln \mathcal{P}_\zeta(k, \eta)}{d \ln k} = \frac{d \ln V}{d \ln k} \Big|_{k=aH} - \frac{d \ln \epsilon}{d \ln k} \Big|_{k=aH} \quad (6.79)$$

$$= \frac{d \ln V}{d \ln(aH)} - \frac{d \ln \epsilon}{d \ln(aH)} \quad (6.80)$$

$$= \frac{1}{H} \frac{d \ln V}{dt} - \frac{1}{H} \frac{d \ln \epsilon}{dt} \quad (6.81)$$

But $dN = -Hdt$ hence

$$n - 1 = -\frac{d \ln V}{dN} + \frac{d \ln \epsilon}{dN} = -2\epsilon - 4\epsilon + 2\eta \quad (6.82)$$

where in the first term on the RHS we have replaced V with H through $H^2 = \frac{8\pi G}{3}V$. At last we have the remarkable result that

$$n(k) - 1 = -6\epsilon + 2\eta \quad (6.83)$$

where as we have said the RHS is evaluated at $k = aH$. Using Eqn. (6.74) we see that the constraints on the inflaton potential and its derivatives (assuming single field slow roll inflation of course) become

$$\left(\frac{V}{\epsilon}\right)^{\frac{1}{4}} = (4.9 \pm 0.2) \times 10^{-5}; \quad -6\epsilon + 2\eta = -0.04 \pm 0.03 \quad (6.84)$$

thereby showing how cosmology can directly constrain the potential associated with early universe inflation.

7 Dark energy

In this last chapter we will give a brief introduction to Dark Energy.

7.1 Why Dark energy and what is it?

Dark energy is a fluid which has negative pressure and an equation of state $w < -1/3$ and participates very little if at all on structure formation. This is the only thing we know! In particular, dark energy must contribute to the Friedman equation:

$$3H^2 = 8\pi G(\rho_m + \rho_{DE}) \quad (7.1)$$

so that $\Omega_{DE} \sim 0.7$ today. And it must provide negative pressure so that the acceleration equation which is

$$\frac{\ddot{a}}{a} = -\frac{4\pi G}{3}(\rho_m + \rho_{DE} + 3P_{DE}) \quad (7.2)$$

becomes positive, i.e. we must have $\rho_{DE} + 3P_{DE} < 0$. The equation of state of dark energy $w_{DE} = P_{DE}/\rho_{DE}$ must be smaller than $-1/3$ (actually data show that it is very close to -1). The equation of state can also be time-varying $w_{DE} = w_{DE}(t)$. Indeed in most models of dark energy that is the case (e.g. quintessence). If it is time varying then we may also define an *adiabatic* speed of sound c_a^2 as

$$c_a^2 = \frac{dP}{d\rho} = w - \frac{\dot{w}}{3(1+w)H} \quad (7.3)$$

so that if $\dot{w} = 0$ then $c_a^2 = w$. The background DE energy density evolves as

$$\dot{\rho}_{DE} + 3H(1 + w_{DE})\rho_{DE} = 0 \quad (7.4)$$

Dark energy, with the exception of cosmological constant, should also have perturbations, i.e. it should have a density contrast δ_{DE} , perturbed pressure Π_{DE} , velocity u_{DE} and even anisotropic stress σ_{DE} . In the simplest cases, e.g. quintessence, the anisotropic stress is zero. But the usual relation that holds between Π and δ as in CDM or radiation, does not hold here. The most general relation between Π and δ involves a new free function of space and time called the effective speed of sound c_s^2 . We have

$$\Pi_{DE} = c_s^2 \delta_{DE} + 3(1 + w_{DE})\mathcal{H}(c_s^2 - c_a^2)u_{DE} \quad (7.5)$$

So to characterize dark energy we specify $w_{DE}(t)$ and $c_s^2(t, k)$ (the adiabatic speed of sound is not independent). For instance, Λ has $w = c_a^2 = -1$ and c_s^2 is not really defined but may also be taken to be -1 . Quintessence has $w = w(t)$ and $c_s^2 = 1$. K-essence has $w = w(t)$ and $c_s^2 \neq 1$.

7.1.1 Hints of dark energy

Why do we need Dark Energy? Good question. The common expectation would be that the Universe is filled with ordinary matter and a little bit of radiation. That would mean that

$$\Omega_m = 1 \quad (7.6)$$

That was indeed that expectation up until 1984. Over a period of about 10 years, a number of data sets started pointing to a Universe where $\Omega_m < 1$. There were data which came mostly from virial estimates of cluster masses. A collection of these constraints is shown in figure 7.1 from a paper by Krauss and Turner (1994).

Another, good argument came from Efstathiou, Sutherland and Maddox. They used large scale structure data from the APM survey to show that $\Omega_m \sim 0.3$. They concluded that given that there must have been a period of inflation, then the Universe must be flat. Therefore the rest must be in the simplest possible form: a cosmological constant! This is a giant leap of faith in 1991 given that the acceleration of the Universe was not discovered yet. Up to today, they still may be right.

However, all of these datasets simply show that $\Omega_m \sim 0.3$. But they don't show that there is anything like Dark Energy at that point. It could be curvature of $\Omega_K = 0.7$, it could be scalar fields (but not quintessence), massive neutrinos (the so called $\tau - CDM$ model), warm Dark Matter, and other forgotten possibilities. In the 90's it was popular to try to find models which give an open Universe with $\Omega_K \sim 0.7$. Some models called "open inflation" tried to do that. Yes this is not

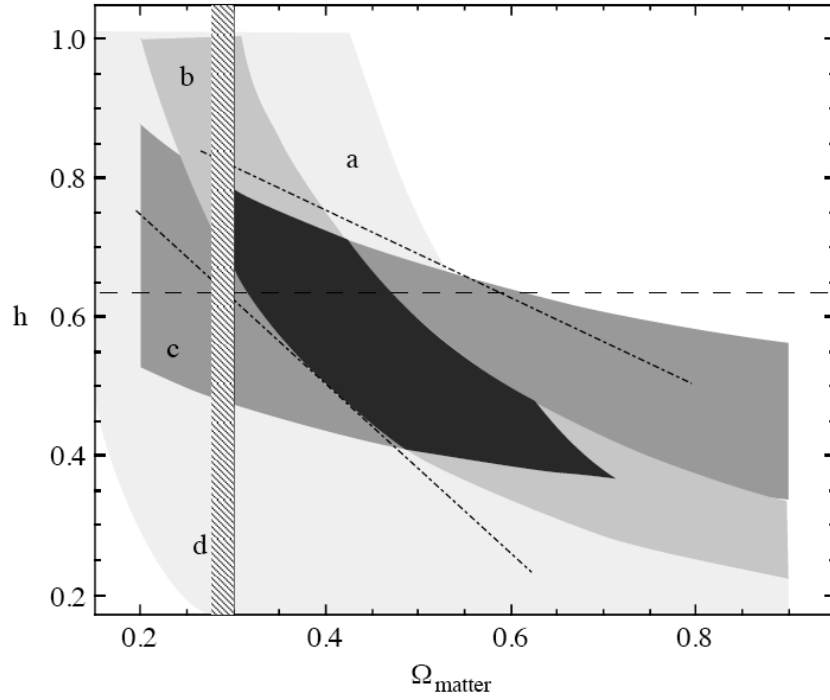
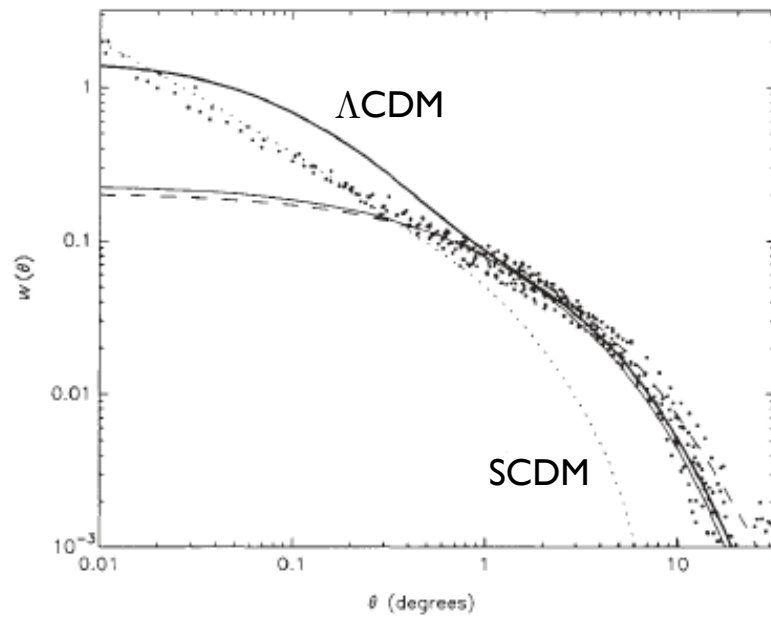


Figure 7.1: Constraints on the Hubble parameter h and Ω_m from data prior to 1994. They are:

- (a) BBN (Big-Bang Nucleosynthesis) limits
- (b) Clustering
- (c) Globular cluster ages
- (d) Virial estimates of cluster masses

All show that $\Omega_m < 1$ is preferred, particularly the virial estimates of cluster masses data. Taken from Krauss and Turner 1994.



Efstathiou, Sutherland, Maddox
Nature 348, 705 (1990)

Figure 7.2: APM survey shows evidence for Dark Energy in 1991 from the galaxy angular correlation function.

a typo. There were inflationary models which tried to create a non-flat Universe. Sounds weird given that inflation was invented to solve the flatness problem! We now know that the Universe is nearly flat. This comes from a variety of data but the killer data set was the Cosmic Microwave Background.

7.1.2 Acceleration observed: the Supernovae data

The killer evidence for the existence of Dark Energy came in 1998 when two independent teams discovered that the expansion of the Universe is accelerating. They eventually shared the Nobel prize in 2011 for their discovery.

To do that they observed the light curves of type-1a supernovae. What are type-1a supernovae? Supernovae come with the following classification

- Type I: Their spectrum does not contain a line of hydrogen. They are further classified as
 - Type Ia: Their spectrum contains a singly ionized silicon line.
 - Type Ib: Their spectrum contains a helium line.
 - Type Ic: Their spectrum lacks both lines.
- Type II: Their spectrum contains a line of hydrogen.

Type 1a supernovae are thought to occur when the mass of a white dwarf in a binary system exceeds the Chandrasekhar mass limit which is $1.38M_{sol}$. The white dwarf continuously accretes material from its companion until its mass exceeds $1.38M_{sol}$. When that happens the star starts to collapse under gravity and the increased temperature leads to a massive supernovae explosion.

What is nice about type-1a supernovae is that their absolute magnitude M is correlated with the width of their light curve. This means that we can use them as standard candles. Standard candles are a set of objects that have known properties so that they can be used to estimate distances. Brighter supernovae have a broader light curve. What one does is to measure the apparent magnitude m and the light curve from which the luminosity distance is determined:

$$m - M = 5 \log \frac{d_L}{1Mpc} + 25 \quad (7.7)$$

The observed supernovae appear brighter in the past so that the luminosity distance is greater than the expectation based on an $\Omega_m = 1$ universe. What is left is to convert d_L to a redshift and then we can extract information on the underlying cosmological model.

7.1.3 Distance measures

The usual evidence for Dark Energy comes from distance measures so it is important to remind ourselves of them. There are two that are widely used. The angular diameter distance is the actual diameter of an object over the angle subtended.

$$d_A = \frac{1}{1+z} \frac{1}{H_0 \sqrt{\Omega_{0K}}} \sinh \left[H_0 \sqrt{\Omega_{0K}} \int_0^z \frac{dz'}{H(z')} \right] \quad (7.8)$$

The luminosity distance squared is the observed luminosity of the source over the observed flux times 4π

$$d_L^2 = \frac{L_s}{4\pi \mathcal{F}} \quad (7.9)$$

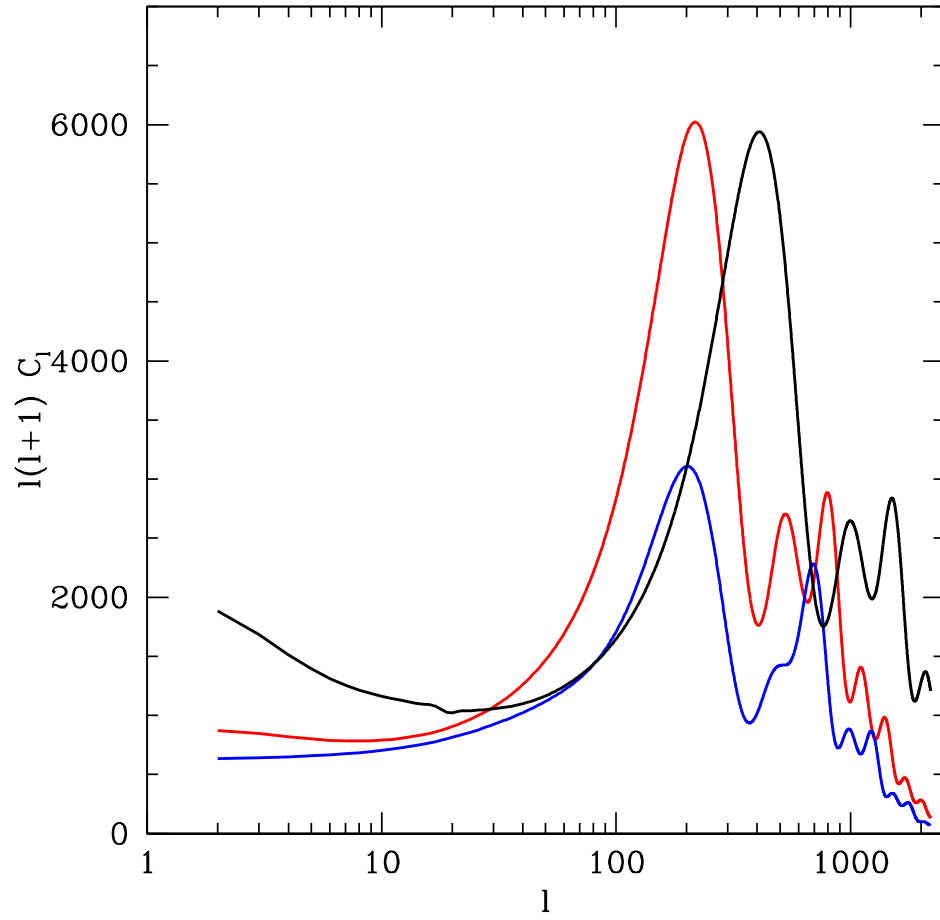


Figure 7.3: CMB angular power spectrum for Λ CDM (red), standar $\Omega = 1$ $SCDM$ (blue) and open $\Omega_K = 0.7$ $OCDM$ (black). Data agree well with the red curve. The other two models are well off the mark.

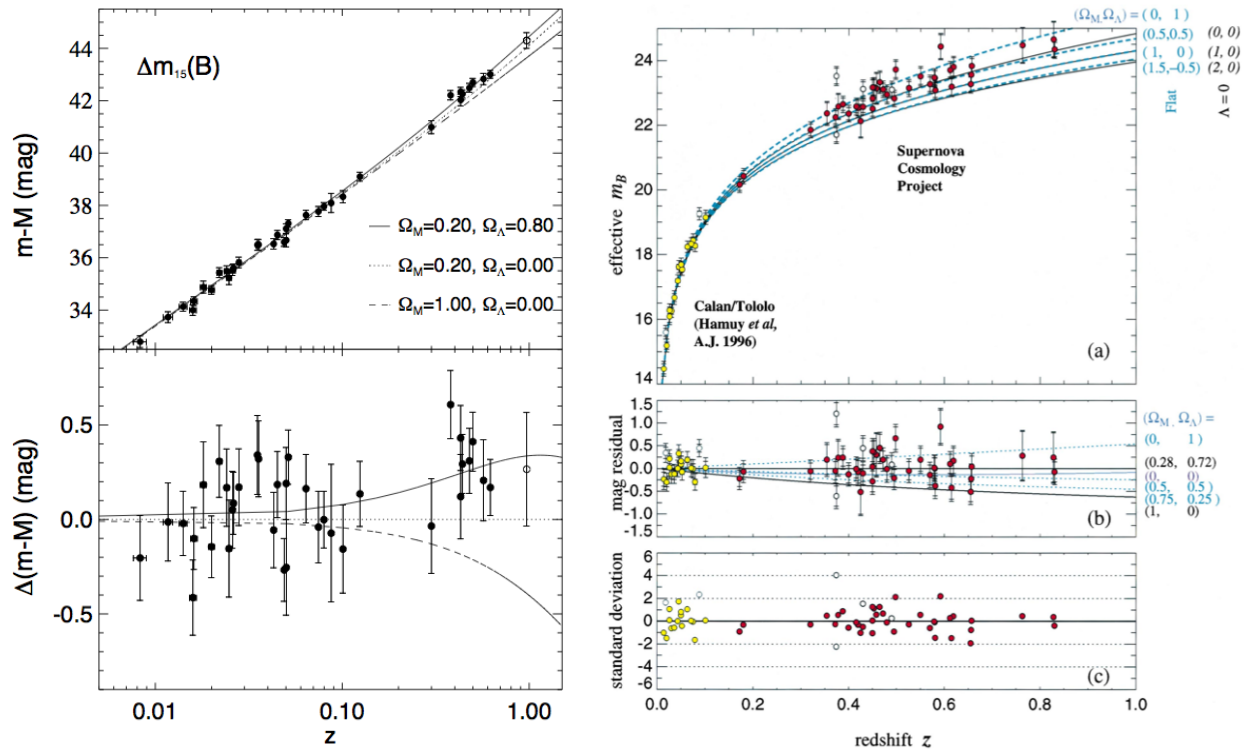


Figure 7.4: The original Type-1a supernovae data from the two independent groups that got the Nobel prize in 2011.

A mathematical theorem due to Etherington in 1931 relates the two for any theory of gravity which permits a description in terms of a spacetime irrespective of the field equations. The only assumption is that photon number is conserved. They are related as

$$d_L = (1 + z)^2 d_A \quad (7.10)$$

7.2 The Λ CDM model, its predictions and possible shortcomings

The supernovae people didn't show evidence for dark energy or acceleration (acceleration is not measured directly). Rather they were more specific: showed evidence for a cosmological constant. This leads to the so called Λ CDM concordance model where the universe today is composed of about $\Omega_\Lambda = 73$, $\Omega_c = 0.23$ and $\Omega_b = 0.04$. The cosmological constant model is quite rigid in its predictions and we have to understand what they are. If any of them fails then we may start to think that this model is not the correct model to use.

7.2.1 Predictions of the Λ CDM model

Some of these are

- We should see an integrated Sachs-Wolfe effect which is correlated in a particular way with large scale structure.
- The growth function f must behave as $f = \Omega_m^\gamma$ where $\gamma = \frac{6}{11} + \frac{15}{2057}\Omega_\Lambda + \dots$
- The CMB angular diameter distance must correlate with the angular diameter distance from Baryon Acoustic Oscillations.
- The matter power spectrum is suppress compared to standard CDM (see fig 7.6).
- Certain relations between observables exist in Λ CDM and one can test for them. For instance, the Chiba-Nakamura relation.
- A cold dark matter should be detected in the near future.

7.2.2 Shortcomings of the Λ CDM model

Perhaps the most serious problem with Λ CDM is the cosmological constant problem: That the observed value of Λ is around 120 orders of magnitude smaller than the naive expectation that it should be of the Planck Mass, M_{Pl}^4 . Super-Symmetric (SUSY) theories can lower this expectation to that of the SUSY breaking scale, but this still required a bare Λ_0 to cancel the vacuum energy coming from the SUSY symmetry breaking scale to about 60 decimal places. One could consider arguing that some unknown physics at high energies may provide a mechanism for achieving this level of fine-tuning, but this seems unlikely as the problem already manifests on low energies.

Suppose that we want to describe all physics up-to scales just above the electron mass. Then the contribution to the vacuum energy Λ will include a bare term Λ_1 , a term coming from the electron and a term coming from the neutrino. This is schematically given by

$$\Lambda = \Lambda_1 + c_e m_e^4 + c_\nu m_\nu^4 \dots,$$

where c_e and c_ν are coefficients. Now if we lower the energy below the electron mass and integrate out the electron, we would instead have

$$\Lambda = \Lambda_0 + c_\nu m_\nu^4 \dots,$$

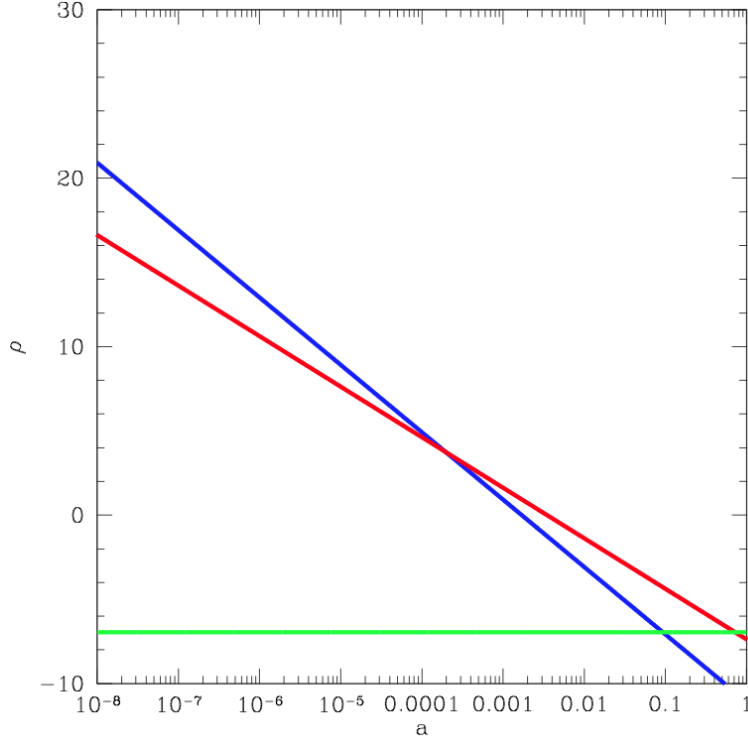


Figure 7.5: The evolution of the energy densities of radiation (blue), matter (red) and Λ (green) showing the coincidence problem.

for a new bare term Λ_0 . To get the same observable vacuum energy Λ , Λ_1 and Λ_0 must cancel to 32 decimal places.

It may be that some mechanism relaxes the effective cosmological constant¹⁰ to zero dynamically but Weinberg [?] shows that this is impossible. Suppose that there is a set of N scalars, ϕ^A , that are responsible for driving the effective Λ to zero. These scalars will contribute an effective potential $V(\phi^A)$ to the cosmological constant. If we are to approach a global Minkowski metric at these energy levels, then $V(\phi^A)$ must cancel the other contributions to Λ to high accuracy as the fields settle to the minimum. However, this is hardly a readjustment mechanism: If the cosmological constant changes slightly, then the mechanism fails. This proof assumes Poincaré invariance in the scalar sector which could be considered an unnecessary assumption (e.q. Horndeski's theory and the Fab Four).

The present value of Λ , as implied by cosmological observations, has another potential problem associated with it: It has an energy density of the same order of magnitude as the average matter density in the Universe today,

$$\rho_\Lambda|_{a=1} \sim \rho_m|_{a=1}.$$

These two quantities scale with the size of the Universe in very different ways, and so their similarity at the present time appears naively to be somewhat of a coincidence. Hence, this problem is sometimes referred to as *the coincidence problem*. It is displayed graphically in figure 7.5.

¹⁰By effective cosmological constant we mean the effective spacetime curvature of the vacuum.

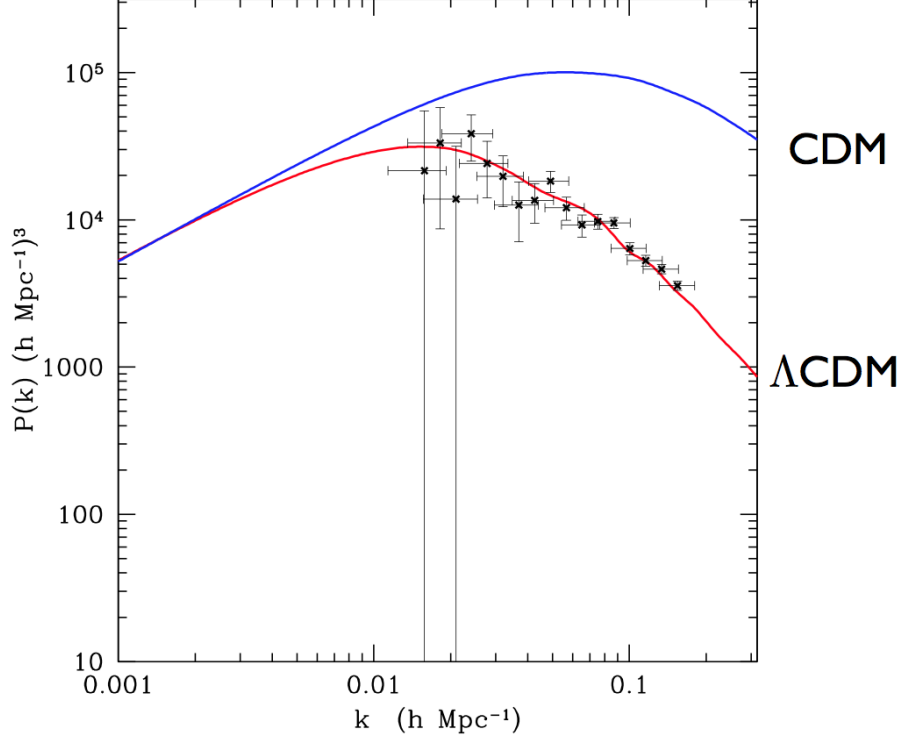


Figure 7.6: The matter power spectrum for Λ CDM and $SCDM$ contrasted. Λ inhibits growth.

7.3 Dynamical dark energy

7.3.1 Phenomenological dark energy

If all we want to do is to study constraints on the equation of state of dark energy then all we need is to specify a $w(z)$ and solve the energy conservation equation to get the dark energy density and then the Hubble rate. The easiest thing to do is to assume that w is constant. Then we may immediately obtain

$$\rho_{DE} = \rho_{0,DE} a^{-3(1+w)} \quad (7.11)$$

If w is not constant then the above relation is not valid anymore. We still get $\rho_{DE}(z)$ though through simple integration. We start from the dark energy conservation equation, and change variables from time t to z . We have that $H = (1+z)\frac{d}{dt}(\frac{1}{1+z}) = -\frac{\dot{z}}{1+z}$ so that

$$d\rho - 3(1+w)\frac{dz}{1+z}\rho = 0 \quad (7.12)$$

which integrate to

$$\rho = \rho_{0,DE} \exp\left[3 \int_0^z \frac{1+w}{1+z'} dz'\right] \quad (7.13)$$

We then have to specify $w(z)$. There are a number of proposals on how to parametrize $w(z)$ apart from the constant value. One idea is to use an expansion

$$w = \sum_n w_n x_n(z) \quad (7.14)$$

where different cases are

- Redshift: $x_n = z^n$.
- Scale factor: $x_n = (1 - a)^n = \left(\frac{z}{1+z}\right)^n$.
- Logarithmic: $x_n = \ln^n(1 + z)$.

Case 2 is of particular interest. Stopping at order $n = 1$, we have

$$w = w_0 + w_1 \frac{z}{1+z} \quad (7.15)$$

It was introduced by Chevallier, Polarski and Linder and is called the CPL parametrization. Although simple it is apparently quite robust and powerful at the same time as it can accommodate a variety of realistic dark energy models.

Alternatively we may try to reconstruct the equation of state from the supernovae data. We start from the Friedman equation and define the dimensionless Hubble rate $E(z)$ as $E = H/H_0$. Then (ignoring radiation)

$$E^2(z) = \Omega_{0K}(1+z)^2 + \Omega_{0m}(1+z)^3 + \Omega_{0,DE} \exp\left[3 \int_0^z \frac{1+w}{1+z'} dz'\right] \quad (7.16)$$

Differentiating with respect to z we find that

$$w = \frac{(1+z)[2EE' + \Omega_{0K}(1+z)] - 3E^2}{3[E^2 - \Omega_{0K}(1+z)^2 - \Omega_{0m}(1+z)^3]} \quad (7.17)$$

Now we can also relate $E(z)$ to the luminosity distance. We find

$$E(z) = \frac{(1+z) \sqrt{\frac{(1+z)^2}{H_0^2} + \Omega_{0K} D_L^2(z)}}{(1+z) D_L'(z) - D_L(z)} \quad (7.18)$$

so given $D_L(z)$ we may reconstruct $w(z)$ provided we know the matter density Ω_{0m} and the curvature Ω_{0K} .

7.3.2 Quintessence

Quintessence was introduced to solve the coincidence problem. It is a classical scalar field ϕ and has a potential $V(\phi)$. The background energy density and pressure are

$$\bar{\rho}_\phi = \frac{1}{2} \dot{\phi}^2 + V(\phi) \quad (7.19)$$

and

$$\bar{P}_\phi = \frac{1}{2} \dot{\phi}^2 - V(\phi) \quad (7.20)$$

respectively. This means that the equation of state is in general time varying:

$$w_\phi = \frac{\dot{\phi}^2 - 2V}{\dot{\phi}^2 + 2V} \quad (7.21)$$

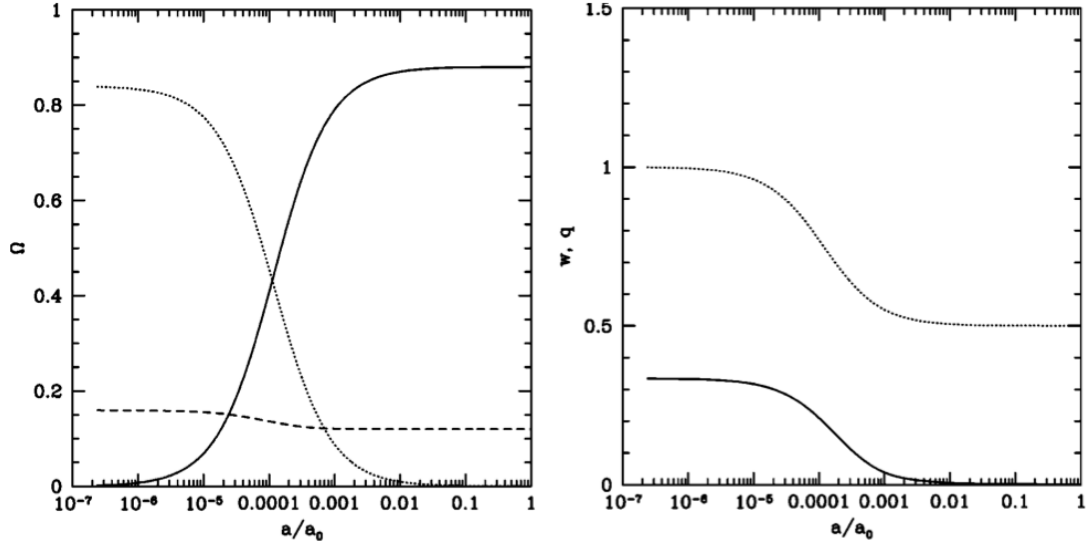


Figure 7.7: Left: The evolution of the relative density of ϕ (dashed), matter (solid) and radiation (dotted) in the simple exponential model. Right: The evolution of the equation of state w (solid) and deceleration parameter (dotted) for the simple exponential model.

The scalar field also obeys a Klein-Gordon equation

$$\ddot{\phi} + 3H\dot{\phi} + \frac{dV}{d\phi} = 0 \quad (7.22)$$

Since the equation of state is time-varying, quintessence has an adiabatic speed of sound c_a^2 not equal to w . In particular we find using $c_a^2 = \dot{P}/\dot{\rho}$ and using the Klein-Gordon equation to eliminate $\ddot{\phi}$ that

$$c_a^2 = 1 + \frac{2 \frac{dV}{d\phi}}{3H\dot{\phi}} \quad (7.23)$$

As you may suspect, all the dynamics of quintessence lie in the potential. Numerous potentials have been proposed and each one is supposed to have a specific purpose, i.e. to solve the coincidence problem, or to be well motivated from particle physics or string theory or anyother thing the authors had in mind.

One particular classification is the following:

- Freezing models. Here the field rolls down the potential in the past but the movement gradually slows down after the system enters the phase of acceleration. Examples are
 - $V = M^{4+n}\phi^{-n}$. This is the Ratra-Peebles (1988) potential, which was later revived by Zlatev, Wang and Steinhardt(1999).
 - $V = M^4 [(\phi/M_p - B)^2 + A] e^{-\lambda\phi/M_p}$. This is the Albrecht-Skordis model (1999) (henceforth called AS model).
 - $V = M_p^4 [e^{\alpha\phi/M_p} + e^{\beta\phi/M_p}]$. This is the double-exponential model of Barreiro-Copeland-Nunes (2001).

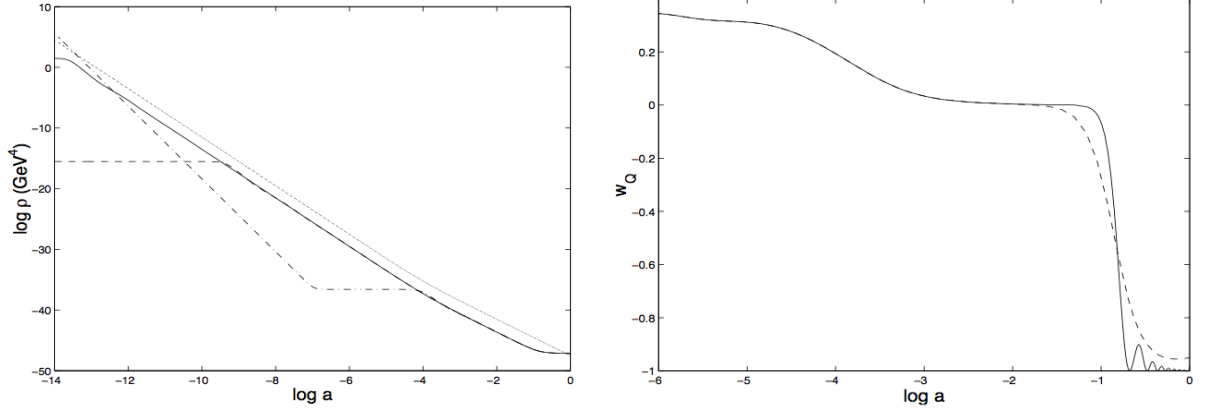


Figure 7.8: Left: The evolution of the energy density for the double exponential model. Dotted line is $\rho_r + \rho_m$ while solid and dashed lines is a double exponential model with three different initial conditions.

Right: The evolution of the equation of state w for the double exponential model (two sets of parameters).

- Thawing models. Here the field is frozen in the past by the Hubble friction (the term $H\dot{\phi}$) until recently when it begins to evolve once H drops below the mass of the field m_ϕ . The equation of state is always -1 in the past and only recently does it start to deviate from -1 . Examples are,
 - $V = V_0 + M^{4-n}\phi^n$, for $n > 0$. This is similar to chaotic inflation ($n = 2, 4$).
 - $V = M^4 \cos^2(\phi/f)$. This is the pseudo-Nambu goldstone model. It is a particle physics motivated model, where ϕ is a pseudo-scalar (e.g. an axion).

There are many more models not considered here. Have a look at Ed's review "Dynamics of Dark Energy" (with Sami and Tsujikawa) for a broader list.

Let's pick two models completely at random: the double exponential and the AS model. They are both based on the simple exponential potential $V = V_0 e^{-\lambda\phi/M_p}$ where V_0 and λ are parameters. The simple exponential potential cannot lead to an accelerating Universe (unless λ is very small) and V_0 is of the order of the cosmological constant. (see Ferreira and Joyce 1995). Typically what happens is that the field ϕ mimics the behaviour of the dominant form of matter so that $w_\phi = w_{dom}$. So in the radiation era ϕ behaves like radiation while in the matter era it behaves like matter. The value of Ω_ϕ is always smaller than the dominant component and is given by

$$\Omega_\phi = \frac{3(1 + w_{dom})}{\lambda^2} \quad (7.24)$$

The above behaviour of the exponential model is independent of initial conditions. This means that the above tracking behaviour is not fine-tuned.

Now to the two quintessence models. Both models look like the simple exponential model in the past, but deviate from it today. What happens in both cases is that the modification of the simple exponential model is to introduce a local minimum in the potential. So the field eventually gets trapped into the minimum and starts behaving as a cosmological constant. As you may suspect, although the fine-tuning of the initial conditions has been removed, there is still fine-tuning left as the $V_{min} \sim \rho_\Lambda$ in order to get the right amount of acceleration.

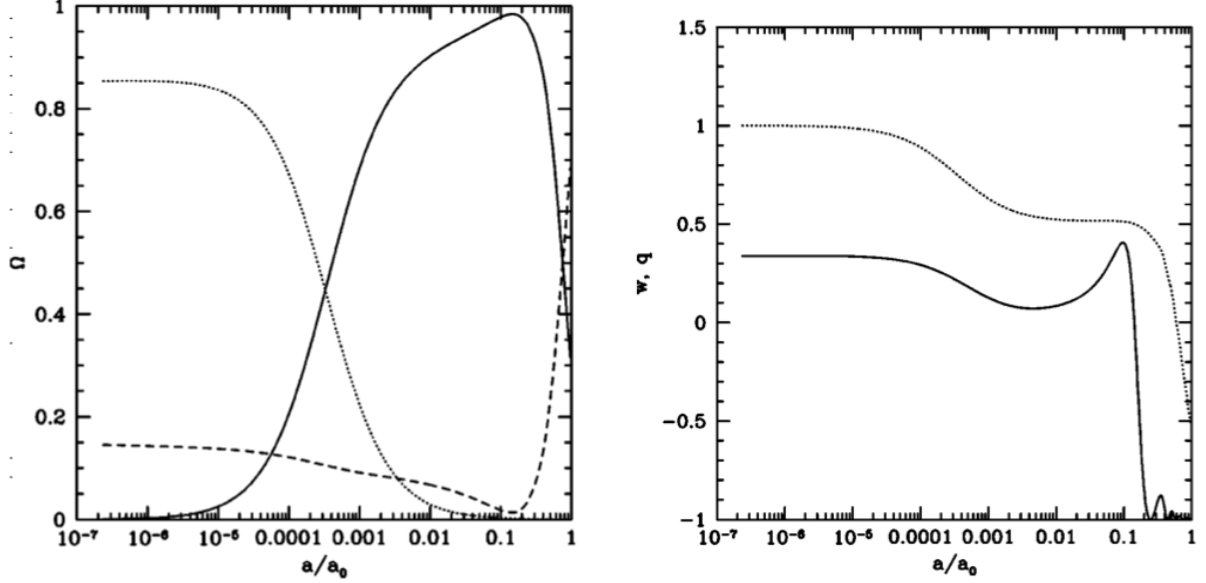


Figure 7.9: Left: The evolution of the relative energy density for the AS model. Dotted line is $\rho_r + \rho_m$ while solid and dashed lines is a double exponential model with three different initial conditions.

Right: The evolution of the equation of state w (solid) and deceleration parameter (dotted) for the AS model.

A generic problem with quintessence is that in general the mass of the field ϕ has to be very small. The mass of the field is $m_\phi^2 \sim \frac{d^2V}{d\phi^2}$. This is of the same order as $\sim V/\phi^2$ but since to get acceleration we need $V(\text{today}) \sim \rho_\Lambda \sim H_0^2 M_p^2$ we get that $m_\phi \sim H_0 \sim 10^{-33} \text{eV}$. This is a tiny mass by all standards which means that quintessence is effectively a massless scalar field. Thus if it couples to anything else (and it should if we include quantum corrections) then it would mediate a 5th force on solar system scales, which has not been observed.

7.3.3 K-essence

K-essence (Armendariz-Picon, Mukhanov and Steinhardt, 2000) is an attempt to use higher powers of the kinetic term rather than a potential to provide for dark energy. A similar idea (by the same people) exists for inflation. The simplest K-essence does not have a potential but rather it has a free function of the kinetic term: $F = F(X)$ where $X = \frac{1}{2}\dot{\phi}^2$. The energy density and pressure are

$$\bar{\rho}_\phi = 2X \frac{dF}{dX} - F \quad (7.25)$$

and

$$\bar{P}_\phi = F \quad (7.26)$$

so that

$$w_\phi = \frac{F}{2X \frac{dF}{dX} - F} \quad (7.27)$$

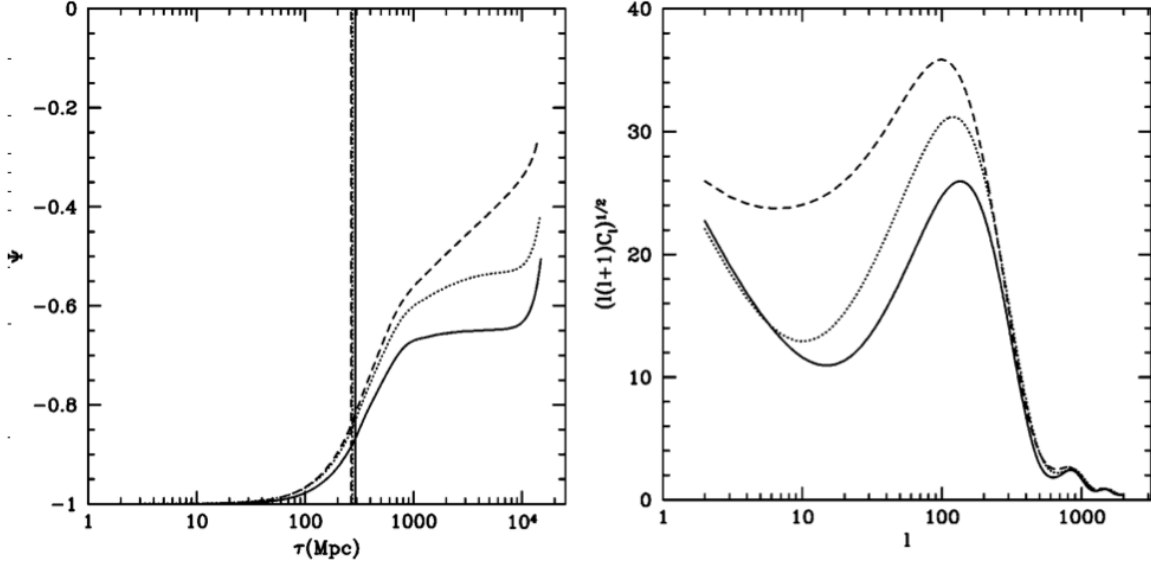


Figure 7.10: Left: The evolution of the Newtonian potential Ψ for Λ CDM (solid), AS model (dotted) and a brane model (dashed). Last scattering is at the vertical line.

Right: The Integrated Sachs-Wolfe effect for the same models as on the left. Note, this is only the ISW part of the C_ℓ spectrum as the primary anisotropies have been removed by hand. The quintessence models have a greater ISW effect than Λ CDM which also extends over a wider range of scales.

If we let $F = \frac{1}{2}\dot{\phi}^2$ we recover a quintessence-like model without a potential. The evolution for a typical k-essence model is shown in figure 7.11. Typically the k-essence field will like radiation in the past, then as cosmological constant and eventually as a constant- w component in the future (called the k-attractor). The value of the constant- w depends on the parameters of $F(X)$.

7.3.4 Coupled dark energy

Since neither the nature of dark energy nor dark matter has been deciphered it is natural to investigate whether we can couple them together. In principle it is possible of course to couple dark energy with anything, however, it is very dangerous to couple it to photons or baryons as that sector is very well constrained. To couple the dark energy to dark matter we introduce a coupling function $Q(t)$. This is in addition to the equation of state of dark energy. Furthermore, we must decide how the coupling manifests. We have seen that photons and baryons, for example, are coupled together via Compton scattering. This has the consequence that the energy momentum tensor for each individual species is not conserved but only the total energy momentum tensor is conserved. Rather we have that dark matter evolves according to

$$\dot{\rho}_m + 3H\rho_m = Q \quad (7.28)$$

and dark energy according to

$$\dot{\rho}_{DE} + 3H(1 + w_{DE})\rho_{DE} = -Q \quad (7.29)$$

so that the total $\rho_m + \rho_{DE}$ is conserved. Using coupled dark energy one can get acceleration even from the simple exponential model (see Amendola 2000).

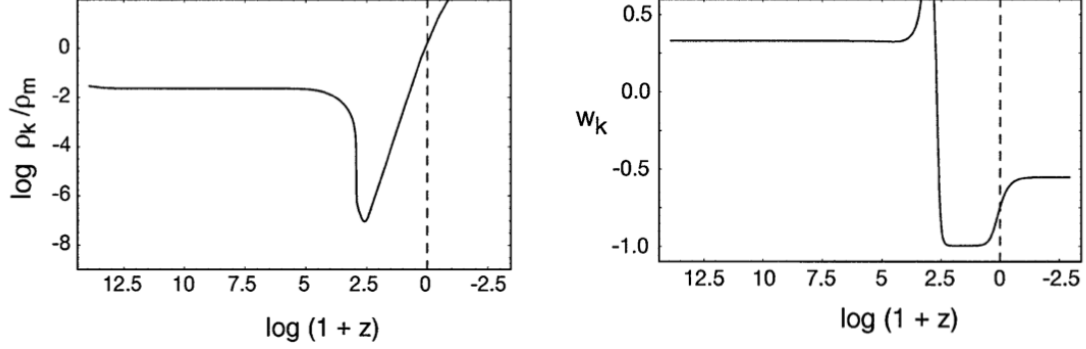


Figure 7.11: Left: The evolution of the ratio of k-essence energy density to the matter energy density versus redshift. The k-essence energy density is at a fixed ratio in the past but after a short dip starts to dominate today.

Right: The evolution of the equation of state w of k-essence versus redshift. K-essence evolves like radiation in the past, then as cosmological constant and eventually as a constant- w component in the future.

7.4 Modifications of gravity

Since we have no idea what may be the dark energy, it has been suggested and is now becoming a popular topic, that perhaps General Relativity breaks down on cosmological scales. One popular idea (but otherwise boring, not to mention ruled out) is $f(R)$ gravity. Here the Einstein-Hilbert action is modified to be

$$S = \frac{1}{16\pi G} \int d^4x \sqrt{-g} f(R) \quad (7.30)$$

where $f(R)$ is a free function of R . These kind of theories have been well constrained by the data so that $f(R) \approx R - 2\Lambda + \text{very small correction}$, i.e. the data permits them to only look like GR + cosmological constant. Moreover $f(R)$ theories typically lead to extra long range forces in the solar system and need a screening mechanism to save them from complete embarrassment.

A different kind of modified gravity model is the very popular Dvali-Gabadadze-Porrati (DGP) model. Here the universe is 5 dimensional and we live on a 3-brane. The 5th dimension becomes manifest on large scales and the effect is cosmic acceleration. The scale corresponding to the manifestation of the 5th dimension is r_c so that the Friedmann equation gets modified to

$$3H^2 - \frac{3H}{r_c} = 8\pi G\rho \quad (7.31)$$

For $H \gg r_c^{-1}$ we recover the usual Friedmann equation but as $H \sim r_c^{-1}$ the r_c term becomes important and eventually the Universe enters an era of acceleration with an effective cosmological constant given by $\rho_\Lambda = \frac{3}{8\pi G r_c^2}$. This is called the self-accelerating Universe.

The DGP model has two severe problems. The first is that the self-accelerating solution has a ghost instability and cannot be a realistic solution. The second is that current data strongly disfavour the DGP model.