Modern Cosmology

Anisotropies and Inhomogeneities in the Universe

Scott Dodelson

Contents

Chapter 1

Pillars of Big Bang Cosmology

Einstein's discovery of general relativity in the last century enabled us for the first time in history to come up with a compelling, testable theory of the universe. Big Bang cosmology posits a smooth, expanding universe. This very simple framework, together with knowledge of sub-atomic physics, allows us to extrapolate backwards in time to understand the history of the universe. The realization that the universe is expanding and was once much hotter and denser leads to a set of profound questions about early times. How did the elements form? Why is there more matter than anti-matter in the universe? Why is the universe so smooth? These questions and many like them have answers, answers that can be found only by combining our knowledge of particle and nuclear physics with our understanding of the conditions in the early universe.

This chapter describes the idea of an expanding universe, without using the equations of general relativity. Then the consequences of the Big Bang theory for the early universe are outlined, culminating in the timeline of figure 1.4.

All of this would be pure speculation if not for the observational evidence. The success of the Big Bang rests of three observational pillars: the Hubble diagram exhibiting expansion; light element abundances which are in accord with Big Bang Nucleosynthesis; and the black-body radiation left over from the first few hundred thousand years, the Cosmic Microwave Background. The final three sections of the chapter introduce these pieces of evidence and set the stage for our discussions in the rest of the book about perturbations around the zero order, smooth universe.

1.1 The Expanding Universe

We have good evidence that the universe is expanding. This means that early in its history the distance between us and distant galaxies was smaller than it is today. It is convenient to describe this effect by introducing the scale factor a, whose present value is set to one. At earlier times a was smaller than it is today. We can picture space as a grid as in figure 1.1 which expands uniformly as time evolves. Points on the grid maintain their coordinates, so the "comoving distance" between two points – which just measures the difference between coordinates – remains constant. However, the physical distance is proportional to the scale factor, and the physical distance does evolve with time.

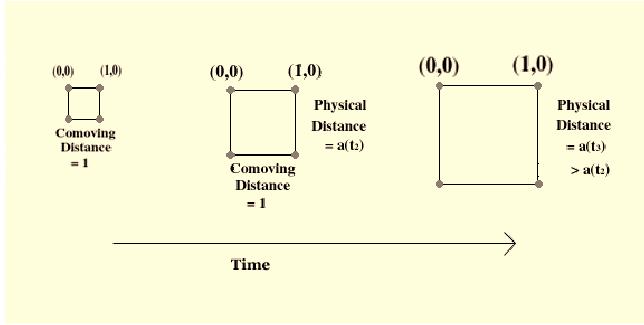


Figure 1.1: Expansion of the universe. The comoving distance between points on a hypothetical grid remains constant as the universe expands. The physical distance is proportional to the comoving distance times the scale factor, so it gets larger as time evolves.

In addition to the scale factor and its evolution, the smooth universe is characterized by one other parameter, its geometry. There are three possibilities: flat, open, or closed. A *flat* universe is Euclidean: parallel lines remain parallel, and light travels in straight lines. General relativity connects geometry to energy. Accordingly, a flat universe is one in which the energy density is equal to a critical value, which we will soon see is approximately 10^{-29} g cm^{-3}. If the density is higher than this value, then the universe is *closed*: parallel lines converge, just as all lines of constant longitude meet at the North and South Poles. Such a universe will eventually stop expanding, start contracting, and end in a Big Crunch. Finally, a low density universe is *open*, so that parallel lines diverge, as would two marbles rolling off a saddle.

To understand the history of the universe, we must determine the evolution of the scale factor a. Again, general relativity provides the connection between this evolution and the energy in the universe. If the universe is flat and dominated by non-relativistic matter, then

$$a(t) = \left(\frac{t}{t_0}\right)^{2/3}. \tag{1.1}$$

(Here and throughout the subscript $_0$ refers to the present values of quantities.) We can differentiate equation 1.1 to get a measure of how fast the universe is expanding. The Hubble rate is a measure of this rate:

$$H(t) \equiv \frac{da/dt}{a} = \frac{2}{3}\frac{1}{t^{1/3}t_0^{2/3}}. \tag{1.2}$$

The last equality holds only in the simple cosmology in which the universe is flat and matter dominated. Thus a powerful test of this cosmology is to measure separately the Hubble

rate today, H_0, and the age of the universe today. In a flat, matter-dominated universe, the product of the two should equal 2/3.

In general, the evolution of the scale factor is determined by Einstein's equation, which is conveniently written as

$$\frac{H^2(t)}{H_0^2} = \frac{\Omega_m}{a^3(t)} + \frac{\Omega_R}{a^4(t)} + \Omega_\Lambda + \frac{1-\Omega_\Lambda-\Omega_m-\Omega_R}{a^2(t)}, \tag{1.3}$$

where Ω_m is the ratio of the energy density of non-relativistic matter in the universe to the critical density required to make the universe flat. Similarly, Ω_Λ is ratio of the energy density in the vacuum to the critical density, and Ω_R is the same ratio for relativistic species such as photons and neutrinos. The critical density is

$$\rho_{\rm cr} \equiv \frac{3H_0^2}{8\pi G} \tag{1.4}$$

where G is Newton's constant.

In §2.1 we will derive these equations from first principles; now though, let's try to understand the the scaling with a of the different terms in equation [1.3]. These terms are all energy densities divided by the critical density, so the question of how they scale reduces to the question of how energy densities of the different components scale as the universe expands.

The photons which comprise the Cosmic Microwave Background (CMB) today have a well-measured temperature $T_0 = 2.728 \pm 0.002K$. A photon with an energy of order $k_B T_0$ today has a wavelength of order $\hbar c/k_B T_0$. Early on, when the scale factor was smaller than it is today, this wavelength would have been correspondingly smaller. Since the energy of a photon is inversely proportional to its wavelength, the photon energy would have been larger than today by a factor of $1/a$. This argument applied to the thermal bath of photons implies that the temperature of the plasma as a function of time is:

$$T(t) = T_0/a(t). \tag{1.5}$$

At early times the temperature was higher than it is today. This has dramatic implications: for one thing, it implies that number densities, which scale as T^3, were also much higher in the early universe. The energy density of radiation, the product of number density times average energy per particle, therefore scales as a^{-4}. This explains the scaling of the Ω_R term in equation [1.3]. To understand the matter term, recall that the energy per particle of non-relativistic matter is simply the rest mass, which remains constant as the universe expands. Hence, the energy density of matter scales as a^{-3}.

The last two terms in equation [1.3] are more speculative. Recent evidence from distant Supernovae suggest that there may well be energy besides that found in ordinary matter. Equation 1.3 implicitly assumes that this excess energy density is constant at all times, the canonical "cosmological constant" first introduced by Einstein. This assumption is not necessary to account for the data, so cosmologists have explored other forms of excess energy , many of which behave very differently than the cosmological constant. We will see more of this in later chapters. Finally, I have allowed for the possibility that the universe is not

flat: if it were flat, the sum of all the energy densities would equal the critical density, and the numerator in the last term in equation [1.3] would vanish. If the universe is not flat, the *curvature* energy scales as $1/a^2$. In most of this book we will work within the context of a flat universe. In such a universe, the evolution of perturbations is much easier to calculate than in open universes ($\Omega_0 < 1$). Further there have been a number of recent experiments which strongly support the flatness of the universe. More on this in Chapter 7.

Figure 1.2 illustrates how the different terms in equation 1.3 vary with the scale factor. While today matter, and possibly a cosmological constant dominate the landscape, early on, due to the a^{-4} scaling, radiation was the dominant constituent of the universe.

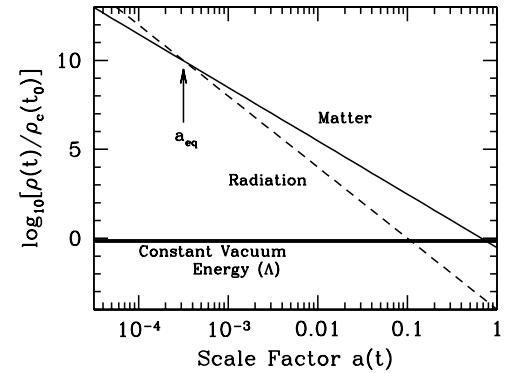


Figure 1.2: Energy density vs. scale factor for different constituents of a flat universe. Shown are non-relativistic matter, radiation, and vacuum energy. All are in units of the critical density today. Even though matter and vacuum energy dominate today, at early times, the radiation density was largest. The epoch at which matter and radiation are equal is $a_{\rm EQ}$.

Let's introduce some numbers. Present measures of the Hubble rate are parameterized by h defined via

$$\begin{aligned} H_0 &= 100h \ \text{km sec}^{-1}\text{Mpc}^{-1} \\ &= 2.133\times 10^{-33} h \ \text{eV}/\hbar \ = \frac{h}{0.98\times 10^{10}\text{years}} \end{aligned} \tag{1.6}$$

where h has nothing to do with Planck's constant $\hbar$. The astronomical length scale of a Megaparsec (Mpc) is equal to 3.0856×10^{24}cm. Current estimates for h range from 0.6 to 0.8. The predicted age for a flat matter dominated universe, $(2/3)H_0^{-1}$, is then of order 8 to 11 Gyr. Different dating techniques get ages ranging from 12 Gyr on up, so this test suggest that a flat, matter dominated universe is barely viable. You will show in Problem (2) that the age of the universe with a cosmological constant is larger (for fixed h); in fact one of the original arguments in favor of Λ was to make the universe older.

Newton's constant in equation [1.4] is equal to $\hbar c^5/(1.22\times 10^{28}\mathrm{eV})^2$. This, together with equation [1.6], enables us to get a numerical value for the critical density:

$$\begin{aligned}\rho_{\rm cr} &= 8.10h^2\times 10^{-11}\mathrm{eV}^4/(\hbar c)^3\\ &= 1.88h^2\times 10^{-29}\mathrm{g\ cm}^{-3}.\end{aligned} \tag{1.7}$$

An important ramification of the higher densities in the past is that reaction rates, which scale as the density, were also much higher early on. Figure 1.3 shows some important rates

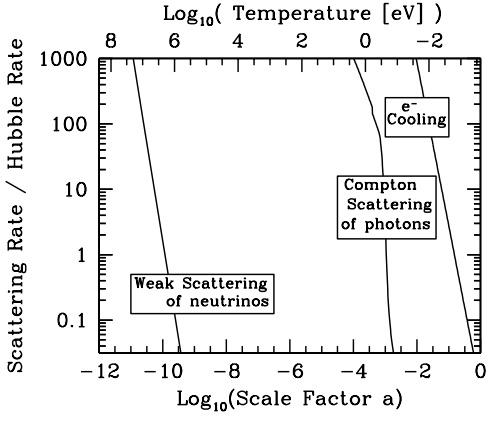


Figure 1.3: Rates as a function of the scale factor. When a given rate becomes smaller than the expansion rate H, that reaction falls out of equilibrium.

as a function of the scale factor. For example, when the temperature of the universe was greater than several MeV, the rate for the process $e^- + \nu_e \leftrightarrow e^- + \nu_e$ was larger than the expansion rate. Thus, before the universe could double in size, a neutrino scattered many times off background electrons. All these scatterings brought the neutrinos into equilibrium with the rest of the cosmic plasma. This is but one example of a very general, profound fact: if a particle scatters with a rate greater than the expansion rate, that particle stays in equilibrium. Once the scattering rate drops beneath the expansion rate, however, the particle loses contact with the rest of the cosmic plasma. Many particles in the early universe underwent this process of *decoupling*. Three examples – including the most important for our purposes, photon decoupling – are shown in figure 1.3. We can use this qualitative understanding of decoupling to trace out a timeline of the universe as the temperature evolves.

1.2 Timeline

We can characterize any epoch in the universe by either the time since the Big Bang; the value of the scale factor at that time; or by the temperature of the cosmic plasma. For example, today, $a = 1$; $t \simeq 15$ billion years; and $T = 2.728°$K. Figure 1.4 shows a timeline of the universe using both time and temperature as markers. The milestones indicated on the timeline range from those about which we are quite certain (e.g. Nucleosynthesis and the CMB) to those much more speculative (e.g. dark matter production and vacuum eras at the beginning and end of the band).

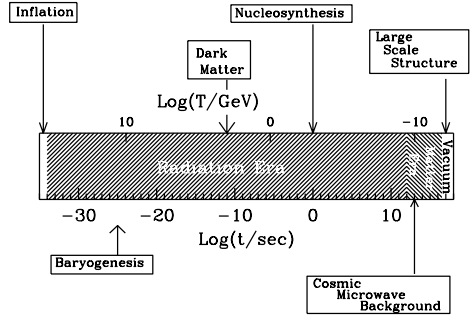


Figure 1.4: One model of the early universe. Any epoch can be associated with either temperature (top scale) or time (bottom scale). The white regions at either end of the timeline indicate vacuum-dominated epochs.

1.2.1 Dark Matter

We start with the production of dark matter when the temperature of the universe was on the order of 100 GeV. Dark matter refers to mass in the universe that contributes to the gravitational field but does not emit light. There are many arguments, most of them indirect but nevertheless quite compelling, that dark matter exists in the universe and further that it is not "ordinary" stuff, i.e. not electrons and protons. Most of this book will assume the existence of such non-baryonic dark matter. Indeed one of the strongest arguments for dark matter is that without it, it is very difficult to account for the observed anisotropies and inhomogeneities in the universe.

The timeline in Figure 1.4 implicitly assumes that the particles have masses on the order of a TeV, such as might characterize supersymmetric candidates. When the temperature was above a TeV, scattering processes for these particles were in equilibrium, so they were as abundant as photons and all other known particles. Once the temperature dropped beneath

their mass, they remained in equilibrium, but the equilibrium abundance was Boltzmann suppressed: $n \propto e^{-m/T}$. If equilibrium was maintained down to arbitrarily low temperatures, the exponential suppression would have eliminated all traces of the particles today.

However, the reaction rates for annihilation of these particles are proportional to their density, n. Since their density was decreasing so rapidly, eventually they fell out of equilibrium. At that point, called either decoupling or *freeze-out*, the abundance stopped its exponential descent. Thereafter, the ratio of these heavy particles to other particles in the plasma such as photons remained fixed (except for some brief moments when additional energy was pumped into the plasma by annihilations of lighter particles). The precise epoch of freeze-out is determined by the cross-sections, but typically $T_{\text{freeze-out}}$ is of order $m/25$. The energy density in such heavy particles today is thus of order $mn_{\text{photons}}e^{-25}$, which is close to the critical density for TeV masses.

1.2.2 Proton–Anti-proton annihilation

The annihilation dance repeated as the temperature dropped below the mass of the proton. However, here there is one twist. Initially there was a small excess, one part in ten billion, of protons over anti-protons. So even before standard freeze-out occurred, all anti-protons were eliminated, leaving the remaining protons with no partners with which to annihilate. This small excess, presumably set by primordial processes labeled on the timeline *baryogenesis*, therefore determines the proton (and electron since charge is conserved) abundance today. By the time the temperature dropped to ten MeV, there were no anti-protons, and the proton/photon ratio was set to its present value on the order of 10^{-10}.

1.2.3 Neutrino decoupling

Once the temperature dropped below an MeV, neutrinos lost contact with the rest of the plasma. Neutrinos interact only weakly with electrons and nuclei, so their interaction rate dropped beneath the Hubble rate even without the Boltzmann suppression which drives the freeze-out of heavy particles. After decoupling, the neutrinos maintained the thermal spectrum which had been set up by collisions. In keeping with our previous argument, though, the temperature characterizing this spectrum continued to fall as $1/a$. If no more energy was input into the photon gas, both neutrinos and photons would have thermal distributions with identical temperatures today, despite the fact that they decoupled very early on.

1.2.4 Electron–Positron Annihilation

After neutrinos decoupled, the photons remained coupled to electrons and positrons, all of which had similar abundances at temperatures of order an MeV. As the temperature dropped below the electron mass (.511 MeV), electrons and positrons began annihilating into photons. Just as with protons, only one in ten billion electrons survived. All of the annihilation energy went into photons. This served to heat the photons up with respect to the neutrinos, so the

photon temperature thereafter is slightly larger – by a factor of $(4/11)^{1/3}$ – than the neutrino temperature.

1.2.5 Big Bang Nucleosynthesis

Around this time, the rate for the process

$$p + e^- \leftrightarrow n + \nu_e \tag{1.8}$$

became smaller than the expansion rate. The ratio of protons to neutrons was fixed when this reaction froze out. Most of the neutrons ended up in helium atoms (some decayed into protons and a trace ended up in other light elements) so the n/p ratio at this time directly determines the helium/hydrogen ratio in the universe today, as well as the other light element abundances. We will see shortly that the confirmation of this prediction is one of the great successes of the Big Bang.

1.2.6 Decoupling at Recombination

For our purposes, the most important incidence of decoupling occurred when the expansion rate became larger than the rate for photons to scatter off electrons via Compton scattering:

$$\gamma + e^- \leftrightarrow \gamma + e^-. \tag{1.9}$$

We will shortly calculate this rate explicitly; for now we note that it falls below the Hubble rate when $k_B T \sim 1/3$ eV corresponding to $a \sim 10^{-3}$. After this, the photons decoupled, maintaining a thermal spectrum with a falling temperature.

This brief list of the highlights of the early universe illustrates the importance of decoupling to the evolution of the universe. The epoch of photon decoupling will come to dominate our landscape, but it is important to keep in mind that it is but one of a series of decouplings that occurred as the universe cooled.

Once the photons decouple, they travel freely without interacting with any other particles. When we observe them today, they literally come from the earliest moments of time. They are therefore the most powerful probes of the early universe. The cosmic microwave background offers us a look at the universe when it was only 300,000 years old.

The most important fact we learned from our first twenty five years of surveying the CMB was that the early universe was very smooth. No anisotropies were detected in the CMB. This period, while undoubtedly frustrating for observers searching for anisotropies, solidified the view of the Big Bang summarized in the previous few pages. We are now moving on. We have discovered anisotropies in the CMB, indicating that the early universe was not completely smooth. There were small perturbations in the cosmic plasma.

1.2.7 Structure Formation

The small perturbations in the universe began to grow when the universe became dominated by matter. The dark matter grew more and more clumpy, simply due to the attractive nature

of gravity. An overdensity of dark matter of one part in a thousand when the temperature was one eV grew to one part in a hundred by the time the temperature dropped to a tenth of an eV. Eventually, perturbations in the matter ceased to be small; they became non-linear. The resultant structure is rich and varied, and we are probably only beginning to explore structure in the universe. However, all this richness traces back to the initial perturbations. If we are to understand our universe, we must first understand the initial perturbations. To do this, we must at first try to avoid scales dominated by non-linearities. As an extreme example, we can never hope to understand cosmology by carefully examining rock formations on Earth. The intermediate steps – collapse of matter into a galaxy; molecular cooling ; star formation; planetary formation; etc. – are much too complicated to allow to say anything definitive about the initial conditions. The best places to look if we want to focus on the initial perturbations are the CMB, where the anisotropies are still very small, and the matter distribution on large scales. As we will see, while perturbations on small scales (less than about ten Mpc) have grown non-linear, large scale perturbations are still small. So they have been processed much less than the corresponding small scale structure. The focus of this book then will be anisotropies in the CMB and inhomogeneities in the matter distribution on large scales.

1.2.8 Energy Content

The timeline in figure 1.4 illustrates the dominant component of the universe at various times. Early on, most of the energy in the universe was in the form of radiation. Eventually, since the energy of a relativistic particle falls as $1/a$ while that of non-relativistic matter remains constant at m, matter overtook radiation. The epoch at which they are equal is so important for structure formation that we will shortly calculate it in great detail.

First, though, notice the ends of the timeline. At relatively recent times, the universe appears to have become dominated not by pressureless matter, but by some vacuum energy which has negative pressure. The evidence for this unexplained form of energy is new and certainly not conclusive, but it is very suggestive. One of the most fascinating aspects of this hypothesis is that it is eminently testable, and perhaps the best test will come from measurements of the CMB and large scale structure.

The early end of the timeline also indicates an early vacuum dominated period, the epoch of inflation. Until recently, there was little evidence for inflation. It survived as a viable theory mainly because of its aesthetic appeal. The discoveries of the last several years have changed this. First, they have by and large confirmed some of the basic predictions of inflation. More importantly, they have alerted us to the testability of the theory. Inflation makes robust predictions about the nature of the primordial perturbations. These predictions can and will be tested within the next decade. This is a remarkable statement in light of the energy scale associated with inflation. It suggests that we can learn something about physics at energies of order 10^{15} GeV, length scales of order 10^{-30} cm, scales far beyond those accessible to accelerators.

Some of the elements in the timeline in figure 1.4 are probably incorrect. However, since most of these ideas are testable, the data which will be taken during the coming decade will tell us which parts of the timeline are correct and which need to be discarded. This in itself

seems a sufficient reason to study the CMB and large scale structure.

1.3 The Hubble Diagram

If the universe is expanding as depicted in figure 1.1, then galaxies should be moving away from each other. We should therefore see galaxies receding from us. Recall that the wavelength of light or sound emitted from a receding object is stretched out so that the observed wavelength is larger than the emitted one. It is convenient to define this stretching factor as the redshift z:

$$1 + z \equiv \frac{\lambda_{\rm obs}}{\lambda_{\rm emit}} = \frac{1}{a}. \tag{1.10}$$

For low redshifts, the standard Doppler formula applies and $z \simeq \frac{v}{c}$. So a measurement of the amount by which absorbtion and/or emission lines are redshifted is a direct measure of how fast they are receding from us.

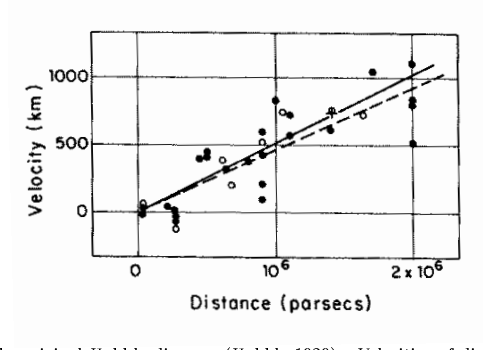


Figure 1.5: The original Hubble diagram (Hubble 1929). Velocities of distant galaxies (units should be km sec^{-1}) are plotted vs. distance. Solid (dashed) line is best fit to the filled (open) points which are corrected (uncorrected) for the sun's motion.

Hubble first found that distant galaxies are in fact receding from us. He also noticed the trend that the velocity increases with distance. This is exactly what we expect in an expanding universe, for the physical distance between two galaxies is $d = ax$ where x is the comoving distance. In the absence of any comoving motion ($\dot{x} = 0$, no *peculiar* velocity) the relative velocity, $v = \dot{d}$ is therefore equal to $\dot{a}x = Hd$. Therefore, velocity should increase linearly with distance (at least at relatively low redshift) with a slope given by H, the Hubble constant. Hubble's Hubble constant can be read off of figure 1.5. It is simply $H = 1000/2$ km sec^{-1} Mpc^{-1}, a factor of almost ten higher than current estimates. Also notice that Hubble's data went out to redshift $z = 1000$km sec$^{-1}/3 \times 10^5$km sec$^{-1} = 0.003$.

The Hubble diagram is still the most direct evidence we have that the universe is expanding. Current incarnations use the same principle as the original: find the distance and the

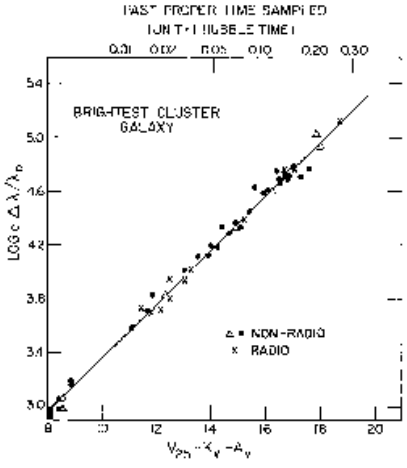


Figure 1.6: Hubble diagram from the late 1960's (Sandage 1968). The $y-$ axis is cz while the $x-$ axis is the apparent magnitude of the brightest galaxy in different clusters. Hubble's initial data fit in the small rectangular box at bottom left.

redshift of distant objects. Measuring redshifts is relatively easy; the hard part is determining distances for objects of unknown intrinsic brightness. One of the most popular techniques is to try to find a *standard candle*, a type of object whose intrinsic brightness can be shown to be almost constant. Figure 1.6 shows a Hubble diagram from the 1960's wherein the standard candle is the brightest galaxy in a cluster. The relatively small dispersion around the straight line is proof that this is a reasonable standard candle.

In recent years, one of the most exciting and unexpected developments in the field has been the emergence of Type Ia Supernovae as reliable standard candles. Figure 1.7 shows a recent Hubble diagram using distant supernovae. Since they are so bright, supernovae can be used to extend the Hubble diagram out to very large redshifts (the current record is of order $z \simeq 1$), a regime where the simple Doppler law ceases to work. In the next chapter, we will derive the correct expression for the distance (in this case the *luminosity* distance) as a function of redshift. For now, you should note that this expression depends on the energy content of the universe. The three curves in the top panel of figure 1.7 depict three different possibilities: flat matter dominated; open; and Λ dominated. The high redshift data is now apparently good enough to distinguish among these possibilities, strongly disfavoring the previously favored flat, matter dominated universe. The current best fit is a universe with about 70% of the energy in the form of a cosmological constant, or some other form of energy with negative pressure.

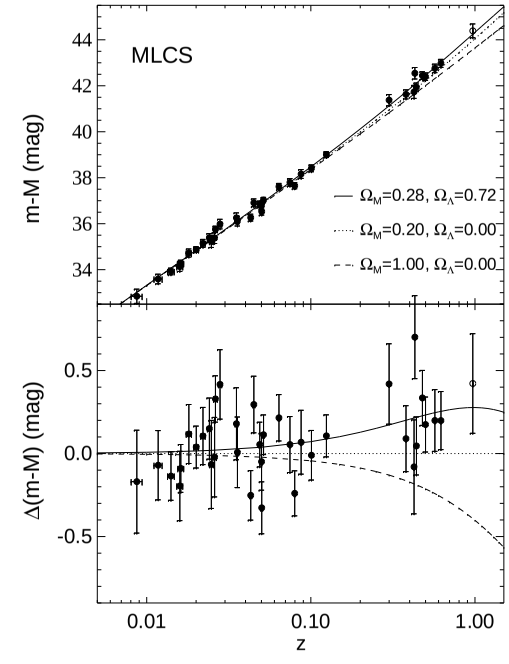


Figure 1.7: Hubble diagram from distant Type Ia supernovae. Top panel shows apparent magnitude (an indicator of the distance) vs. redshift. Lines show the predictions for different energy contents in the universe. Bottom panel plots the residuals, making it clear that the high redhisft supernovae favor a Λ dominated universe over a matter dominated one.

1.4 Big Bang Nucleosynthesis

Figure 1.8 shows the predictions of Big Bang Nucleosynthesis (BBN) for the light element abundances. Note that the four displayed abundances vary over ten orders of magnitude. As the baryon density increases, the reactions necessary to build up ^{4}He (e.g. $pn \to D\gamma$) proceed more rapidly. Therefore, a large baryon density leads to earlier formation of ^{4}He, at a time when there are more neutrons. producing more helium. This explains the monotonic increase in helium as the baryon density increases. The traces of D and ^{3}He are left over because these small amounts cannot be processed into ^{4}He once the reaction rates become too low. If the baryon density is high, fewer of these light elements are left unprocessed: they can all be turned into ^{4}He. Therefore, as the baryon density increases, the abundance of deterium and ^{3}He decreases. It is especially important that the D abundance decreases so dramatically as a function of baryon density.

The boxes and arrows in figure 1.8 show the current estimates for the light element abundances. These are consistent with the predictions, and this consistency test provides yet another ringing confirmation of the Big Bang. The measurements do even more though: they provide perhaps the best estimate to date of the baryon density. In particular, the measurement of primordial deuterium pins down the baryon density extremely accurately to

$$\Omega_B h^2 = 0.019 \pm 0.002. \tag{1.11}$$

(The label at the top of figure 1.8 takes $h = 0.65$.) This suggests that ordinary matter contributes at most 5% of the critical density. Since the total matter density today is almost certainly larger than this – direct estimates give values of order $20 - 30\%$ – nucleosynthesis provides another argument for non-baryonic dark matter. We will see in later chapters that the baryon density also affects the anisotropies in the CMB and, to a lesser extent, the distribution of matter in the universe. There is great hope that future measurements of the CMB and large scale structure will therefore measure Ω_B, testing this prediction of nucleosynthesis.

The deuterium measurements are the new developments in the field. These measurements are so exciting because they explore the deuterium abundance at redshifts of order $3-4$, well before much processing could have altered the primordial abundances. Figure 1.9 shows one such detection. The basic idea is that light from distant QSOs is absorbed by intervening neutral hydrogen systems. The key absorbtion feature is the $n = 1 \to n = 2$ transition, which produces a trough in the spectrum at $\lambda = 1215.7$Å, redshifted by a factor of $1 + z$. The corresponding line from deuterium should be (i) shifted over by $0.33(1+z)$Å(see Problem (5)) and (ii) much less damped since there is much less deuterium. Figure 1.9 shows just such a system; there are now half-dozen with detections precisely in the neighborhood shown in figure 1.8. Note that the steep decline in D as a function of baryon density helps here: even relatively large errors in D measurements translate into small errors on the baryon density.

1.5 The Cosmic Microwave Background

The photons in the Cosmic Microwave Background last scattered at redshift 1100; since then they have travelled freely through space. We will spend an inordinate amount of time in

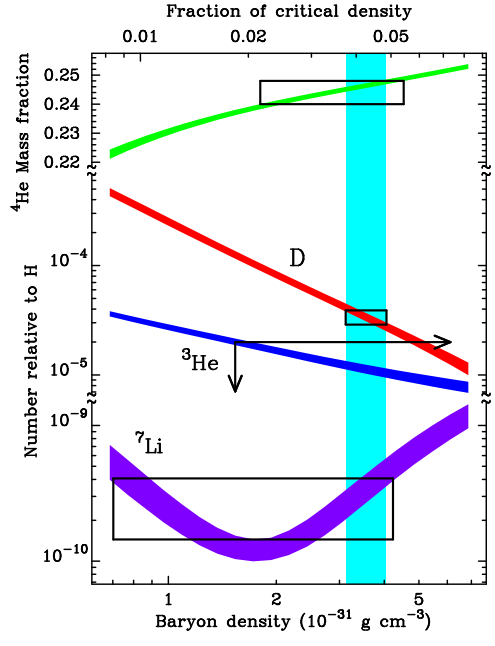


Figure 1.8: Contraint on the baryon density from Big Bang Nucleosynthesis (Burles, Nollett, and Turner 1999). Predictions are shown for four light elements spanning a range of ten orders of magnitude. Solid vertical band is fixed by measurements of primoridial deuterium.

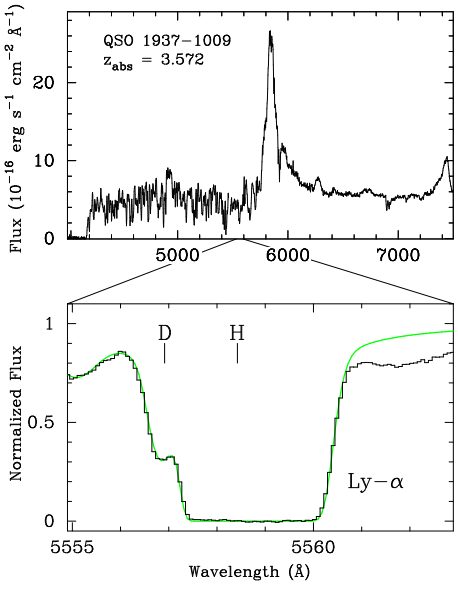


Figure 1.9: Spectrum from a distant QSO (Burles, Nollett, and Turner 1999). Absorbtion of photons with rest wavelength 1216Åcorresponding to the $n = 1$ to $n = 2$ state of hydrogen is redshifted up to 1216(1 + 3.572)Å. Bottom panel provides details of the spectrum in this range, with the the presence of deuterium clearly evident.

this book working through the details of what happened before the epoch of last scattering and also developing the mathematics of the freestreaming process since then. The most important qualitative fact about this history, though, is that the collisions with electrons before last scattering ensured that the photons were in equilibrium. That is, they should have a blackbody spectrum.

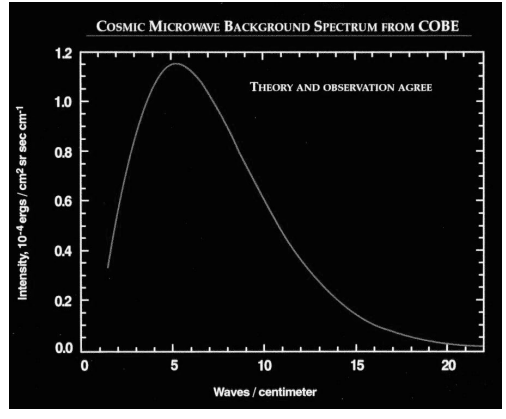


Figure 1.10: Intensity of cosmic microwave radiation as a function of wavenumber. Hidden in the theoretical blackbody curve are dozens of measured points, all of which have uncertainties smaller than the thickness of the curve!

The specific intensity of a gas of photons with a blackbody spectrum is

$$I_\nu = \frac{4\pi\hbar\nu^3/c^2}{\exp\{2\pi\hbar\nu/k_BT\} - 1}. \tag{1.12}$$

Figure 1.10 shows the remarkable agreement between this prediction (see Problem (6)) of Big Bang cosmology and the observations by the Firas instrument aboard the COBE spacecraft. We have been told* that detection of the 3°K background by Penzias and Wilson in the mid-60's was sufficient evidence to decide the controversy in favor of the Big Bang over the Steady State universe. Penzias and Wilson though measured the radiation at just one wavelength. If even their one-wavelength result was enough to tip the scales, the current data depicted in figure 1.10 should send skeptics from the pages of physics journals to the far reaches of radical internet chat groups.

*For a fascinating first-hand account of the history of the discovery of the CMB, see chapter 1 of Partridge(1995).

Suggested Reading

There are many good textbooks covering the homogeneous Big Bang. I am most familiar with *The Early Universe* (Kolb and Turner), which has especially good discussions on nucleosynthesis and inflation. Peacock's *Cosmological Physics* is the most up-to-date and perhaps the broadest of the standard cosmology texts, with more of an emphasis on extra-galactic astronomy than either *The Early Universe* or this book. A popular account which still captures the essentials of the homogeneous Big Bang (testifying to the success of the model: it hasn't changed that much in thirty years) is *The First Three Minutes* (Weinberg).

Problems

1.1 Plug in numbers to get the second of equation [1.7]. If there is a cosmological constant today with energy density comparable to the critical density, what is the ratio of ρ_Λ to the critical density when the temperature of the universe was $10^{19} GeV/k_B$, i.e. at the Planck time?

1.2 Assume the universe today is flat with both matter and a cosmological constant. Integrate equation [1.3] to find the present age of the universe. That is, rewrite equation [1.3] as

$$\frac{da}{a}\left[\Omega_\Lambda + \frac{1-\Omega_\Lambda}{a^3}\right]^{-1/2} = H_0 dt$$

and integrate the left side from $a = 0$ (when $t = 0$) until today at $a = 1$. Plot the age as a function of $\Omega_L ambda$ for fixed h.

1.3 Assume that there is only matter and radiation in the universe (no cosmological constant) and that the universe is flat ($\Omega_0 = 1$). Integrate equation [1.3] to determine the times when the cosmic temperature was 1 Mev and 1/3 eV.

1.4 Determine the mass today of a critical density dark matter candidate with annihilation rate $n < \sigma v >= \hbar^2 n/m^2 c$.

(a) First determine the temperature at which this rate equals the Hubble rate, as e.g. given by equation [1.3].

(b) Assume that annihilation stops abruptly after freeze-out, so that the dark matter density scales as the photon density thereafter. What is the ratio of the dark matter density to the photon density?

(c) Given this ratio, what is Ω_{DM} today? For what value of m is this equal to one?

Note that if m is greater than this critical value, $\Omega_{DM} > 1$, which is ruled out. This is a strong argument against stable particles (and therefore dark matter candidates) with masses above this critical value. It is not possible to escape this bound by raising the cross section, because $\sigma = \hbar^2/m^2c^2$ is an upper limit imposed by unitarity.

1.5 Using the fact that the reduced mass of the electron-nucleus in the D atom is larger than in hydrogen, and the fact that the Lyman α ($n = 1 \to n = 2$) transition in H has a wavelength $1215.67\mathring{A}$, find the wavelength of the photon emitted in the corresponding transtion in D. Astronomers often define

$$v \equiv c\frac{\Delta\lambda}{\lambda} \tag{1.13}$$

to characterize the splitting of two nearby lines. What is v for the H-D pair?

1.6 Convert the specific intensity in equation [1.12] into an expression for what is plotted in figure 1.10, the energy per square centimeter per steradian per second, in the number of waves in a centimeter. Note that the $x-$ axis is $1/\lambda$, the wavelength of the photons. Show that the peak of a 2.73°K blackbody spectrum does lie at $1/\lambda = 5\text{cm}^{-1}$

Chapter 2

Navigating the Expanding Universe

Just as the early navigators of the great oceans required sophisticated tools to help them find their way, we will need modern technology to help work through the ramifications of an expanding universe. One of these tools is general relativity, a field so profound that we risk getting carried away by the beauty of the instrument at the expense of finding our way. Another tool, the Boltzmann equation, is so agile that it can be used on many problems of physical interest. Here the danger is that we might take our tool to other lands (e.g. stellar physics, galactic dynamics, even solid state physics) instead to trying to reach our primary goal. Yet another tool is required to deal with the notion of distance in an expanding universe. Different distances are probed when measuring angles on the sky as opposed to fluxes from distant objects.

In this chapter I introduce some of the needed tools, trying very hard not to get distracted by the beauties of general relativity or the ubiquitousness of the Boltzmann equation. Instead, I will use the tools to derive some of the basic results laid down in Chapter 1: the expansion law of equation [1.3], the epoch of equality $a_{\rm eq}$ shown in figure 1.2, the epoch of decoupling a_* from which emerges the CMB, and the luminosity distance needed to understand the implications of the supernovae diagram in figure 1.7.

In this chapter, I also begin using units in which

$$\hbar = c = k_B = 1. \tag{2.1}$$

Many theoretical papers employ these units, so it is important to get accustomed to them. Please work through Problem (1) if you are uncomfortable with the idea of setting the speed of light to one.

2.1 General Relativity

Most of cosmology can be learned with only a passing knowledge of general relativity. One must be familiar with the concept of a metric, understand geodesics, and be able to apply Einstein's equation to the Friedmann–Robertson–Walker (FRW) metric thereby relating the parameters in the metric to the density in the universe. Equation [1.3] is the result of applying Einstein's equation to the zero order universe. We will derive it in this section. Chapters 3 and 4 apply Einstein's equation to the perturbed universe. With the experience

we gain in this section, there will be nothing difficult about these subsequent chapters. The principles are identical; only the algebra will be a touch harder.

2.1.1 The Metric

Let us return to the grid depicted in figure 1.1. We said earlier that two grid points at rest move away other from each as the scale factor. So, if the comoving distance today is x_0, the physical distance between the two points at some earlier time t was $a(t)x_0$. (This simple expression for the physical distance is true only if the universe is flat, which I will assume in this section. See Problem (4) for a generalization to an open universe.) However, we know that any statement about distance is necessarily coordinate dependent. Even in special relativity, different observers measure different distances and times depending upon their velocities with respect to each other. In special relativity the proper distance $ds^2 = -dt^2 + d\vec{x}^2$ is coordinate-independent, so is identical in all frames. The presence of a gravitational field requires a generalization of the proper distance. More generally, it can be written as

$$ds^2 = \sum_{\mu,\nu=0}^{3} g_{\mu\nu} dx^\mu dx^\nu \tag{2.2}$$

where the indices μ and ν range from 0 to 3, with the first one reserved for the time- like coordinate ($dx^0 = dt$) and the last three for spatial coordinates. Here I have explicitly written down the summation sign, but from now on we will use the convention that repeated indices are summed over. The *metric*, $g_{\mu\nu}$, is necessarily symmetric, so in principle has four diagonal and six off-diagonal components. It provides a measure of proper distance. Flat space-time has a Minkowski metric: $g_{\mu\nu} = \eta_{\mu\nu}$, with

$$\eta_{\mu\nu} = \begin{pmatrix} -1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \end{pmatrix}. \tag{2.3}$$

What is the metric which describes the expanding universe? At least in a flat universe, the matric is almost identical to the Minkowski metric, except that distance must be multiplied by the scale factor. This suggests that the metric in an expanding, flat universe is

$$g_{\mu\nu} = \begin{pmatrix} -1 & 0 & 0 & 0 \\ 0 & a^2 & 0 & 0 \\ 0 & 0 & a^2 & 0 \\ 0 & 0 & 0 & a^2 \end{pmatrix}. \tag{2.4}$$

As noted in equation [1.3], which we will shortly derive, the evolution of the scale factor depends on the density in the universe. When perturbations are introduced, the metric will become more complicated, and the perturbed part of the metric will be determined by the inhomogeneities in the matter and radiation.

2.1.2 The Geodesic Equation

In Minkowski space, particles travel in straight lines, unless they are acted on by a force. Not surprisingly, the paths of particles in curved space-time are more complicated. The notion of a straight line gets generalized to a *geodesic*, the path followed by a particle in the absence of any forces. To express this in equations, we must generalize Newton's law with no forces, $d^2\vec{x}/dt^2 = 0$, to curved space-times. The generalization is the geodesic equation

$$\frac{d^2x^\mu}{d\lambda^2} = -\Gamma^\mu_{\alpha\beta}\frac{dx^\alpha}{d\lambda}\frac{dx^\beta}{d\lambda} \tag{2.5}$$

where λ parametrizes the particle's path as in figure 2.1. The Christoffel symbol, $\Gamma^\mu_{\alpha\beta}$, is

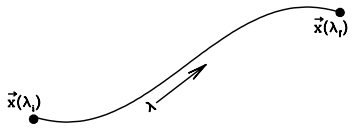


Figure 2.1: A particle's path is parametrized by λ, which monotonically increases from its initial value λ_i to its final value λ_f.

zero in flat space-time, so equation [2.5] reduces to Newton's equation, and particles travel in straight lines. In curved space-time, the Christoffel symbol depends on the metric. A convenient formula expressing this dependence is

$$\Gamma^\mu_{\alpha\beta} = \frac{g^{\mu\nu}}{2}\Big[\frac{\partial g_{\alpha\nu}}{\partial x^\beta} + \frac{\partial g_{\beta\nu}}{\partial x^\alpha} - \frac{\partial g_{\alpha\beta}}{\partial x^\nu}\Big]. \tag{2.6}$$

Note that the raised indices on $g^{\mu\nu}$ are important: $g^{\mu\nu}$ is different than $g_{\mu\nu}$. In general, indices are lowered and raised with the metric (e.g. $x_\mu = g_{\mu\nu}x^\nu$). Raising the indices on the metric itself leads to the relation $g^{\mu\nu} = g^{\mu\alpha}g^{\nu\beta}g_{\alpha\beta}$ which implies that $g^{\nu\beta}g_{\alpha\beta} = \delta^\nu_\alpha$, the Kronicker delta. So $g^{\mu\nu}$ for the flat, FRW metric is identical to $g_{\mu\nu}$ except that its spatial elements are $1/a^2$ instead of a^2.

Using the general expression in equation [2.6] and the FRW metric in equation [2.4], we can derive the Christoffel symbol in an expanding, homogeneous universe. First we compute the components with upper index equal to zero, $\Gamma^0_{\alpha\beta}$. Since the metric is diagonal, the factor of $g^{0\nu}$ vanishes unless $\nu = 0$ in which case it is -1. Therefore,

$$\Gamma^0_{\alpha\beta} = \frac{-1}{2}\Big[\frac{\partial g_{\alpha 0}}{\partial x^\beta} + \frac{\partial g_{\beta 0}}{\partial x^\alpha} - \frac{\partial g_{\alpha\beta}}{\partial x^0}\Big]. \tag{2.7}$$

The first two terms here reduce to derivatives of g_{00}. Since the FRW metric has constant g_{00}, these terms vanish, and we are left with

$$\Gamma^0_{\alpha\beta} = \frac{1}{2}\frac{\partial g_{\alpha\beta}}{\partial x^0}, \tag{2.8}$$

The derivative is non-zero only if α and β are spatial indices, which will be identified with Roman letters i, j running from one to three. Since $x^0 = t$, we have

$$\begin{aligned}\Gamma^0_{00} &= 0\\ \Gamma^0_{0i} &= \Gamma^0_{i0} = 0\\ \Gamma^0_{ij} &= \delta_{ij}\dot{a}a\end{aligned} \tag{2.9}$$

where overdots indicate derivatives with respect to time. It is a straightforward, but useful, exercise to show that $\Gamma^i_{\alpha\beta}$ is non-zero only when one of its lower indices is zero and one is spatial, so that

$$\Gamma^i_{0j} = \Gamma^i_{j0} = \delta_{ij}\frac{\dot{a}}{a} \tag{2.10}$$

with all other $\Gamma^i_{\alpha\beta}$ zero.

The Christoffel symbol allows us to examine the nature of geodesics. In particular we can use it to show that the energy of a massless particle scales as $1/a$. To apply the geodesic equation, we choose the parameter λ in equation [2.5] such that the four dimensional energy-momentum vector $P^\alpha = (E, \vec{P})$ is

$$P^\alpha = \frac{dx^\alpha}{d\lambda}. \tag{2.11}$$

This is an implicit definition of λ. Fortunately, one never needs to find λ explicitly for it can be directly eliminated by noting that

$$\begin{aligned}\frac{d}{d\lambda} &= \frac{dx^0}{d\lambda}\frac{d}{dx^0}\\ &= E\frac{d}{dt}.\end{aligned} \tag{2.12}$$

The zeroth component of the geodesic equation [2.5] then becomes

$$E\frac{dE}{dt} = -\Gamma^0_{ij}P^iP^j \tag{2.13}$$

where the equality holds since only the spatial components of $\Gamma^0_{\alpha\beta}$ are non-zero. Inserting these components leads to a right hand side equal to $-\delta_{ij}a\dot{a}P^iP^j$. A massless particle has energy-momentum* vector $(E, \vec{P})$ with zero magnitude:

$$-E^2 + \delta_{ij}a^2P^iP^j = 0 \tag{2.14}$$

which enables us to write the right hand side of equation [2.13] as $-(\dot{a}/a)E^2$. Therefore, the geodesic equation yields

$$\frac{dE}{dt} + \frac{\dot{a}}{a}E = 0, \tag{2.15}$$

*Note that $\vec{P}$ measures motion on the comoving (non-expanding) grid. The physical momentum which measures changes in physical distance is related to $\vec{P}$ by a factor of a. Hence, the factor of a^2 in equation [2.14].

the solution to which is

$$E \propto \frac{1}{a}. \tag{2.16}$$

This confirms our hand-waving argument earlier on that the energy of a massless particle should decrease as the universe expands since it is inversely proportional to its wavelength, which is being stretched along with the expansion. In the next chapter we will rederive this result in yet another, albeit related, way using the Boltzmann equation.

2.1.3 Einstein's Equation

If you did a word search on the previous two subsections, you might be surprised to discover that the words "general relativity" are not mentioned. The concept of a metric and the realization that non-trivial metrics affect geodesics both exist completely independently of general relativity. The main content of general relativity is the relation between the metric and energy. This relation is described by one equation, Einstein's equation, which reads

$$G_{\mu\nu} \equiv R_{\mu\nu} - \frac{1}{2} g_{\mu\nu} R = 8\pi G T_{\mu\nu} \tag{2.17}$$

where $R_{\mu\nu}$ is the *Ricci tensor* which depends on the metric and its derivatives; R, the *Ricci scalar*, is the contraction of the Ricci tensor ($R \equiv g^{\mu\nu} R_{\mu\nu}$); G is Newton's constant; and $T_{\mu\nu}$ is the energy-momentum tensor. The left hand side is a function of the metric, the right a function of the energy: Einstein's equation relates the two.

The Ricci tensor is most conveniently expressed in terms of the Christoffel symbol,

$$R_{\mu\nu} = \Gamma^{\alpha}_{\mu\nu,\alpha} - \Gamma^{\alpha}_{\mu\alpha,\nu} + \Gamma^{\alpha}_{\beta\alpha}\Gamma^{\beta}_{\mu\nu} - \Gamma^{\alpha}_{\beta\nu}\Gamma^{\beta}_{\mu\alpha}. \tag{2.18}$$

Here commas denote derivatives with respect to x. So, for example, $\Gamma^{\alpha}_{\mu\nu,\alpha} \equiv \partial\Gamma^{\alpha}_{\mu\nu}/\partial x^{\alpha}$. While this expression looks formidable, we have already done the hard work by computing the Christoffel symbol in an FRW universe. (Of course the real hard work was done eighty years ago by Einstein.) It turns out that there are only two sets of non-vanishing components of the Ricci tensor: one with $\mu = \nu = 0$ and the other with $\mu = \nu = i$.

Consider

$$R_{00} = \Gamma^{\alpha}_{00,\alpha} - \Gamma^{\alpha}_{0\alpha,0} + \Gamma^{\alpha}_{\beta\alpha}\Gamma^{\beta}_{00} - \Gamma^{\alpha}_{\beta 0}\Gamma^{\beta}_{0\alpha}. \tag{2.19}$$

Recall that the Christoffel symbol vanishes if its two lower indices are zero, so the first and third terms on the right vanish. Similarly, the indices α and β in the second and fourth terms must be spatial. We are left with

$$R_{00} = -\Gamma^{i}_{0i,0} - \Gamma^{i}_{j0}\Gamma^{j}_{0i}. \tag{2.20}$$

Using equation [2.10] leads directly to

$$\begin{aligned} R_{00} &= -\delta_{ii}\frac{\partial}{\partial t}\left(\frac{\dot{a}}{a}\right) - \left(\frac{\dot{a}}{a}\right)^2 \delta_{ij}\delta_{ij} \\ &= -3\frac{\ddot{a}}{a}. \end{aligned} \tag{2.21}$$

The factor of three arises here since δ_{ii} means sum over all three spatial indices, counting one for each. I will leave the space-space component as an exercise; it is

$$R_{ij} = \delta_{ij}\left[2\dot{a}^2 + a\ddot{a}\right]. \tag{2.22}$$

The next ingredient in the Einstein equation is the Ricci scalar, which we can now compute since

$$\begin{aligned} R &\equiv g^{\mu\nu} R_{\mu\nu} \\ &= -R_{00} + \frac{1}{a^2} R_{ii}. \end{aligned} \tag{2.23}$$

Again the sum over i leads to a factor of three, so

$$R = 6\left[\frac{\ddot{a}}{a} + \left(\frac{\dot{a}}{a}\right)^2\right]. \tag{2.24}$$

To write down Einstein's equation in a homogeneous universe, we need consider only the time-time component:

$$R_{00} - \frac{1}{2} g_{00} R = 8\pi G T_{00}. \tag{2.25}$$

The time-time component of the energy-momentum tensor is simply the energy density ρ. So we finally have

$$\left(\frac{\dot{a}}{a}\right)^2 = \frac{8\pi G}{3}\rho. \tag{2.26}$$

To get this into the form of equation [1.3], recall that the left hand side here is the square of the Hubble rate and that the critical density was defined as $\rho_{\rm cr} \equiv 3H_0^2/8\pi G$. So, dividing both sides by H_0^2 leads to

$$\frac{H^2(t)}{H_0^2} = \Omega. \tag{2.27}$$

Here $\Omega \equiv \rho/\rho_{\rm cr}$ counts the energy density from all species: matter, radiation, and the vacuum. Putting in the scaling of each of these leads directly to equation [1.3] except for the *curvature* term there, the one which scales as a^{-2}. In our derivation, we have assumed the universe is flat, so this term vanishes. I leave it as an exercise to derive the Einstein equation in an open universe.

2.2 Distances

We can anticipate that measuring distance in an expanding universe will be a tricky business. Referring back to the expanding grid of figure 1.1, we immediately see two possible ways to measure distance, the comoving distance which remains fixed as the universe expands or the physical distance which grows simply because of the expansion. Frequently, neither of these two measures accurately describes the process of interest. For example light leaving a distant QSO at redshift three starts its journey towards us when the scale factor was only a quarter of its present value and ends it today when the universe has expanded by a factor

of four. Which distance do we use in that case to relate say the luminosity of the QSO to the flux we see?

The fundamental distance measure, from which all others may be calculated, is the distance on the comoving grid. If the universe is flat, as we will assume through most of this book, then computing distances on the comoving grid is easy: the distance between two points $\vec{x}_1$ and $\vec{x}_2$ is equal to $[(x_1-x_2)^2+(y_1-y_2)^2+(z_1-z_2)^2]^{1/2}$.

One very important comoving distance is the distance light could have traveled (in the absence of interactions) since $t=0$. In a time dt, light travels a comoving distance $dx = dt/a$, so the total comoving distance light could have traveled is

$$\eta \equiv \int_0^t \frac{dt'}{a(t')}. \tag{2.28}$$

The reason this distance is so important is that no information could have propagated further (again on the comoving grid) than η since the beginning of time. Therefore, regions separated by distance greater than η are not causally connected. If they appear similar, we should be suspicious! We can think of η then as the *comoving horizon*. Since we are setting $c=1$, we can also think of η, which is monotonically increasing, as a time-variable and call it the *conformal time*. Just like the time t, the temperature T, the redshift z, and the scale factor a, η can be used to discuss the evolution of the universe. In fact, for most purposes η is the most convenient time variable, so when we begin to study the evolution of perturbations, we will use it instead of t. For a universe with just matter and radiation (no cosmological constant), the conformal time can be calculated analytically; you will show (Problem (6)) that

$$\frac{\eta}{\eta_0} = \sqrt{a+a_{\rm EQ}} - \sqrt{a_{\rm EQ}}. \tag{2.29}$$

Although equation [2.29] is only true in a matter+radiation universe and therefore may not be accurate at recent times, it is almost always true early on ($z>10$ say). The limits – $\eta \propto a$ in a radiation dominated universe and $\eta \propto a^{1/2}$ in a matter dominated era – are good to remember.

Another important comoving distance is that between a distant emitter and us. In that case, the comoving distance out to an object at scale factor a (or redshift $z=1/a-1$) is

$$d_C(a) = \int_a^1 \frac{da'}{a'^2 H(a')}. \tag{2.30}$$

Here I have changed the integration over t' in equation [2.28] to one over a', which brings in the additional factor of $da/dt = aH$ in the denominator. Typically we can see objects out to $z <= 5$; at these late times radiation can be ignored (recall figure 1.2). If the universe is purely matter dominated at these late times, then $H \propto a^{-3/2}$ and we can do the integral in equation [2.30] analytically,

$$\begin{aligned} d_C^{\rm flat,MD}(a) &= \frac{2}{H_0}\left[1-a^{1/2}\right] \\ &= \frac{2}{H_0}\left[1-\frac{1}{\sqrt{1+z}}\right]. \end{aligned} \tag{2.31}$$

This comoving distance goes as z/H_0 for small z (verifying our hand waving discussion of the low-z Hubble diagram in §1.3) and then asymptotes to $2/H_0$ as z gets very large.

A classic way to measure distances in astronomy is to measure the angle θ subtended by an object of known physical size d. The distance to that object (assuming the angle subtended is small) is then

$$d_A = \frac{d}{\theta}. \tag{2.32}$$

The subscript $_A$ here denotes *angular diameter distance*. To compute the angular diameter distance in an expanding universe, we first note that the comoving size of the object is d/a. The comoving distance out to the object is given by equation [2.30], so the angle subtended is $\theta = (d/a)/d_C(a)$. Comparing with equation [2.32], we see that the angular diameter distance is

$$d_A^{\rm flat} = a d_C = \frac{d_C}{1+z}. \tag{2.33}$$

The superscript here is a warning that this result holds only in a flat universe. See Problem (7) for a generalization to an open universe. Note that the angular diameter distance is equal to the comoving distance at low redshift, but actually decreases at very large redshift. At least in a flat, matter dominated universe, objects at large redshift appear larger than they would at intermediate redshift!

Another way of measuring distances in astronomy is to measure the flux from an object of known luminosity. Recall that (forgetting about expansion for the moment), the observed flux F a distance d from a source of known luminosity L is

$$F = \frac{L}{4\pi d^2} \tag{2.34}$$

since the total luminosity through a spherical shell with area $4\pi d^2$ is constant. How does this result generalize to an expanding universe? Again it is simplest to work on the comoving grid, this time with the source centered at the origin. The flux we observe is

$$F = \frac{L(d_C)}{4\pi d_C^2(a)} \tag{2.35}$$

where $L(d_C)$ is the luminosity through a (comoving) spherical shell with radius $d_C(a)$. To further simplify, let's assume that the photons are all emitted with the same energy. Then $L(d_C)$ is this energy multplied by the number of photons passing through a spherical (comoving) shell per second. In one second, photons travel farther on the comoving grid at early times than at late times since the associated physical distance at early times is smaller. Therefore, the number of photons crossing a shell in one second will be smaller today than at emission, smaller by a factor of a. Similarly, the energy of of the photons will be smaller today than at emission due to expansion. Therefore, the energy per second passing through a comoving shell a distance $d_C(a)$ (i.e. our distance) from the source will be a factor of a^2 smaller than the luminosity at the source. The flux we observe therefore will be

$$F = \frac{La^2}{4\pi d_C^2(a)} \tag{2.36}$$

where L is the luminosity ar the source. We can keep† equation [2.34] in an expanding universe as long as we define the *luminosity distance*

$$d_L \equiv \frac{d_C}{a}. \tag{2.37}$$

Figure 2.2 shows the three different distances.

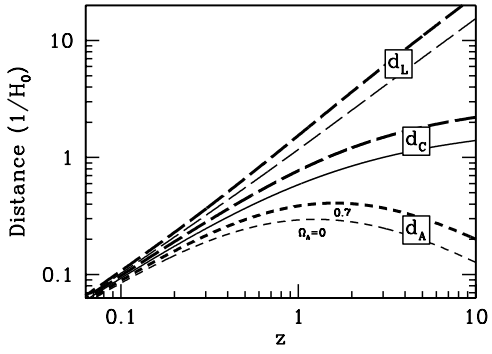


Figure 2.2: Three distance measures in a flat expanding universe. From top to bottom, the luminosity distance, the comoving distance, and the angular diameter distance. The pair of lines in each case is for a flat universe with matter only (light curves) and seventy percent cosmological constant (heavy curves). In a Lambda dominated universe, distances out to fixed redshift are larger than in a mtter dominated universe.

The luminosity distance of equation [2.37] can be used to find the apparent magnitude m of a source with absolute magnitde M. Magnitudes are related to fluxes via $m = -(5/2)\log(F)+$ constant. Since the flux scales as d_L^{-2}, the apparent magnitude $m = M + 5\log(d_L)+$ constant. The convention is that

$$m - M = 5\log\left(\frac{d_L}{10\text{pc}}\right) + K \tag{2.38}$$

where K is a correction for the shifting of the spectrum into or out of the wavelength range measured due to expansion. Using equation [2.38] you can reproduce the theory curves on the famous Supernova plot in figure 1.7.

†Actually there is one more difference that needs to be accounted for: the observed luminosity is related to the emitted luminosity at a different wavelength. Here we have assumed a detector which counts all the photons.

2.3 Epoch of Matter-Radiation Equality

The epoch at which the energy density in matter equals that in radiation is called *matter-radiation equality*. It has a special significance for the generation of large scale structure and for the development of CMB anisotropies, because perturbations grow at different rates in the two different eras.

It is therefore a useful exercise to calculate the epoch of matter- radiation equality. To do this, we need to compute the energy density of both matter and radiation, and then find the value of the scale factor at which they were equal. Let us start by calculating the energy density of the radiation. As far as we know, there are two components of this radiation: the microwave photons and the neutrinos. First the photons. It is convenient to introduce the occupation number, or *distribution function*, of the photon gas. This counts the number of photons in a given region in phase space around position $\vec{x}$ and momentum $\vec{p}$. In the "zero order" universe, the photon distribution function $f = [e^p/T - 1]^{-1}$. This zero order distribution is homogeneous (does not depend on $\vec{x}$) and anisotropic (does not depend on $\hat{p}$). The energy density of photons, ρ_γ, is the sum of f over all momenta. In the infinite volume limit this sum becomes an integral and

$$\rho_\gamma = 2\int \frac{d^3p}{(2\pi)^3}\,\frac{1}{e^{p/T}-1}\,p. \tag{2.39}$$

The factor of two in front of equation 2.39 accounts for the two spin states of the photon. The integral is over momentum $\vec{p}$, and the energy of a given state is simply equal to p since the photon is massless. Since there is no angular dependence in the integrand of equation 2.39, the angular integral yields a factor of 4π and we are left with a one dimensional integral. Define a dummy variable $x \equiv p/T$; then

$$\rho_\gamma = \frac{8\pi T^4}{(2\pi)^3}\int_0^\infty \frac{dx\; x^3}{e^x - 1}. \tag{2.40}$$

The integral can be expressed in terms of the Riemann ζ function; it is $6\zeta(4) = \pi^4/15$, so

$$\rho_\gamma = \frac{\pi^2}{15}T^4. \tag{2.41}$$

It will be useful to have all energy densities in the same units. The simplest way to do this is to divide all energy densities by the critical density today‡. Thus,

$$\begin{aligned}\frac{\rho_\gamma}{\rho_{\rm cr}} &= \frac{\pi^2}{15}\left(\frac{2.728^\circ\text{K}}{a}\right)^4 \frac{1}{8.10\times10^{-11}\text{eV}^4} \\ &= \frac{2.48\times10^{-5}}{h^2a^4}\end{aligned} \tag{2.42}$$

where to get the last line, it is useful to remember the conversion between degrees Kelvin and eV: 11605° K $= 1$ eV. According to equation 2.42 the photon energy density depends

‡The critical density – just as the Hubble rate which defines it – changes with time. However, it is common to define $\rho_{\rm cr}$ to be a constant, the critical density today, and I will follow this convention.

on time via the scale factor, but has no spatial dependence. This is because we have used the zero order distribution function, the Planck function, for the photons. In fact there are small perturbations around this zero order distribution function. These do have a spatial dependence and correspond to the anisotropies in the CMB.

Returning to our calculation, the next component we need consider is the neutrino. The neutrino is a massless particle so its density is comparable to that of the photon. However, there are a number of differences between the two, some fundamental and one cosmological:

- One spin degree of freedom for neutrinos
- Neutrino has anti-particle
- Three generations of neutrinos
- Neutrinos are fermions $\rightarrow$ Fermi-Dirac distribution function
- Neutrino temperature is lower since photons are heated by e^+e^- annihilation

The first two factors cancel: the photons have twice as much density since there are two spin states but half as much since they do not have an anti-particle. Three generations of neutrinos means the neutrino density is three times larger than the photon density. The fourth difference means we need to change the denominator in the integrand in equation 2.39 to $e^{p/T}+1$. The Fermi-Dirac integral is then smaller by a factor of 7/8. The final effect is that the neutrino temperature is smaller than the photon temperature by the factor $(4/11)^{1/3}$. This translates to an energy density smaller by a factor $(4/11)^{4/3}$. Putting all these factors together leads to

$$\rho_\nu = 3\,\frac{7}{8}\,\left(\frac{4}{11}\right)^{4/3}\,\rho_\gamma. \tag{2.43}$$

Incidentally, the warning about the photon energy density being spatially independent only at zero order holds here as well. Although it is unlikely we will ever directly detect anisotropies in the cosmic neutrino background, these anisotropies are important because they affect the photon anisotropies. In fact it has been argued that future observations of CMB anisotropy will be so precise that they will be sensitive to the neutrino anisotropy. So we will need to worry about these later on as well. In any event, using equation 2.42, we see that the total energy density in radiation is

$$\frac{\rho_{\rm R}}{\rho_{\rm cr}} = \frac{4.17\times10^{-5}}{h^2a^4} = \frac{\Omega_R}{a^4}. \tag{2.44}$$

Note that Ω_R (and all other Ω's in the book) is the ratio of energy densities today.

We now need to calculate the energy density in matter. This turns out to be no calculation at all because of our ignorance. The abundances of photons and neutrinos are fully specified by their temperatures which we know very accurately. The abundance of protons, say, is much less certain since it depends not only on the protons' temperature but also on the primordial excess of baryons over anti-baryons. Neither of these is known very well. The abundance of a hypothetical cold dark matter candidate is even more uncertain. Since we

have nothing in terms of which to express the energy density, it is simplest to keep Ω_m as our free parameter. The one thing we do know is the scaling, namely that

$$\frac{\rho_{\rm M}}{\rho_{\rm cr}} = \frac{\Omega_m}{a^3(t)}. \tag{2.45}$$

We can now calculate the epoch of matter-radiation equality. We equate equations 2.44 and 2.45 to find

$$a_{\rm EQ} = \frac{4.17\times10^{-5}}{\Omega_m h^2}. \tag{2.46}$$

A different way to express this epoch is in terms of redshift z; the redshift of equality is

$$1+z_{\rm EQ} = 2.4\times10^4\Omega_m h^2. \tag{2.47}$$

Note that – obviously – as the amount of matter in the universe today, $\Omega_m h^2$, goes up, the redshift of equality also goes up. For our purposes it will be very important to remember that typically $z_{\rm EQ}$ is of order 10^4 while the redshift when photons decouple from matter, which we will denote z_*, is of order 10^3. Thus, we expect photons to decouple when the universe is already well into the matter dominated era.

2.4 Decoupling

2.4.1 General Considerations

I have already alluded several times to the important fact that photons decoupled from electrons at a redshift of order 10^3. Since this is such an important time for CMB anisotropies, we now calculate it. This calculation will also introduce some of the Boltzmann-ology that will segue us into the next chapter.

Decoupling occurs roughly when the rate for photons to Compton scatter off electrons becomes smaller than the expansion rate. The scattering rate is $n_e\sigma_T$ where n_e is the *free* (i.e. ionized) electron density and σ_T is the Thomson cross section, equal to $0.665\times10^{-24}{\rm cm}^2$. An electron can either be free, bound into hydrogen, or bound into helium. To simplify our discussion, we will ignore helium for the time being. Then the total number of electrons is just equal to the number of hydrogen nuclei, which is simply the baryon density $n_{\rm B}$ in this no-helium limit. Define the ratio of ionized electrons to the total electrons to be x_e. Then,

$$n_e\sigma_T = x_e n_{\rm B}\sigma_T. \tag{2.48}$$

Since the ratio of the baryon density to the critical density is $m_{\rm proton}n_{\rm B}/\rho_{\rm cr} = \Omega_B/a^3$, n_B can be eliminated in equation [2.48] in favor of the parameter Ω_B:

$$n_e\sigma_T = 7.477\times10^{-30}{\rm cm}^{-1}\frac{x_e\Omega_B h^2}{a^3}. \tag{2.49}$$

Decoupling occurs when $n_e\sigma_T/H$ is less than one. Using equations [1.6] and [2.49], we find that this ratio is

$$\frac{n_e\sigma_T}{H} = 0.0692\frac{x_e\Omega_B h}{a^3(H/H_0)}. \tag{2.50}$$

The ratio on the right depends on the time dependence of the Hubble rate, which is given in equation 1.3. From that equation or from figure 1.2, we see that at early times, the main contribution comes from either matter or radiation. Any vacuum energy or curvature term is negligible at early times. Recalling our calculation from the previous section, we expect matter to be the dominant component at redshift 1000, so we can safely set $H/H_0 \simeq \Omega_m^{1/2}/a^{3/2}$. Then,

$$\frac{n_e \sigma_T}{H} \simeq .0692 \frac{\Omega_B h}{\Omega_m^{1/2}} \frac{x_e}{a^{3/2}}. \tag{2.51}$$

In order to go further we need to find the ionization history of the electrons, x_e. However, already at this stage we can say something interesting about decoupling.

2.4.2 Eternally Ionized Universe

If the universe remains ionized for all times, $x_e = 1$, then equation 2.51 can be trivially solved to find the redshift of decoupling. Setting the right hand side to one leads to

$$1 + z_* = 43 \left(\frac{0.02}{\Omega_B h^2}\right)^{2/3} \left(\frac{\Omega_m h^2}{0.15}\right)^{1/3} \qquad \text{(No Recombination)}. \tag{2.52}$$

where I have normalized the baryon density $\Omega_B h^2$ with a best guess from nucleosynthesis arguments and the matter density $\Omega_m h^2$ with another currently popular number. Equation [2.52] tells us that even if the electrons remained ionized throughout the history of the universe, eventually the photons decoupled simply because expansion made it more difficult to find the increasingly dilute electrons. In theory, we do not expect the electrons to remain ionized throughout, and we will shortly calculate what we do expect for their history. However, equation 2.52 is relevant for a more general reason. We do expect that at some late time, the electrons were re-ionized[§]. We expect this because the universe we observe back to redshift $z \sim 5$ appears to be ionized. If the universe was reionized at very late times, much after the z_* of equation [2.52], there would not be a huge change in the CMB anisotropy pattern. However, if the universe was reionized earlier than this redshift, multiple scattering of the photons would dramatically alter the primordial anisotropy pattern set up at $z \sim 1000$.

2.4.3 Standard Ionization History

We now consider a more realistic ionization history. At temperatures well above the ionization energy of hydrogen, $\epsilon_0 = 13.6$ eV, any time an electron combined with a proton to produce a neutral hydrogen atom, a photon from the cosmic background immediately dissociated the atom. As the temperature dropped below ϵ_0, this still held true because there were so many more photons than electrons: even the small number of photons in the high energy tail of the distribution was enough to keep the universe ionized. Finally, as the temperature dropped to roughly 1/3 eV (corresponding to $z \sim 1000$), there were too few

[§]Reionization means the electrons become ionized again, just as they were at very early times. No one has ever been able to figure out what *re*-combination refers to.

photons to maintain full ionization. Soon after this, almost all of the electrons and protons were swept up into neutral hydrogen atoms.

To quantify this history, it is convenient to use the Saha approximation, which assumes that the collision rate for $ep \to H\gamma$ is the same as its inverse $H\gamma \to ep$. In reality, this equality, a manifestation of equilibrium, held until just below 1/3 eV. So for most of the time of interest, the Saha approximation is a reasonable one. Instead of simply writing down the Saha equation, though, I want to derive it using some Boltzmann tools. These will come in handy in the next chapter when we consider Compton scattering.

The rate for any process is obtained by summing over the momenta of all particles, weighting by the number of particles in the initial state and an amplitude, $|\mathcal{M}|^2$, for the transition. In principle, one can calculate the amplitude using quantum mechanics. Here we will not need to know anything about this amplitude except that it is symmetric: it is the same if the initial and final states are reversed. Equating the rates for recombination and ionization:

$$n_p \sum_{\vec{p},\vec{q}} \frac{f_e(q)}{E_\gamma(p)E_e(q)} \delta_{p-\epsilon_0, q^2/2m_e} |\mathcal{M}|^2 = n_H \sum_{\vec{p},\vec{q}} \frac{f_\gamma(p)}{E_\gamma(p)E_e(q)} \delta_{p-\epsilon_0, q^2/2m_e} |\mathcal{M}|^2 \tag{2.53}$$

where f_e and f_γ are the electron and photon distribution functions respectively; n_H and n_p are the number densities of neutral hydrogen and free protons, both of which are so heavy that their momentum is irrelevant, and the Dirac delta function enforces energy conservation. To emphasize the similarity between the the way photons and electrons are treated, I have written the denominators[¶] as E_e and E_γ, but we will soon replace these by m_e and p respectively, since electrons are very non-relativistic at decoupling and photons are massless.

The functions f_e and f_γ can be set to a Maxwell-Boltzmann and a Planck distribution respectively. Thus,

$$f_e(q) = n_e \left(\frac{2\pi}{m_e T}\right)^{3/2} e^{-q^2/2m_e T} \tag{2.54}$$

where the prefactor is set by requiring that the integral over d^3q gives n_e, the electron density[‖]. Although the photon distribution function is the Bose-Einstein one used in equation 2.39 we can make use of the fact that only the highest energy photons (those with $p/T >> 1$) contribute to the ionization. For these photons the distribution becomes:

$$\lim_{p/T >> 1} f_\gamma(p) = e^{-p/T}. \tag{2.55}$$

So the Saha condition is

$$n_p n_e \left(\frac{2\pi}{m_e T}\right)^{3/2} \sum_{\vec{p},\vec{q}} \frac{e^{-q^2/2m_e T}}{p} \delta_{p-\epsilon_0, q^2/2m_e} |\mathcal{M}|^2 = n_H \sum_{\vec{p},\vec{q}} \frac{e^{-p/T}}{p} \delta_{p-\epsilon_0, q^2/2m_e} |\mathcal{M}|^2 \tag{2.56}$$

[¶]The factors of energy in the denominator arise because one initially sums over the full four component energy-momentum vector $(q_0, \vec{q})$. The sum (or integral) over q_0 is performed with a delta function, $\delta(q_0^2 - q^2 - m^2)$, and results in the derivative of the delta function in the denominator.

[‖]The Maxwell-Boltzmann distribution in equation [2.54] is the non-relativistic limit of the full Fermi-Dirac distribution: $f_e = [e^{(E-\mu)/T} + 1]^{-1}$ with the energy $E \simeq m + q^2/2m_e$ and the chemical potential given by

$$\mu = m + T \ln \left[n_e (2\pi/m_e T)^{3/2}\right]$$

If we needed to, we could perform the sums in equation [2.56], but there is no need. Simply replace $q^2/2m_e$ in the exponential on the left with $p - \epsilon_0$, since the delta function enforces this equality. Then, we immediately see that the sums are identical apart from a factor of $e^{\epsilon_0/T}$ on the left, which can be taken outside. Therefore, equating the rates leads to

$$n_e n_p e^{\epsilon_0/T} \left(\frac{2\pi}{m_e T}\right)^{3/2} = n_H, \tag{2.57}$$

the Saha equation.

We have defined $x_e \equiv n_e/(n_e + n_H)$, so $n_e/n_H = x_e/(1 - x_e)$. The free proton density is of course equal to the free electron density: $n_p = x_e n_{\rm B}$, so

$$\frac{x_e^2}{1 - x_e} = \frac{e^{-\epsilon_0/T}}{n_{\rm B}} \left(\frac{m_e T}{2\pi}\right)^{3/2}. \tag{2.58}$$

This history is plotted in figure 2.3. At high temperatures, x_e is very close to one. As the temperature drops, the exponential becomes very small and x_e falls dramatically. After this, the denominator on the left side of equation [2.58] can be set to one, so at late times

$$x_e \to \frac{e^{-\epsilon_0/2T}}{n_{\rm B}^{1/2}} \left(\frac{m_e T}{2\pi}\right)^{3/4}. \tag{2.59}$$

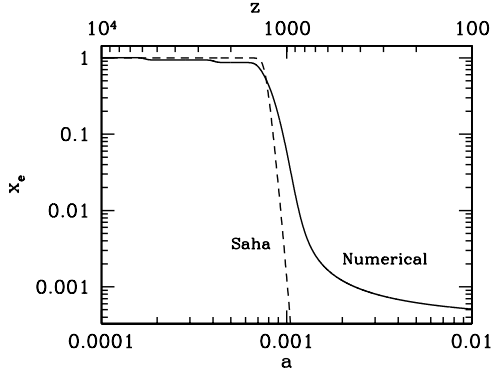


Figure 2.3: Ionization history of the universe. Plotted is the fraction of electrons which remain free as a function of redshift for both the Saha approximation and the full numerical solution.

Using the analytic expression in equation [2.59], we can find a good approximation for the epoch of decoupling. Recall that decoupling occurs when $n_e \sigma_T / H$ falls below unity. Using

equations (2.51) and (2.59), we see that the Saha approximation tells us that for standard recombination

$$\begin{aligned} \frac{n_e \sigma_T}{H} &\simeq .069\, \frac{\Omega_B h}{\Omega_m^{1/2}} \left(\frac{m_e T_0}{2\pi n_{B,0}^{2/3}}\right)^{3/4} \frac{e^{-a\epsilon_0/2T_0}}{a^{3/4}} \\ &= 6.8 \times 10^{13} \left(\frac{\Omega_B}{\Omega_m}\right)^{1/2} \frac{\exp\{-2.89 \times 10^4 a\}}{a^{3/4}}. \end{aligned} \tag{2.60}$$

Requiring this ratio to be equal to one leads to an expression for the epoch of decoupling:

$$a_* \simeq 1.3 \times 10^{-3} \left[1 + \frac{\ln\left(\left(\frac{\Omega_B}{0.2\Omega_m}\right)^{1/2} (1000 a_*)^{-3/4}\right)}{36}\right]. \tag{2.61}$$

The logarithmic correction is negligible for all reasonable models, so we find an expected decoupling at $z_* \simeq 800$, in rough agreement with more detailed numerical work.

The Saha approximation has several shortcomings as applied here. If the universe stayed in equilibrium throughout, as Saha assumes, x_e would be exponentially small today. In fact, these reactions fall out of equilibrium and x_e freezes out to some constant value at late times. Second, we have assumed that hydrogen is a one level atom and we have not included the (non-equilibrium) distribution of photons that emerges from recombination itself. That is, every time an electron and proton combine, they emit a photon with energy slightly above ϵ_0. These non-equilibrium photons can easily ionize any H atom in the ground state. Thus recombination directly to the ground state is not effective. The results of a more correct treatment are shown in figure 2.3. A final feature of the full treatment is that helium recombination rids the universe of two free electrons for every helium atom at epochs well before hydrogen recombination. Note the dip in x_e at $z \simeq 2500$ in the full numerical solution.

A final comment on the accuracy of our estimate for the epoch of decoupling. We will see in Chapter 7 that decoupling really takes place at $z \simeq 1100$. The main reason for the discrepency is that here we simply take the ratio of the scattering rate to the Hubble rate. In Chapter 7 we will be a bit more sophisticated and define the *visibility function* which is the probability a photon last scattered at redshift z. This turns out to be sharply peaked at $z \simeq 1100$.

Suggested Reading

My favorite book on general relativity at this level is *A First Course in General Relativity* (Schutz), which gives many simple examples to introduce the seemingly profound ideas of general relativity. Slightly more advanced is *Gravitation and Cosmology : Principles and Applications of the General Theory of Relativity* (Weinberg), which also has a nice discussion of the early universe in Chapter 15. Two more advanced books are *General Relativity* (Wald) and the classic *Gravitation* (Misner, Thorne and Wheeler). Some of the Boltzmann formalism introduced in this, and especially the next, chapter is presented in *Kinetic Theory in the Expanding Universe* (Bernstein). The distance formulae of §2.2 are covered in all standard texts.

A number of papers treat the topics in this chapter at an accessible level. An especially coherent review of all the different distance measures is given by Hogg (1999). As I tried to indicate in the text, the process of recombination is very rich; it involves some subtle physics. The original paper which worked through all the details was by Peebles (1968). Ma and Bertschinger (1996) however managed to describe the physics succintly in just one page in their §5.8. Recently Seager, Sasselov, and Scott (1999) presented a more accurate treatment (although, as they emphasize, Peebles' more intuitive work holds up remarkably well), including many small effects previously neglected.

Problems

2.1 Convert the following quantities by inserting the appropriate factors of $c, \hbar$, and k_B:

- $T_0 = 2.728°\mathrm{K} \to \mathrm{eV}$
- $\rho_\gamma = \pi^2 T^4/15 \to \mathrm{eV}$ and $\mathrm{g\ cm^{-3}}$
- $1/H_0 \to \mathrm{cm}$
- $m_{\mathrm{Pl}} \equiv 1.2 \times 10^{19}\ \mathrm{GeV} \to °\mathrm{K}$, $\mathrm{cm^{-1}}$, $\mathrm{sec^{-1}}$

2.2 Fill in some of the blanks left in our derivation of Einstein's equation.

(a) Compute the Christoffel symbol $\Gamma^i_{\alpha\beta}$ for a flat FRW metric.

(b) Compute the spatial components of the Ricci tensor in a flat FRW universe, R_{ij}. Show that the space-time component, R_{i0}, vanishes.

2.3 Show that the space-space component of Einstein's equation in a flat universe is

$$\frac{d^2a/dt^2}{a} + \frac{1}{2}\left(\frac{da/dt}{a}\right)^2 = -4\pi G P \tag{2.62}$$

where P is the pressure, the T^i_i (no sum over i) component of the stress-energy tensor.

2.4 Derive Einstein's equation in the case of an open universe. The proper distance in an open universe is

$$ds^2 = -dt^2 + a^2(t)\left\{d\eta^2 + \frac{\sinh^2(\Omega_k^{1/2}H_0\eta)}{\Omega_k H_0^2}(d\theta^2 + \sin^2\theta d\phi^2)\right\} \tag{2.63}$$

where η, θ, ϕ are the standard 3D spherical coordinates, and Ω_k the curvature density is equal to $1 - \Omega_m - \Omega_\Lambda$.

2.5 Show that the geodesic equation we derived in a flat universe implies that

$$\frac{d^2\vec{x}}{d\eta^2} = 0 \tag{2.64}$$

where η is the conformal time.

2.6 Consider a universe with only matter and radiation, with equality at a_{eq}. Find an analytic expression for η as a function of a. What is the conformal time today? At decoupling?

2.7 Using the metric in Problem (4), find the angular diameter distance in an open universe with curvature density, Ω_k. Show that the comoving size of the object with physical size d is no longer equal to d/a. Incorporate this change to show that the angular diameter distance in an open universe is equal to

$$d_A^{\mathrm{open}} = \frac{a}{H_0\sqrt{\Omega_k}}\sinh\left(\sqrt{\Omega_k}H_0\eta\right). \tag{2.65}$$

2.8 Plot $m - M$ for a flat, matter dominated universe (this can be analytically) and for a flat universe with $\Lambda = 0.7, \Omega_m = 0.3$ (for this you need to evaluate numerically a 1D integral). Neglect the $K-$ correction. Compare with figure 1.7.

2.9 Show that the number density of one generation of neutrinos and anti-neutrinos in the universe today is

$$n_\nu = \frac{3}{11}n_\gamma = 112\mathrm{cm}^{-3}.$$

2.10 An important parameter for CMB anisotropies is the sound speed at decoupling. This is determined by the ratio of baryons to photons.

(a) Find

$$R \equiv \frac{3\rho_B}{4\rho_\gamma}$$

as a function of a. Evaluate it at decoupling. Your answer should depend on $\Omega_b h^2$.

(b) We will see in Chapter 7 that the sound speed of the combined photon/baryon fluid is

$$c_s = \sqrt{\frac{1}{3(1+R)}}.$$

Use your answer from (a) to compute the sound speed at decoupling.

Chapter 3

The Boltzmann Equations

We are interested in the anisotropies in the cosmic distribution of photons and inhomogeneities in the matter. Figure 3.1 shows why these are complicated to calculate. The photons are affected by gravity and by Compton scattering with free electrons. The electrons are tightly coupled to the protons. Both of these of course are also affected by gravity. The metric which determines the gravitational forces is influenced by all these components plus the neutrinos and the dark matter. Thus to solve for the photon and dark matter distributions, we need to simultaneously solve for all the other components.

There is a systematic way to account for all of these couplings. We write down a Boltzmann equation for each species in the universe. Schematically, the Boltzmann equation is

$$\frac{df}{dt} = C[f]. \tag{3.1}$$

The right hand side of the Boltzmann equation contains all possible collision terms. These terms in general are complicated functionals of the distribution functions of the various components. In the absence of collisions, the distribution function obeys $df/dt = 0$. This seemingly innocent equation says that the number of particles in a given element of phase space does not change with time. The catch is that the phase space elements themselves are moving in time in complicated ways due to the non-trivial metric. This catch makes the problem more difficult than it seems from equation 3.1. Nonetheless, we can still progress systematically by re-expressing the full derivative in terms of partial derivatives.

In this Chapter, we derive the Boltzmann equations for photons, eletrons, protons, dark matter, and massless neutrinos. This set of equations governs the evolution of perturbations in the universe.

3.1 Boltzmann Equation for the Harmonic Oscillator

Before tackling the problem of interest – the Boltzmann equation for all species in an expanding universe – let us treat a much simpler example of the Boltzmann equation: the non-relativistic harmonic oscillator. This simple example is very similar to the full general relativistic version we will encounter in the next section, but the algebra is much less cumbersome. So here the physics will be quite transparent. It will be useful to keep this example

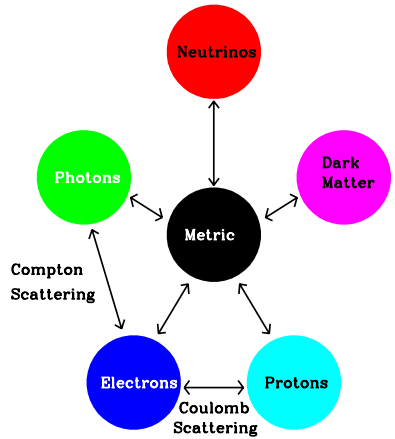


Figure 3.1: The ways in which the different components of the universe interact with each other. These connections are encoded in the coupled Boltzmann equations.

in mind when the algebra threatens to obscure the physics in the next section.

Consider a one dimensional harmonic oscillator with energy

$$E = \frac{p^2}{2m} + \frac{1}{2}kx^2. \tag{3.2}$$

The distribution function of the harmonic oscillator depends on time t, position x, and momentum p. Thus, the full time derivative in equation [3.1] can be rewritten as

$$\frac{df(t,x,p)}{dt} = \frac{\partial f}{\partial t} + \frac{\partial f}{\partial x}\frac{dx}{dt} + \frac{\partial f}{\partial p}\frac{dp}{dt}. \tag{3.3}$$

Figure 3.2 illustrates the movement through phase space of a distribution of collisionless ($C = 0$) oscillators. The full time derivative df/dt vanishes since the number of particles in the bunch at t_1 equals that at t_2. What has changed is the location of the phase space elements $x(t)$ and $p(t)$ themselves. Alternatively, we can think of x and p as independent variables (not dependent on t) and take partial derivatives of f with respect to t, x, and p. All of these partial derivatives are non-zero, but the appropriate weighted sum of the three vanishes.

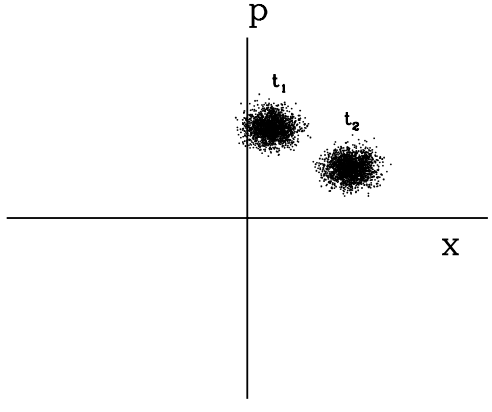


Figure 3.2: Distribution function for a set of collisionless harmonic oscillators. The initial distribution at t_1 moves in phase space by time t_2. The distribution function $f(t,x,p)$ remains constant as long as the evolution of $x(t)$ and $p(t)$ is accounted for.

To determine the coefficients dx/dt and dp/dt we must use the equations of motion. By the definition of momentum,

$$\frac{dx}{dt} \equiv \frac{p}{m}. \tag{3.4}$$

This equation will be generalized to a fully relativistic, three-dimensional version in the next section. Indeed we already got a preview of this when we defined $p^\mu \equiv dx^\mu/d\lambda$ in the previous chapter. Newton's equation governing the motion of the oscillator is

$$\frac{dp}{dt} = -kx. \tag{3.5}$$

The analogue of this familiar equation in the next section will be the geodesic equation of general relativity.

The collisionless Boltzmann equation for the harmonic oscillator is thus

$$\frac{\partial f}{\partial t} + \frac{p}{m}\frac{\partial f}{\partial x} - kx\frac{\partial f}{\partial p} = 0. \tag{3.6}$$

The second term here governs how rapidly the oscillator moves in real space; the coefficient in front is just the velocity, p/m. The last term governs how quickly particles lose momentum.

In order to solve the Boltzmann equation, we need to know the initial conditions on the distribution function. Even without these, though, the Boltzmann equation offers some useful physics. Consider the equilibrium distribution, wherein $\partial f/\partial t = 0$. A general solution for the equilibrium distribution is

$$f(p,x) = f_{\rm EQ}(E); \tag{3.7}$$

that is, f is a function only of energy E. To see that this is indeed a solution, consider

$$\begin{aligned} \frac{p}{m}\frac{\partial f(E)}{\partial x} - kx\frac{\partial f(E)}{\partial p} &= \frac{df}{dE}\left[\frac{p}{m}\frac{\partial E}{\partial x} - kx\frac{\partial E}{\partial p}\right] \\ &= 0. \end{aligned} \tag{3.8}$$

So any function of the energy alone is an equilibrium distribution. Of course, in general, there will be interactions, or collisions. The only way for the full Boltzmann equation to be satisfied is if the collision terms also vanish. This will in general drive f to one of the familiar equilibrium distributions, e.g $e^{-E/T}$ for the classical Maxwell-Boltzmann distribution.

3.2 Collisionless Boltzmann Equation for Photons

Let us begin then by considering the left hand side of equation 3.1 for massless photons. First we must specify the form of the metric, accounting for perturbations around the smooth universe described by equation [2.4]. Whereas the smooth universe is characterized by a single function, $a(t)$, which depends only on time and not on space, the perturbed universe requires two more functions, Ψ and Φ, both of which depend on both space and time. In terms of them, the metric can be written as be

$$\begin{aligned} g_{00}(\vec{x},t) &= -1 - 2\Psi(\vec{x},t) \\ g_{0i}(\vec{x},t) &= 0 \\ g_{ij}(\vec{x},t) &= a^2\delta_{ij}\left(1 + 2\Phi(\vec{x},t)\right). \end{aligned} \tag{3.9}$$

In the absence of Ψ and Φ, equation 3.9 is simply the FRW metric of the zero order homogeneous, flat cosmology. The perturbations to the metric are Ψ, which corresponds to the Newtonian potential and Φ, the perturbation to the spatial curvature. Since the perturbations in the universe are small at the times and scales of interest, we will treat these Ψ and Φ as small quantities, dropping all terms quadratic in them.

There are two technical points about the metric in equation [3.9] which you don't need to worry about for most of this book, but which nonetheless are important to be aware of, if only to better understand the literature. First, one can break up perturbations into those behaving as scalars, vectors, and tensors. Equation [3.9] contains only scalar perturbations. In principle, it is possible that the metric of our universe also has vector or tensor perturbations. If so, g would require other functions besides Ψ and Φ to fully describe all perturbations. For example, the off-diagonal elements become non-zero if there are vector perturbations. Indeed, there are many cosmological theories wherein there are both tensor and vector perturbations. For example, inflation tends to predict that there will be tensor perturbations, while models based on topological defects tend to produce large vector perturbations . For now we focus solely on the scalar perturbations; these are the only ones that couple to matter perturbations and are the most important that couple to photon perturbations as well.

The other feature of equation [3.9] worth noting is that its form corresponds to a choice of *gauge*. That is, even if only scalar perturbations are considered, there is still considerable freedom in the variables one chooses to describe the fluctuations. While any physical results must be insensitive to the gauge choice, it is possible to use a gauge which looks quite different from equation [3.9] and still describe the same physics. For the record, the gauge in equation [3.9] is called the *Conformal Newtonian* gauge*.

We now want to re-express the total derivative in equation 3.1 as a sum of partial derivatives. The distribution function depends on the space-time point $x^\mu = (t, \vec{x})$ and also on the momentum vector defined as

$$P^\mu \equiv \frac{dx^\mu}{d\lambda} \tag{3.10}$$

where λ again parametrizes the particle's path, as in equation [2.5] (and again we will not need to specify λ explicitly). Thus, in principle, f is a function defined in an 8− dimensional space. However, not all the components of the momentum vector are independent since the masslessness of the photon implies that

$$P^2 \equiv g_{\mu\nu} P^\mu P^\nu = 0. \tag{3.11}$$

So there are only three independent components of the momentum vector. Before we choose which three we will use, let us enforce the constraint of equation 3.11, using the metric of equation 3.9.

$$P^2 = 0 = -(1+2\Psi)(P^0)^2 + p^2 = 0 \tag{3.12}$$

where I have defined

$$p^2 \equiv g_{ij} P^i P^j. \tag{3.13}$$

*Historically, the initial ground-breaking work on CMB fluctuations was carried out in synchronous gauge . Recently, the physics of the anisotropies has been elucidated best by using Conformal Newtonian gauge. Problem (2) works out the some of the relevant equations in synchronous gauge.

We can use the constraint equation then to eliminate the time component of P^μ:

$$P^0 = \frac{p}{\sqrt{1+2\Psi}} = p(1-\Psi). \tag{3.14}$$

This last equality holds since we are doing first order perturbation theory in the small quantity Ψ.

Equation 3.14 is the generalization of the relativistic expression $E = pc$ to a perturbed Robertson-Walker metric. It allows us to express P^0 whenever it occurs in terms of p, the generalized magnitude of the momentum. Recall that in the harmonic oscillator case, we did not include a term proportional to $\partial f/\partial E$ in equation [3.3]. Here too, we do not need to include a term proportional to $\partial f/\partial P^0$ when expanding the total time derivative. We need include only the dependence of f on p and the angular direction. Let us therefore define a unit vector γ^i which specifies the direction of the momentum.

$$P^i \equiv \alpha \gamma^i \tag{3.15}$$

To determine the coefficient α, we can use equation 3.13:

$$\begin{aligned} p^2 &= g_{ij}\gamma^i\gamma^j\alpha^2 \\ &= a^2(1+2\Phi)\delta_{ij}\gamma^i\gamma^j\alpha^2 \\ &= a^2(1+2\Phi)\alpha^2 \end{aligned} \tag{3.16}$$

where the last equality holds because the direction vector is a unit vector. Equation 3.16 tells us that $\alpha = p(1-\Phi)/a$ so whenever we encounter P^i, we can always eliminate it in terms of p, γ^i via

$$P^i = p\gamma^i \frac{1-\Phi}{a}. \tag{3.17}$$

The total time derivative can now be expressed as

$$\frac{df}{dt} = \frac{\partial f}{\partial t} + \frac{\partial f}{\partial \vec{x}} \cdot \frac{d\vec{x}}{dt} + \frac{\partial f}{\partial p}\frac{dp}{dt} + \frac{\partial f}{\partial \gamma^i} \cdot \frac{d\gamma^i}{dt}. \tag{3.18}$$

The easiest term in equation 3.18 is the last one since it does not contribute at first order in perturbation theory. To see this, first recall that the zero order distribution function is simply the Bose-Einstein function which depends only on p, not on the direction γ^i. Therefore, $\partial f/\partial\gamma^i$ is non-zero only if we consider the perturbation to the zero order f; i.e. it is a first order term. But so is the term which multiplies it, $d\gamma^i/dt$, for the direction only changes due to the potentials Φ and Ψ. In the absence of these potentials, a given element of phase space moves in a straight line. Thus the last term is the product of two first order terms, rendering it a second order term. We can neglect it.

Next let us re-express the second term on the right hand side of equation 3.18 by recalling that (equation 3.10) $P^i \equiv dx^i/d\lambda$ and $P^0 \equiv dt/d\lambda$. Therefore,

$$\frac{dx^i}{dt} = \frac{dx^i}{d\lambda}\frac{d\lambda}{dt}$$

$$= \frac{P^i}{P^0}$$

$$= \frac{\gamma^i}{a}\,(1+\Psi-\Phi) \tag{3.19}$$

where the last equality follows from equations 3.14 and 3.17. We will see that an overdense region typically has $\Psi < 0$ and $\Phi > 0$. So, equation [3.19] says that a photon slows down (dx/dt becomes smaller) when travelling through an overdense region. This makes perfect sense: we expect the gravitational force of an overdense region to slow down even photons. Having said that, I now claim that we can neglect the potentials in equation 3.19. For, in the Boltzmann equation they multiply $\partial f/\partial x^i$ which is a first order term. (Again, the zero order distribution function does not depend on position.) So collecting terms up to this point, we have

$$\frac{df}{dt} = \frac{\partial f}{\partial t} + \frac{\gamma^i}{a}\frac{\partial f}{\partial x^i} + \frac{\partial f}{\partial p}\frac{dp}{dt}. \tag{3.20}$$

The remaining term to be calculated is dp/dt. Alas, unlike the harmonic oscillator, here $dp/dt \neq -kx$. Rather we will need the geodesic equation from general relativity and more fortitude to compute dp/dt for photons in a perturbed FRW metric.

To begin, let us recall that the time component of the geodesic equation can be written as:

$$\frac{dP^0}{d\lambda} = -\Gamma^0_{\alpha\beta}P^\alpha P^\beta. \tag{3.21}$$

We can rewrite the derivative with respect to λ as a derivative with respect to time, multiplying by the Jacobian $dt/d\lambda = P^0$. Also, we can use equation 3.14 to eliminate P^0 in terms of our favored variable p. Then the geodesic equation reduces to

$$\frac{d}{dt}\left[p(1-\Psi)\right] = -\Gamma^0_{\alpha\beta}\frac{P^\alpha P^\beta}{p}(1+\Psi). \tag{3.22}$$

Expand out the time derivative to get

$$\frac{dp}{dt}(1-\Psi) = p\frac{d\Psi}{dt} - \Gamma^0_{\alpha\beta}\frac{P^\alpha P^\beta}{p}(1+\Psi). \tag{3.23}$$

Now we multiply both sides by $(1+\Psi)$; drop all terms quadratic in Ψ; and reexpress the total time derivative of Ψ in terms of partial derivatives so that

$$\frac{dp}{dt} = p\left\{\frac{\partial\Psi}{\partial t} + \frac{\gamma^i}{a}\frac{\partial\Psi}{\partial x^i}\right\} - \Gamma^0_{\alpha\beta}\frac{P^\alpha P^\beta}{p}(1+2\Psi). \tag{3.24}$$

In order to evaluate dp/dt then we need to evaluate the product $\Gamma^0_{\alpha\beta}P^\alpha P^\beta/p$. Recall that the Christoffel symbol is best written as a sum of derivatives of the metric (equation [2.6]). Here we are interested only in the $\Gamma^0_{\alpha\beta}$ component. It multiplies $P^\alpha P^\beta$, which is symmetric in α, β. Thus, the first two metric derivatives contribute equally, and we have

$$\Gamma^0_{\alpha\beta}\frac{P^\alpha P^\beta}{p} = \frac{g^{0\nu}}{2}\left[2\frac{\partial g_{\nu\alpha}}{\partial x^\beta} - \frac{\partial g_{\alpha\beta}}{\partial x^\nu}\right]\frac{P^\alpha P^\beta}{p}. \tag{3.25}$$

Now $g^{0\nu}$ is non-zero only when $\nu = 0$, in which case it is simply the inverse of g_{00}, so

$$\Gamma^0_{\alpha\beta}\frac{P^\alpha P^\beta}{p} = \frac{-1+2\Psi}{2}\left[2\frac{\partial g_{0\alpha}}{\partial x^\beta} - \frac{\partial g_{\alpha\beta}}{\partial t}\right]\frac{P^\alpha P^\beta}{p}. \tag{3.26}$$

Once again, $g_{0\alpha}$ in the first term in brackets is non-zero only when $\alpha = 0$, in which case its derivative is $-2\partial\Psi/\partial x^\beta$. The second term in brackets multiplied by the product of momenta is

$$\begin{aligned} -\frac{\partial g_{\alpha\beta}}{\partial t}\frac{P^\alpha P^\beta}{p} &= -\frac{\partial g_{00}}{\partial t}\frac{P^0P^0}{p} - \frac{\partial g_{ij}}{\partial t}\frac{P^iP^j}{p} \\ &= 2\frac{\partial\Psi}{\partial t}p - a^2\delta_{ij}\left[2\frac{\partial\Phi}{\partial t} + 2\frac{\dot a}{a}(1+2\Phi)\right]\frac{P^iP^j}{p}. \end{aligned} \tag{3.27}$$

So we have

$$\begin{aligned} \Gamma^0_{\alpha\beta}\frac{P^\alpha P^\beta}{p} &= \frac{-1+2\Psi}{2}\Big[-4\frac{\partial\Psi}{\partial x^\beta}P^\beta + 2p\frac{\partial\Psi}{\partial t} \\ &\qquad -p\left\{2\frac{\partial\Phi}{\partial t} + 2\frac{\dot a}{a}(1+2\Phi)\right\}(1-2\Phi)\Big] \\ &= \frac{-1+2\Psi}{2}\left[-4(\frac{\partial\Psi}{\partial t}p + \frac{\partial\Psi}{\partial x^i}\frac{p\gamma^i}{a}) + 2p\frac{\partial\Psi}{\partial t} - p\Big\{2\frac{\partial\Phi}{\partial t} + 2\frac{\dot a}{a}\Big\}\right] \\ &= \{-1+2\Psi\}\left[-\frac{\partial\Psi}{\partial t}p - 2\frac{\partial\Psi}{\partial x^i}\frac{p\gamma^i}{a} - p\Big\{\frac{\partial\Phi}{\partial t} + \frac{\dot a}{a}\Big\}\right]. \end{aligned} \tag{3.28}$$

We can insert this into equation 3.24 to get

$$\frac{dp}{dt} = p\left\{\frac{\partial\Psi}{\partial t} + \frac{\gamma^i}{a}\frac{\partial\Psi}{\partial x^i}\right\} - \frac{\partial\Psi}{\partial t}p - 2\frac{\partial\Psi}{\partial x^i}\frac{p\gamma^i}{a} - p\Big\{\frac{\partial\Phi}{\partial t} + \frac{\dot a}{a}\Big\}. \tag{3.29}$$

Collecting terms, we finally have

$$\frac{1}{p}\frac{dp}{dt} = -\frac{\dot a}{a} - \frac{\partial\Phi}{\partial t} - \frac{\gamma^i}{a}\frac{\partial\Psi}{\partial x^i}. \tag{3.30}$$

Equation [3.30] is what we were after. It describes the change in the photon energy as it moves through a perturbed FRW universe. The first term accounts for the loss of energy due to expansion. The second says that a photon in a deepening gravitational well ($\partial\Phi/\partial t > 0$) loses energy. This is understandable: the deepening well makes it more difficult for the photon to emerge, thereby increasing the magnitude of the red-shift. Finally, a photon travelling into a well ($\gamma^i\partial\Psi/\partial x^i < 0$) gains energy because it is being pulled towards the center.

We are now in a position to write down the Boltzmann equation for photons. Using equation 3.30 in equation 3.20 leads to

$$\frac{df}{dt} = \frac{\partial f}{\partial t} + \frac{\gamma^i}{a}\frac{\partial f}{\partial x^i} - p\frac{\partial f}{\partial p}\left[\frac{\dot a}{a} + \frac{\partial\Phi}{\partial t} + \frac{\gamma^i}{a}\frac{\partial\Psi}{\partial x^i}\right]. \tag{3.31}$$

This equation incorporates much of the physics with which we are already familiar, such as the fact that photons redshift in an expanding universe. It also leads directly to the equations governing anisotropies. Working through the terms on the right, the first two are familiar from standard hydrodynamics; when integrated, they lead to the Euler equations. The third term dictates that photons lose energy in an expanding universe. We saw some of this in the last chapter when considering geodesics. Shortly, we see how the Boltzmann formalism enforces this result. Finally, the last two encode the affect of overdense regions on the photon distribution function.

To go further we must now expand the photon distribution function f about its zero order Bose-Einstein value. I will do this in a way that may seem odd at first. Let us write

$$f(\vec{x}, p, \hat{\gamma}, t) = \left[\exp\left\{\frac{p}{T^{(0)}(t)[1+\Theta(\vec{x},\hat{\gamma},t)]}\right\} - 1\right]^{-1}. \tag{3.32}$$

Here the zero order temperature $T^{(0)}$ is a function of time only (i.e. scales as a^{-1}), not space. The perturbation to the distribution function is encompassed in Θ, which could also be called $\delta T/T$. There is one assumption built into equation 3.32. I have explicitly written down that Θ depends on $\vec{x}, \hat{\gamma}$, and t. This assumes that it does *not* depend on the magnitude of the momentum p. We will soon see that this is a valid assumption, following directly from that fact that the magnitude of the photon momentum is virtually unchanged during a Compton scatter. The perturbation Θ is small, so we can expand (again keeping only terms up to first order)

$$\begin{aligned} f &= \left[\exp\left\{\frac{p}{T^{(0)}}\right\} - 1\right]^{-1} + \frac{\partial}{\partial T^{(0)}}\left[\exp\left\{\frac{p}{T^{(0)}}\right\} - 1\right]^{-1} T^{(0)}\Theta \\ &= f^{(0)} - p\frac{\partial f^{(0)}}{\partial p}\Theta. \end{aligned} \tag{3.33}$$

In the last line I have defined the zero order distribution function as

$$f^{(0)} \equiv \left[\exp\left\{\frac{p}{T^{(0)}}\right\} - 1\right]^{-1} \tag{3.34}$$

and made use of the fact that for this function $T^{(0)}\partial f^{(0)}/\partial T^{(0)} = -p\partial f^{(0)}/\partial p$.

3.2.1 Zero Order Equation

We can now set about systematically collecting the terms of similar order in equation 3.31. Let us start with the zero order terms, those with no Φ, Ψ, or Θ. These lead immediately to

$$\left.\frac{df}{dt}\right|_{\text{zero order}} = \frac{\partial f^{(0)}}{\partial t} - \frac{\dot{a}}{a}p\frac{\partial f^{(0)}}{\partial p} = 0. \tag{3.35}$$

I have set df/dt here equal to zero. I could justify this by claiming that we are now looking only at the collisionless Boltzmann equation. But there is a much deeper justification. In fact, even when we come around to including collisions, we will see that there is no zero order

collision term. That is, the collision terms will be proportional to Θ and other perturbatively small quantities. There is a profound reason for this: the zero order distribution function is set precisely by the requirement that the collision term vanishes. Another, perhaps more familiar way, of saying this is to point out that any collision term includes the rate for the given reaction and for its inverse. If the distribution functions are set to their equilibrium values, the rate for the reaction precisely cancels the rate for its inverse. If a given component is out of equilibrium, collisions will drive it towards its equilibrium distribution. This is the reason we expected a Bose-Einstein distribution in the first place. Its observation is convincing evidence that photons were at one point in the early universe tightly coupled to the electrons.

Returning to equation 3.35, we can rewrite the time derivative as

$$\frac{\partial f^{(0)}}{\partial t} = \frac{\partial f^{(0)}}{\partial T^{(0)}}\frac{dT^{(0)}}{dt} = -\frac{dT^{(0)}/dt}{T^{(0)}}p\frac{\partial f^{(0)}}{\partial p}$$

so that the zero order equation becomes

$$\left[-\frac{dT^{(0)}/dt}{T^{(0)}} - \frac{da/dt}{a}\right]\frac{\partial f^{(0)}}{\partial p} = 0 \tag{3.36}$$

Thus $dT^{(0)}/T^{(0)} = -da/a$ or

$$T^{(0)} \propto \frac{1}{a}. \tag{3.37}$$

This is precisely what we expected from the heuristic argument about the photon's wavelength getting stretched as the universe expands (§1.1) and the more concrete argument of §2.1. It is reassuring to see this result emerge from the Boltzmann treatment.

3.2.2 First Order Equation

We now return to equation 3.31 and extract the equation for the deviation of the photon temperature from its zero order value, i.e. an equation for Θ. To do this, everywhere we encounter f in equation 3.31, we insert the expansion of equation [3.33]

$$\begin{aligned} \left.\frac{df}{dt}\right|_{\text{first order}} &= -p\frac{\partial}{\partial t}\left[\frac{\partial f^{(0)}}{\partial p}\Theta\right] - p\frac{\gamma^i}{a}\frac{\partial\Theta}{\partial x^i}\frac{\partial f^{(0)}}{\partial p} + \frac{\dot{a}}{a}p\Theta\frac{\partial}{\partial p}\left[p\frac{\partial f^{(0)}}{\partial p}\right] \\ &\quad -p\frac{\partial f^{(0)}}{\partial p}\left[\frac{\partial\Phi}{\partial t} + \frac{\gamma^i}{a}\frac{\partial\Psi}{\partial x^i}\right]. \end{aligned} \tag{3.38}$$

Consider the first and third terms on the right hand side here. The time derivative in the first can be rewritten as a temperature derivative so these two terms are

$$\begin{aligned} -p\frac{\partial f^{(0)}}{\partial p}\frac{\partial\Theta}{\partial t} &- p\Theta\frac{dT^{(0)}}{dt}\frac{\partial^2 f^{(0)}}{\partial T^{(0)}\partial p} + \frac{\dot{a}}{a}p\Theta\frac{\partial}{\partial p}\left[p\frac{\partial f^{(0)}}{\partial p}\right] \\ &= -p\frac{\partial f^{(0)}}{\partial p}\frac{\partial\Theta}{\partial t} + p\Theta\frac{dT^{(0)}/dt}{T^{(0)}}\frac{\partial}{\partial p}\left[p\frac{\partial f^{(0)}}{\partial p}\right] + \frac{\dot{a}}{a}p\Theta\frac{\partial}{\partial p}\left[p\frac{\partial f^{(0)}}{\partial p}\right] \end{aligned}$$

$$= \quad -p\frac{\partial f^{(0)}}{\partial p}\frac{\partial \Theta}{\partial t}. \tag{3.39}$$

We can finally write down the equation governing the perturbation Θ:

$$\left.\frac{df}{dt}\right|_{\text{first order}} = -p\frac{\partial f^{(0)}}{\partial p}\left[\frac{\partial \Theta}{\partial t} + \frac{\gamma^i}{a}\frac{\partial \Theta}{\partial x^i} + \frac{\partial \Phi}{\partial t} + \frac{\gamma^i}{a}\frac{\partial \Psi}{\partial x^i}\right]. \tag{3.40}$$

The first two terms here account for "free streaming," which translates into anisotropies on increasingly small scales as the universe evolves. The last two account for the effect of gravity. Note that every time x appears it is coupled with a, the scale factor. This must happen, for physical distances are ax.

3.3 Collision Terms: Compton Scattering

Our task in this section is to determine the influence Compton scattering has on the photon distribution function. The scattering process of interest (3.3) is

$$e^-(\vec{q}) + \gamma(\vec{p}) \leftrightarrow e^-(\vec{q}\,') + \gamma(\vec{p}\,'), \tag{3.41}$$

where I have explicitly indicated the momentum of each particle.

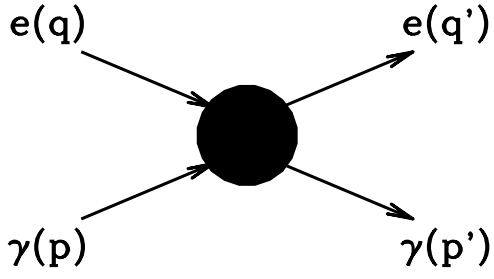


Figure 3.3: Compton scattering of a photon off an electron.

We are interested in the change of distribution of photons with momentum $\vec{p}$ (with magnitude p and direction $\hat{\gamma}$). Therefore we must sum over all other momenta $(\vec{q}, \vec{q}\,', \vec{p}\,')$ which affect $f(\vec{p})$. Schematically, then, the collision term is

$$C[f(\vec{p})] = \sum_{\vec{q},\vec{q}\,',\vec{p}\,'} |\text{Amplitude}|^2 \left\{f_e(\vec{q}\,')f(\vec{p}\,') - f_e(\vec{q})f(\vec{p})\right\}. \tag{3.42}$$

The amplitude is reversible so it multiplies both the reaction and its inverse. The products of the electron distribution function f_e and the photon distribution function simply count the number of particles with the given momenta. I have neglected stimulated emission and Pauli

blocking, which would lead to factors of $1+f$ and $1-f_e$ with the appropriate momenta. At first order this turns out to be a valid assumption. If one were to go to second order though, stimulated emission would have to be included. Pauli blocking is never important after electron-positron annihilation because the occupation numbers f_e are very small (Problem 2.4).

Unfortunately, the collision term becomes a bit messier than the schematic version when we put in all the factors of 2π to properly account for the sums over phase space. Explicitly, the collision term is

$$\begin{aligned} C[f(\vec{p})] &= \frac{1}{p}\int \frac{d^3q}{(2\pi)^3 2E_e(q)}\int \frac{d^3q'}{(2\pi)^3 2E_e(q')}\int \frac{d^3p'}{(2\pi)^3 2E(p')}|\mathcal{M}|^2(2\pi)^4 \\ &\quad \times \delta^3\left[\vec{p}+\vec{q}-\vec{p}\,'-\vec{q}\,'\right]\delta\left[E(p)+E(q)-E(p')-E(q')\right] \\ &\quad \times \left\{f_e(\vec{q}\,')f(\vec{p}\,') - f_e(\vec{q})f(\vec{p})\right\}. \end{aligned} \tag{3.43}$$

Here the delta functions enforce energy momentum conservation. The energies at this order are the simple ones: $E(p) = p$ and $E_e(q) = m_e + q^2/(2m_e)$. Since the kinetic energy of the electrons is very small at the epochs of interest compared with their rest energy, the factors of E_e in the denominator of equation [3.43] may be replaced with m_e. Then using the three dimensional momentum delta function to eliminate $\vec{q}\,'$, we have

$$\begin{aligned} C[f(\vec{p})] &= \frac{\pi}{4m_e^2 p}\int \frac{d^3q}{(2\pi)^3}\int \frac{d^3p'}{(2\pi)^3 p'}\,\delta\left[p + \frac{q^2}{2m_e} - p' - \frac{(\vec{q}+\vec{p}-\vec{p}\,')^2}{2m_e}\right] \\ &\quad \times \ |\mathcal{M}|^2\left\{f_e(\vec{q}+\vec{p}-\vec{p}\,')f(\vec{p}\,') - f_e(\vec{q})f(\vec{p})\right\}. \end{aligned} \tag{3.44}$$

To go further, we need to understand the kinematics of non-relativistic Compton scattering. The most important feature of this process for our purposes is that very little energy is transferred. In particular,

$$\begin{aligned} E_e(q) - E_e(\vec{q}+\vec{p}-\vec{p}\,') &\simeq \frac{(\vec{p}\,'-\vec{p})\cdot\vec{q}}{m_e} \\ &= \mathcal{O}\left(\frac{T^{(0)}q}{m_e}\right) \end{aligned} \tag{3.45}$$

where the last equality holds since both p and p' are of order $T^{(0)}$. In the early universe, the momentum of the electrons[†] is due to their bulk velocity $\vec{v}$. Thus the change in the electron energy is of order $T^{(0)}v$. Since v is a small quantity (we will treat it as the same order as our other perturbative quantities), the ratio of the change in energy to the temperature is small.

It makes sense, therefore, to expand the final electron kinetic energy $(\vec{q}+\vec{p}-\vec{p}\,')^2/(2m_e)$ around its zero order value of $q^2/(2m_e)$. The delta function can be expanded as

$$\delta\left[p + \frac{q^2}{2m_e} - p' - \frac{(\vec{q}+\vec{p}-\vec{p}\,')^2}{2m_e}\right] = \delta(p-p') + \frac{(\vec{p}-\vec{p}\,')\cdot\vec{q}}{m_e}\frac{\partial\delta(p-p')}{\partial p'}. \tag{3.46}$$

[†]There is another contribution to the momentum coming from the thermal motion of the electrons, but this does not contribute to the first order collision term.

This formal expansion appears ill-defined at present, but when integrating over momenta, the derivatives of delta functions can be handled by integrating by parts. With this expansion, and using the fact that $f_e(\vec{q}+\vec{p}-\vec{p}\,') \simeq f_e(\vec{q})$, the collision term becomes

$$\begin{aligned} C[f(\vec{p})] &= \frac{\pi}{4m_e^2 p}\int d^3q \frac{f_e(\vec{q})}{(2\pi)^3}\int \frac{d^3p'}{(2\pi)^3 p'}|\mathcal{M}|^2 \\ &\times \left\{\delta(p-p') + \frac{(\vec{p}-\vec{p}\,')\cdot\vec{q}}{m_e}\frac{\partial\delta(p-p')}{\partial p'}\right\}\{f(\vec{p}\,') - f(\vec{p})\}. \end{aligned} \tag{3.47}$$

To proceed further, we will take the amplitude squared to be constant

$$|\mathcal{M}|^2 = 8\pi\sigma_T m_e^2. \tag{3.48}$$

This is wrong, and it is wrong for two reasons. First of all, the amplitude squared has an angular dependence $\propto (1+\cos^2[\hat{p}\cdot\hat{p}\,'])$. Ignoring this angular dependence, as I now propose to do, makes a small difference in the final collision term. It needs to be included in calculations which aspire to one percent accuracy. But it would simply distract us here, so let us ignore it for the present. The second reason a constant amplitude is wrong is a little more subtle and, when properly accounted for, opens up a whole new branch of CMB study. In particular, the amplitude squared has a polarization dependence ($\propto |\hat{\epsilon}\cdot\hat{\epsilon}'|^2$, where $\hat{\epsilon}$ and $\hat{\epsilon}'$ are the polarizations of the incoming and outgoing photons) which I have implicitly summed over here. The dependence on polarization means that at a small level the CMB will be polarized due to Compton scattering, . It turns out that the information carried by the polarization spectrum is as valuable as that carried by the temperature spectrum . At present there is therefore a big push to devote some of the sensitivity of the upcoming satellites to measure polarization. Even if we were not concerned with polarization, the temperature anisotropies are coupled to the polarization field, so an accurate determination of the former requires a treatment of the latter. Again, though, I will neglect this small effect here in the derivation of the collision term. It is straightforward to include both the effects of polarization and the angular dependence of Compton scattering using the same formalism we are now in the midst of. The algebra is simply a bit more tedious.

Once we have assumed that $|\mathcal{M}|^2$ is constant, we can multiply out the terms in brackets in equation [3.47] keeping only first order terms. Also, the $\vec{q}$ integral simply gives a factor of n_e (or $n_e\vec{v}$ for the term which has a factor of $\vec{q}/m_e$). So,

$$\begin{aligned} C[f(\vec{p})] &= \frac{2\pi^2 n_e\sigma_T}{p}\int\frac{d^3p'}{(2\pi)^3 p'}\left\{\delta(p-p') + (\vec{p}-\vec{p}\,')\cdot\vec{v}\frac{\partial\delta(p-p')}{\partial p'}\right\} \\ &\quad\times\{f(\vec{p}\,') - f(\vec{p})\} \\ &= \frac{n_e\sigma_T}{4\pi p}\int_0^\infty dp'p'\int d\Omega'\Big[\delta(p-p')\left(-p'\frac{\partial f^{(0)}}{\partial p'}\Theta(\hat{\gamma}\,') + p\frac{\partial f^{(0)}}{\partial p}\Theta(\hat{\gamma})\right) \\ &\quad + (\vec{p}-\vec{p}\,')\cdot\vec{v}\frac{\partial\delta(p-p')}{\partial p'}(f^{(0)}(p') - f^{(0)}(p))\Big] \end{aligned} \tag{3.49}$$

where Ω' is the solid angle spanned by $\hat{\gamma}\,'$. To get the second equality here, I have broken up the difference $f(\vec{p}\,') - f(\vec{p})$ into a zeroth order piece – which doesn't contribute when multiplying $\delta(p'p')$ – and a first order part – which can be neglected when multiplying the velocity term.

There are only two terms in equation [3.49] which depend on $\hat{\gamma}\,'$. First, there is the perturbation to the distribution function, $\Theta(\hat{\gamma}\,')$. It is convenient at this stage to introduce the notation:

$$\Theta_0 \equiv \frac{1}{4\pi}\int d\Omega'\Theta(\hat{\gamma}\,'). \tag{3.50}$$

So Θ_0 does not depend on the direction vector; it is an integral of the perturbation over all directions. In other words, it is the monopole part of the perturbation. Note that we cannot absorb this monopole into the definition of the zero order temperature since the latter is constant over all space. The perturbation Θ_0 therefore represents the deviation of the monopole at a given point in space from its average in all space. Later on we will generalize equation [3.50] to all other multipoles.

The second term in equation [3.49] which depends on $\hat{\gamma}\,'$ is the explicit factor $\hat{\gamma}\,'\cdot\vec{v}$. This term integrates to zero since $\vec{v}$ is a fixed vector. Thus, the integration over $\hat{\gamma}\,'$ leaves

$$\begin{aligned} C[f(\vec{p})] &= \frac{n_e\sigma_T}{p}\int_0^\infty dp'p'\Big[\delta(p-p')\left(-p'\frac{\partial f^{(0)}}{\partial p'}\Theta_0 + p\frac{\partial f^{(0)}}{\partial p}\Theta(\hat{\gamma})\right) \\ &\quad + \vec{p}\cdot\vec{v}\frac{\partial\delta(p-p')}{\partial p'}(f^{(0)}(p') - f^{(0)}(p))\Big]. \end{aligned} \tag{3.51}$$

Now the p' integral can be done; in the first line by trivially integrating over the delta function and in the second by integrating by parts. We are left with

$$C[f(\vec{p})] = -p\frac{\partial f^{(0)}}{\partial p}n_e\sigma_T\left[\Theta_0 - \Theta(\hat{\gamma}) + \hat{\gamma}\cdot\vec{v}\right]. \tag{3.52}$$

Already, we can anticipate the effect of Compton scattering on the photon distribution. In the absence of a bulk velocity for the electrons ($v=0$), the collision terms serve to drive Θ to Θ_0. That is, when Compton scattering is very efficient, only the monopole perturbation survives; all other moments are washed out. The situation changes slightly if the electrons carry a bulk velocity. In that case, the photons will also have a dipole moment, fixed by the amplitude and direction of the electron velocity. Even in this case, though, all higher moments vanish. Thus Compton scattering produces a photon distribution which is extremely simple to categorize: it has only a non-vanishing monopole and dipole. This is equivalent to saying that the photons behave like a fluid. Indeed, strong scattering, or tight coupling, produces a situation wherein the photons and electrons behave as a single fluid.

3.4 The Boltzmann Equation for Photons

We can now collect the left and right hand sides of the Boltzmann equations from the previous two sections. A few more definitions will complete the first goal of this chapter,

a linear equation for the perturbation to the photon distribution. Equating equation [3.40] and equation [3.52] leads to

$$\frac{\partial\Theta}{\partial t}+\frac{\gamma^i}{a}\frac{\partial\Theta}{\partial x^i}+\frac{\partial\Phi}{\partial t}+\frac{\gamma^i}{a}\frac{\partial\Psi}{\partial x^i}=n_e\sigma_T\left[\Theta_0-\Theta+\hat{\gamma}\cdot\vec{v}\right]. \tag{3.53}$$

At this point, it is convenient to re-introduce the conformal time η, defined in equation [2.28], as our time variable. In terms of the conformal time, the Boltzmann equation becomes

$$\dot{\Theta}+\gamma^i\frac{\partial\Theta}{\partial x^i}+\dot{\Phi}+\gamma^i\frac{\partial\Psi}{\partial x^i}=n_e\sigma_T a\left[\Theta_0-\Theta+\hat{\gamma}\cdot\vec{v}\right]. \tag{3.54}$$

Here,and from now on, overdots represent derivatives with respect to conformal time.

Equation 3.54 is a partial differential linear equation coupling Θ to other variables Φ, Ψ, and $\vec{v}$ which also behave linearly. If we Fourier transform all these variables, the resulting Fourier amplitudes obey ordinary differential equations, which are much simpler to solve. Note that this simplification arises because the perturbations are small (equivalently the equations are linear). In this case, the different Fourier modes all evolve independently. Perturbations to the CMB remain small at all cosmological epochs, so Fourier transforms are very useful. In contrast, perturbations to matter are more complicated. Initially they are small, and they remain small until relatively recently. The largest scales today are still in the linear regime, so Fourier transforming is certainly useful for the matter perturbations as well. However, to completely characterize the matter field today requires accounting for non-linearities, and for this purpose, Fourier transforms lose much of their appeal. Different Fourier modes couple when non-linear behavior becomes important, so the codes which follow matter perturbations all the way until today work in real space. Even these codes, however, start at $z\sim 20$ with the initial conditions set by linear evolution.

Our Fourier convention will be

$$\Theta(\vec{x})=\int\frac{d^3k}{(2\pi)^3}e^{i\vec{k}\cdot\vec{x}}\Theta(\vec{k}). \tag{3.55}$$

I have avoided writing the Fourier amplitude as $\bar{\Theta}$, which may cause a little confusion in equation [3.55]. From now on, though, we will almost never refer to real space variables, so there should be no further ambiguity.

Before rewriting equation [3.54] in terms of Fourier modes, let us make two final definitions. First, define the cosine of the angle between the wavenumber $\vec{k}$ and the photon direction γ to be

$$\mu\equiv\frac{\vec{k}\cdot\hat{\gamma}}{k}=\frac{\vec{k}\cdot\vec{p}}{kp}. \tag{3.56}$$

Note that, if we assume that the velocity is irrotational (as we will), then $\vec{v}$ is parallel to $\vec{k}$, so $\vec{v}\cdot\hat{\gamma}=v\mu$. So a good way to think about the variable μ is that it characterizes the angle between the photon direction and an arbitrary $\hat{z}-$ axis chosen to lie in the direction of the electron velocity vector. Next, we define the optical depth

$$\tau(\eta)\equiv\int_\eta^{\eta_0}d\eta'\;n_e\sigma_T a. \tag{3.57}$$

At late times, the free electron density is small, so $\tau << 1$, while at early times, it is very large. Note that I have defined the limits of integration in such a way that

$$\dot{\tau}\equiv\frac{d\tau}{d\eta}=-n_e\sigma_T a. \tag{3.58}$$

With this definition, we are finally left with

$$\dot{\Theta}+ik\mu\Theta+\dot{\Phi}+ik\mu\Psi=-\dot{\tau}\left[\Theta_0-\Theta+\mu v\right]. \tag{3.59}$$

3.5 Boltzmann Equation for Cold Dark Matter

We can apply the formalism developed in the previous sections to derive the Boltzmann equation for any other constituent in the universe. Of particular importance is the evolution of the dark matter. In almost all currently popular models of structure formation, dark matter plays an important role in structure formation and in determining the gravitational field in the universe.

There are several ways in which the dark matter distribution differs from that of the photons. First, by definition, "dark" matter does not interact with any of the other constituents in the universe. Thus we need not deal with any collision terms. Second, cold dark matter, in contrast to the photons, is non-relativistic. So we need to redo some of the kinematics which led to the left side of the Boltzmann equation. In particular, the constraint equation 3.11 now becomes

$$g_{\mu\nu}P^\mu P^\nu=-m^2 \tag{3.60}$$

where m is the mass of the dark matter particle. It is also useful to define the energy as

$$E\equiv\sqrt{p^2+m^2}, \tag{3.61}$$

where p is defined exactly as in equation [3.13]: $p^2=g_{ij}P^iP^j$. In the massless case, of course, equation [3.61] says that $E=p$, so E is superfluous. Here it will be convenient to let E replace p as one of the variables on which the distribution function depends (in addition to position $\vec{x}$, time t, and the direction vector $\hat{\gamma}$). We can now derive the equivalent of equations 3.14 and 3.17 for the four-momentum of a massive particle:

$$P^\mu=\left[E(1-\Psi),p\gamma^i\frac{1-\Phi}{a}\right]. \tag{3.62}$$

Only the time component is different than that of a massless particle, with E replacing p.

Using E as one of the dependent variables means that the total time derivative of the distribution function is

$$\frac{df_{\rm DM}}{dt}=\frac{\partial f_{\rm DM}}{\partial t}+\frac{\partial f_{\rm DM}}{\partial\vec{x}}\frac{d\vec{x}}{dt}+\frac{\partial f_{\rm DM}}{\partial E}\frac{dE}{dt}+\frac{\partial f_{\rm DM}}{\partial\gamma^i}\frac{d\gamma^i}{dt}. \tag{3.63}$$

Once again, the last term here vanishes since it is the product of two first order terms. Due to the change in the constraint equation, the coefficients of the derivatives of the distribution

function with respect to x^i and E are slightly different than they were in the massless case. Working through the algebra, which is otherwise identical to the calculation presented in section 2.2, leads to the collisionless Boltzmann equation for non-relativistic matter:

$$\frac{\partial f_{\rm DM}}{\partial t} + \frac{\gamma^i}{a}\frac{p}{E}\frac{\partial f_{\rm DM}}{\partial x^i} - \frac{\partial f_{\rm DM}}{\partial E}\left[\frac{da/dt}{a}\frac{p^2}{E} + \frac{p^2}{E}\frac{\partial \Phi}{\partial t} + \frac{\gamma^i p}{a}\frac{\partial \Psi}{\partial x^i}\right] = 0. \tag{3.64}$$

Equation 3.64 reduces to equation [3.31] in the massless limit as it must. The main difference between the two is the presence of factors of p/E, or velocity. For dark matter particles, these velocity factors suppress any free streaming, as we will shortly see.

In the massless case, to proceed further we used our knowledge of the distribution function. Namely, we knew that the zero order distribution function was Planckian, and we perturbed around this zero order solution. For cold dark matter particles, we do not need such detailed information about the zero order distribution function. All we need to know is that these particles are very non-relativistic. As an example consider a typical supersymmetric dark matter candidate with mass of order one TeV. For such a particle, the average value of p/E due to thermal motion alone is expected to be

$$\begin{aligned} < \frac{p}{E} > \; &\simeq \; < p/m > \simeq \sqrt{T/m} \\ &= \; 10^{-6}\left(\frac{T}{1\mathrm{eV}}\right)^{1/2}. \end{aligned} \tag{3.65}$$

This is much smaller than the expected amplitude of our "small" perturbations: when $T \sim 1$ eV[‡] , the amplitude of the density perturbation, for example, is of order 10^{-4}. As the temperature drops, the thermal motion becomes less and less important since the perturbations are growing and thermal motion is getting weaker. Having understood this, we might be tempted to throw out all terms with factors of p/E in them. This would be a mistake: the density perturbations themselves induce velocity flows in the dark matter via the continuity equation. These cannot be neglected. However, in our linear treatment, it is consistent to neglect terms second order in p/E.

Instead of assuming a form for $f_{\rm DM}$, we will take moments of equation [3.64]. First, multiply both sides by the phase space volume $d^3p/(2\pi)^3$ and integrate. This leads to

$$\begin{aligned} \frac{\partial}{\partial t}\int \frac{d^3p}{(2\pi)^3} f_{\rm DM} \; &+ \; \frac{1}{a}\frac{\partial}{\partial x^i}\int \frac{d^3p}{(2\pi)^3} f_{\rm DM}\frac{p\gamma^i}{E} \; - \; \left[\frac{da/dt}{a} + \frac{\partial \Phi}{\partial t}\right]\int \frac{d^3p}{(2\pi)^3}\frac{\partial f_{\rm DM}}{\partial E}\frac{p^2}{E} \\ &- \frac{1}{a}\frac{\partial \Psi}{\partial x^i}\int \frac{d^3p}{(2\pi)^3}\frac{\partial f_{\rm DM}}{\partial E}\gamma^i p = 0. \end{aligned} \tag{3.66}$$

Note that, since they are independent variables, the integral over p passes through the partial derivatives with respect to x^i and t. The last term here can be neglected since the integral

[‡]Here I have implicitly assumed that the dark matter temperature equals that of the photons. In fact, since the photons get heated as different species annihilate early in the universe, the dark matter temperature is significantly lower than that of the photons. Their thermal motion is even less important than suggested by equation [3.65]

over the direction vector is non-zero only for the perturbed part of $f_{\rm DM}$. Thus the integral is first order and it multiplies the first order term $\partial\Psi/\partial x^i$. The rest of the terms are all relevant though. To simplify, let us recall that the dark matter density is

$$n_{\rm DM} = \int \frac{d^3p}{(2\pi)^3} f_{\rm DM} \tag{3.67}$$

while the velocity is defined as

$$v^i_{\rm DM} \equiv \frac{1}{n_{\rm DM}}\int \frac{d^3p}{(2\pi)^3} f_{\rm DM}\frac{p\gamma^i}{E}. \tag{3.68}$$

The first two terms in equation [3.66] can be simply expressed in terms of the velocity and the density. The third term is a bit more subtle; to relate it to the density, we need to integrate by parts. Since $dE/dp = p/E$, the integrand can be re-expressed as $p\partial f_{\rm DM}/\partial p$. Thus, the integral becomes

$$\begin{aligned} \int \frac{d^3p}{(2\pi)^3} p\frac{\partial f_{\rm DM}}{\partial p} &= \frac{4\pi}{(2\pi)^3}\int_0^\infty dp p^3 \frac{\partial f_{\rm DM}}{\partial p} \\ &= -3\frac{4\pi}{(2\pi)^3}\int_0^\infty dp p^2 f_{\rm DM} \\ &= -3 n_{\rm DM}. \end{aligned} \tag{3.69}$$

So the zeroth moment of the Boltzmann equation leads to the cosmological generalization of the continuity equation:

$$\frac{\partial n_{\rm DM}}{\partial t} + \frac{1}{a}\frac{\partial (n_{\rm DM} v^i_{\rm DM})}{\partial x^i} + 3\left[\frac{da/dt}{a} + \frac{\partial \Phi}{\partial t}\right] n_{\rm DM} = 0. \tag{3.70}$$

The first two terms here are the standard continuity equation from fluid mechanics. The last term arises due to the FRW metric and its perturbations.

To go further, we can collect zeroth order and first order terms in equation [3.70]. The velocity is first order as is Φ so the only zeroth order terms are:

$$\frac{\partial n^{(0)}_{\rm DM}}{\partial t} + 3\frac{da/dt}{a} n^{(0)}_{\rm DM} = 0 \tag{3.71}$$

where $n^{(0)}_{\rm DM}$ is the zeroth order, homogeneous part of the density. Equivalently, we have

$$\frac{d(n^{(0)}_{\rm DM} a^3)}{dt} = 0 \Longrightarrow n^{(0)}_{\rm DM} \propto a^{-3} \tag{3.72}$$

a relation we anticipated early on as a generic characteristic of the expansion.

Now let us extract the first order part of equation [3.70]. All factors of $n_{\rm DM}$ multiplying the first order quantities $v_{\rm DM}$ and Φ may be set to $n^{(0)}_{\rm DM}$. Everywhere else, we need to expand $n_{\rm DM}$ out to include a first order perturbation. In particular, we will set

$$n_{\rm DM} = n^{(0)}_{\rm DM}\left[1 + \delta_{\rm DM}(\vec{x}, t)\right], \tag{3.73}$$

which defines the first order piece as $n_{\rm DM}^{(0)}\delta_{\rm DM}$. After dividing by $n_{\rm DM}^{(0)}$, the first order equation is therefore

$$\frac{\partial\delta_{\rm DM}}{\partial t}+\frac{1}{a}\frac{\partial v_{\rm DM}^i}{\partial x^i}+3\frac{\partial\Phi}{\partial t}=0. \tag{3.74}$$

As it stands, we have introduced two new perturbation variables for the dark matter, the density perturbation $\delta_{\rm DM}$ and the velocity $v_{\rm DM}$. Equation 3.74 is only one equation, though, for these two variables. We need another. To get it, we return to the unintegrated Boltzmann equation, 3.64. We have just taken its zeroth moment; to extract a second equation, let us take its first moment. In particular, multiply equation [3.64] by $d^3p(p/E)\gamma^j/(2\pi)^3$ and then integrate. The first moment equation is then

$$\begin{aligned}\frac{\partial}{\partial t}\int\frac{d^3p}{(2\pi)^3}f_{\rm DM}\frac{p\gamma^j}{E} &+ \frac{1}{a}\frac{\partial}{\partial x^i}\int\frac{d^3p}{(2\pi)^3}f_{\rm DM}\frac{p^2\gamma^i\gamma^j}{E^2}\\ &- \left[\frac{da/dt}{a}+\frac{\partial\Phi}{\partial t}\right]\int\frac{d^3p}{(2\pi)^3}\frac{\partial f_{\rm DM}}{\partial E}\frac{p^3\gamma^j}{E^2}\\ &- \frac{1}{a}\frac{\partial\Psi}{\partial x^i}\int\frac{d^3p}{(2\pi)^3}\frac{\partial f_{\rm DM}}{\partial E}\frac{\gamma^i\gamma^j p^2}{E}=0.\end{aligned} \tag{3.75}$$

The first two terms are straightforward: the first is the time derivative of $n_{\rm DM}v_{\rm DM}^i$ while the second can be safely neglected since it is of order $<(p/E)^2>$. The last sets of terms must be handled more carefully, though, because of the partial derivatives. Since $(p/E)\partial/\partial E=\partial/\partial p$ the third term is actually of order p/E while the last is independent of velocity. Let us do the integration by parts explicitly in the third term. The integral is:

$$\begin{aligned}\int\frac{d^3p}{(2\pi)^3}\frac{\partial f_{\rm DM}}{\partial p}\frac{p^2\gamma^j}{E} &= \int\frac{d\Omega\gamma^j}{(2\pi)^3}\int_0^\infty dp\frac{p^4}{E}\frac{\partial f_{\rm DM}}{\partial p}\\ &= -\int\frac{d\Omega\gamma^j}{(2\pi)^3}\int_0^\infty dp f_{\rm DM}\left(\frac{4p^3}{E}-\frac{p^5}{E^3}\right).\end{aligned} \tag{3.76}$$

The p^5/E^3 term is completely negligible, so the integral is $-4n_{\rm DM}v_{\rm DM}^j$. The same steps carry through for the last term in equation [3.75]; the one additional fact we need is that

$$\int d\Omega\gamma^i\gamma^j=\delta^{ij}\frac{4\pi}{3}. \tag{3.77}$$

So the first moment of the Boltzmann equation is

$$\frac{\partial(n_{\rm DM}v_{\rm DM}^j)}{\partial t}+4\frac{da/dt}{a}n_{\rm DM}v_{\rm DM}^j+\frac{n_{\rm DM}}{a}\frac{\partial\Psi}{\partial x^j}=0. \tag{3.78}$$

This equation has no zero order parts, since the velocity is a first order quantity. Therefore, we need extract only the first order terms, which allows us to set $n_{\rm DM}\to n_{\rm DM}^{(0)}$ everywhere. Using the time dependence we found in equation [3.72] we arrive at

$$\frac{\partial v_{\rm DM}^j}{\partial t}+\frac{da/dt}{a}v_{\rm DM}^j+\frac{1}{a}\frac{\partial\Psi}{\partial x^j}=0. \tag{3.79}$$

Equations 3.74 and 3.79 are the two equations governing the evolution of the density and the velocity of the cold, dark matter. The momentum conservation equation, 3.79 does not have the standard $(\vec{v}\cdot\nabla)\vec{v}$ term, since any term with two factors of v is manifestly second order. An interesting feature of the two equations is generic to this process of integrating the Boltzmann equations to get the fluid equations. Note that the equation for the density depends on the next highest moment, the velocity. This is general: the integrated Boltzmann equation for the $l^{\rm th}$ moment depends on the $l+1$ moment. In principle, then, this process of integrating leads to an infinite heirarchy of equations for the moments of the distribution function. Indeed, we will see that this is one way of solving the Boltzmann equation for the photons, equation [3.59], which we have not yet integrated over. One might expect, then, that the velocity equation would depend on the next highest moment, the quadrupole, of the dark matter distribution. Why doesn't it? The answer lies in our assumption that the dark matter is *cold*. We have explicitly dropped all terms of order $(p/E)^2$ and higher. These terms correspond to the higher moments of the distribution, but since we are dealing with cold, dark matter they are irrelevant. Thus, the set of two equations, 3.74 and 3.79, are a closed set of equations for the cold,dark matter distribution[§]. If we were interested in dark matter particles with much smaller masses, such as massive neutrinos, we would need to keep these higher moments. For now we stick with the simpler set of equations for cold, dark matter.

Let us finally rewrite equations (3.74) and (3.79) in terms of conformal time η and the Fourier transforms. The density equation becomes

$$\dot{\delta}_{\rm DM}+ikv_{\rm DM}+3\dot{\Phi}=0 \tag{3.80}$$

where I have assumed that the velocity is irrotational so $v_{\rm DM}^i=\hat{k}^i v_{\rm DM}$. The velocity equation is

$$\dot{v}_{\rm DM}+\frac{\dot{a}}{a}v_{\rm DM}+ik\Psi=0. \tag{3.81}$$

3.6 Boltzmann Equation for Baryons

The final components of the universe which require a set of Boltzmann equations are the electrons and protons. These components are often grouped together and called *baryons*, nomenclature which is obviously ridiculous (electrons are leptons, not baryons) but nonetheless common.

The fundamental fact about electrons and protons is that they are very tightly coupled by Coulomb scattering. The Coulomb scattering rate is much larger than the expansion rate at all epochs of interest (see Problem (2.9)). This tight coupling forces the electron and proton overdensities to a common value:

$$\frac{\rho_e-\rho_e^{(0)}}{\rho_e^{(0)}}=\frac{\rho_p-\rho_p^{(0)}}{\rho_p^{(0)}}\equiv\delta_B \tag{3.82}$$

[§] Of course, we still need equations for the gravitational potentials Φ and Ψ. These come from Einstein's equations, as does the zero order equation for a.

where we bow to common usage with the subscript B. Similarly the velocities of the two species are forced to a common value,

$$\vec{v}_e = \vec{v}_p \equiv \vec{v}_B. \tag{3.83}$$

We need to derive equations then for δ_B and $\vec{v}_B$. The starting point will be the unintegrated equations for electrons and protons:

$$\frac{df_e(\vec{x},\vec{q},t)}{dt} = < c_{ep} >_{QQ'q'} + < c_{e\gamma} >_{pp'q'} \tag{3.84}$$

$$\frac{df_p(\vec{x},\vec{Q},t)}{dt} = < c_{ep} >_{qq'Q'} . \tag{3.85}$$

The notation here is more compact, and therefore more deceiving, than that in previous sections. We will need this compactness in what follows, so let's walk through it slowly. First, notice that initial and final momenta for the photon are $\vec{p}$ and $\vec{p}'$; for electron $\vec{q}$ and $\vec{q}'$; and the proton has been assigned $\vec{Q}$ and $\vec{Q}'$. Consider the Compton collision term in the equation for the electron distribution function. I have defined the unintegrated part of the collision term as

$$c_{e\gamma} \equiv (2\pi)^4\delta^4(p+q-p'-q')\frac{|\mathcal{M}|^2}{8E(p)E(p')E(q)E(q')}\{f_e(q')f_\gamma(p') - f_e(q)f_\gamma(p)\} \tag{3.86}$$

and the angular brackets denote integration over all momenta in the subscripts:

$$< (\ldots) >_{pp'q'} \equiv \int \frac{d^3p}{(2\pi)^3}\int\frac{d^3p'}{(2\pi)^3}\int\frac{d^3q'}{(2\pi)^3}(\ldots). \tag{3.87}$$

The Coulomb collision term is similar, the main difference being the amplitude for the two processes.

In principle, equation [3.85] should contain a term accounting for scattering of protons off photons. In practice, though, the cross section for this process is much smaller than for Compton scattering off electrons (in each case the cross section is inversely proportional to the mass squared). So the interactions of the combined electron-proton fluid with the photons is driven by Compton scattering of electrons, and the proton-photon process can be ignored.

With this notation defined, we can now proceed and derive equations for δ_B and $\vec{v}_B$. First, multiply both sides of equation [3.84] by the phase space volume $d^3q/(2\pi)^3$ and integrate. The left hand side then becomes identical to the left hand side we derived for dark matter in equation [3.70]. So we can immediately write

$$\frac{\partial n_{\rm B}}{\partial t} + \frac{ikn_{\rm B}v_{\rm B}}{a} + 3\left[\frac{da/dt}{a} + \frac{\partial\Phi}{\partial t}\right]n_{\rm B} = < c_{ep} >_{QQ'q'q} + < c_{e\gamma} >_{pp'q'q} \tag{3.88}$$

where the second term accounts for the fact that $v_{\rm B}^i = \hat{k}^i v_{\rm B}$. Both terms on the right vanish. The mathematical way to see this is to realize that the integration measure is completely

symmetric under the interchange of $p \leftrightarrow p'$ and $q \leftrightarrow q'$. Because of the factors of the distribution function, the integrands – in both c_{ep} and $c_{e\gamma}$ – are anti-symmetric under this interchange. So the full integral vanishes. More intuitively, the processes we are considering conserve electron number so they can certainly not contribute to dn/dt. That is, the integral over $f_e(q')f_p(Q')$ counts the total number of electrons that are produced in Coulomb scattering. But this is obviously equal to the integral over $f_e(q')f_p(Q')$, which counts the number of electrons lost in Coulomb scattering. More generally, any time we multiply an unintegrated collision term by a conserved quantity and then integrate we will get zero.

The equation for the zero order electron density is therefore equivalent to that of dark matter particles, $n_e^{(0)} \propto a^{-3}$. The perturbation equation is also identical to equation [3.74]:

$$\dot{\delta}_B + ikv_{\rm B} + 3\dot{\Phi} = 0 \tag{3.89}$$

written now with conformal time as the variable.

The final equation for the baryons is obtained by taking the first moments of both equation [3.84] and equation [3.85] and then adding the two integrated equations. Here though we will take the moments by first multiplying the unintegrated equations by $\vec{q}$ (and $\vec{Q}$ for the protons) instead of by $\vec{q}/E$ as we did for the dark matter. Therefore, our results from the dark matter case carry over as long as we multiply them by a factor of m. Since the proton mass is so much larger than the electron mass, the sum of the two left hand sides will be dominated by the protons. So, following equation [3.78], we have

$$\begin{aligned} m_p\frac{\partial(n_{\rm B}v_{\rm B}^j)}{\partial t} &+ 4\frac{da/dt}{a}m_pn_{\rm B}v_{\rm B}^j + \frac{m_pn_{\rm B}ik^j}{a}\Psi \\ &= < c_{ep}(\vec{q}+\vec{Q}) >_{QQ'q'q} + < c_{e\gamma}\vec{q} >_{pp'q'q} . \end{aligned} \tag{3.90}$$

Once again we can use a conservation law, this time conservation of momentum, to argue that the integral of $c_{ep}(\vec{q}+\vec{Q})$ over all momenta vanishes. By the same token, an easy way to evaluate the Compton integral is to note that

$$< c_{e\gamma}\vec{q} >_{pp'q'q} = - < c_{e\gamma}\vec{p} >_{pp'q'q} . \tag{3.91}$$

The first order equation for the baryon velocity is therefore

$$\frac{\partial v_{\rm B}}{\partial t} + \frac{da/dt}{a}v_{\rm B} + \frac{ik}{a}\Psi = -\frac{1}{\rho_B} < c_{e\gamma}p\mu >_{pp'q'q}, \tag{3.92}$$

where the baryon density is $\rho_B = n_{\rm B}m_p$ and we have multiplied by $\hat{k}^j$. We have already computed $< c_{e\gamma} >_{p'q'q}$ in equation [3.52]. We need simply multiply this by $p\mu$ and integrate over all p to find the right hand side of equation [3.92]. Plugging in, we have

$$\begin{aligned} -\frac{< c_{e\gamma}p\mu >_{pp'q'q}}{\rho_B} &= \frac{n_e\sigma_T}{\rho_B}\int\frac{d^3p}{(2\pi)^3}p^2\frac{\partial f^{(0)}}{\partial p}\mu[\Theta_0 - \Theta(\mu) + v_{\rm B}\mu] \\ &= \frac{n_e\sigma_T}{\rho_B}\int_0^\infty\frac{dp}{2\pi^2}p^4\frac{\partial f^{(0)}}{\partial p}\int_{-1}^1\frac{d\mu}{2}\mu[\Theta_0 - \Theta(\mu) + v_{\rm B}\mu] . \end{aligned} \tag{3.93}$$

The integral over p can be done by integrating by parts: it is $-4\rho_\gamma$. The $\mu-$ integration over the first and third terms is straightforward (first term vanishes and second gives $v_{\rm B}/3$). The second term is the first moment of the perturbation Θ. Recall that the zeroth moment was defined as Θ_0. It makes sense therefore to define the first moment as

$$\Theta_1 \equiv i\int_{-1}^{1}\frac{d\mu}{2}\mu\Theta(\mu) \tag{3.94}$$

where the factor of i is a convention.

We now have an expression for the collision term which can be inserted into equation [3.92], and after switching to conformal time, we have:

$$\dot{v}_{\rm B} + \frac{\dot{a}}{a}v_{\rm B} + ik\Psi = \dot{\tau}\frac{4\rho_\gamma}{3\rho_B}\left[3i\Theta_1 + v_{\rm B}\right]. \tag{3.95}$$

Until now, we have worked under the assumption that free electrons and protons are tightly coupled via Coulomb scattering. This is the origin of the factor of ρ_B in the denominator. Physically, it arises from the fact that moving electrons is much more difficult when they are coupled to protons. If the proton was infinitely heavy, so $\rho_B \to \infty$, Compton scattering would not change the electron velocity at all. One might argue that, since Coulomb scattering couples only free electrons and protons, the factor of ρ_B should include only ionized gas. In fact, though, even neutral hydrogen and helium are tightly coupled to electrons and protons (see Problem (7)), so all baryons should be included.

3.7 Summary

We now collect the equations we have derived for the photons, dark matter, and baryons and supplement them with a trivial extension to massless neutrinos.

$$\dot{\Theta} + ik\mu\Theta = -\dot{\Phi} - ik\mu\Psi - \dot{\tau}\left[\Theta_0 - \Theta + \mu v_{\rm B} - \frac{1}{2}P_2(\mu)\Pi\right] \tag{3.96}$$

$$\Pi = \Theta_2 + \Theta_{P2} + \Theta_{P0} \tag{3.97}$$

$$\dot{\Theta}_P + ik\mu\Theta_P = -\dot{\tau}\left[-\Theta_P + \frac{1}{2}(1 - P_2(\mu))\Pi\right] \tag{3.98}$$

$$\dot{\delta}_{\rm DM} + ikv_{\rm DM} = -3\dot{\Phi} \tag{3.99}$$

$$\dot{v}_{\rm DM} + \frac{\dot{a}}{a}v_{\rm DM} = -ik\Psi \tag{3.100}$$

$$\dot{\delta}_{\rm B} + ikv_{\rm B} = -3\dot{\Phi} \tag{3.101}$$

$$\dot{v}_{\rm B} + \frac{\dot{a}}{a}v_{\rm B} = -ik\Psi + \frac{\dot{\tau}}{R}\left[v_{\rm B} + 3i\Theta_1\right] \tag{3.102}$$

$$\dot{\Theta}_\nu + ik\mu\Theta_\nu = -\dot{\Phi} - ik\mu\Psi. \tag{3.103}$$

Equation 3.96 is the Boltzmann equation for photons we have derived. The one change from our derivation is the last term $(1/10)P_2\Pi$, which requires some explanation. First, note that it is proportional to the second Legendre polynomial, $P_2(\mu) = (3\mu^2 - 1)/2$. This term accounts for the angular dependence of Compton scattering and polarization, which we explicitly ignored earlier. The quantity Π here depends on the quadrupole moment of the temperature field. In general, it will be useful to define the lth multipole moment of the temperature field as

$$\Theta_l \equiv \frac{1}{(-i)^l}\int_{-1}^{1}\frac{d\mu}{2}P_l(\mu)\Theta(\mu). \tag{3.104}$$

The quadrupole corresponds to $l = 2$, and we can now recognize the earlier definitions of the monopole (equation [3.50]) and dipole (equation [3.94]) as the $l = 0$ and $l = 1$ cases. The higher Legendre polynomials have structure on smaller scales (see figure 3.4), so the higher moments capture information about the small scale structure of the temperature field.

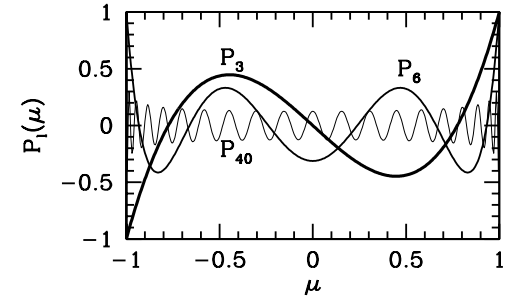


Figure 3.4: Some Legendre polynomials. Note that the higher order ones vary on smaller scales than do the low order ones. In general P_l crosses zero l times between -1 and 1.

Going back to equation [3.96], we see that the temperature field is also coupled to the polarization field Θ_P which obeys equation [3.98]. Note that Θ_P is sourced by the quadrupole, Θ_2, and none of the other temperature moments.

In the equation for the baryon velocity, [3.102], the ratio of photon to baryon density has been defined as

$$\frac{1}{R} \equiv \frac{4\rho_\gamma^{(0)}}{3\rho_B^{(0)}}. \tag{3.105}$$

The neutrino equation, 3.103, is identical to the photon equation, except of course there is no scattering term. Here I have assumed that the neutrinos are massless. If any of the neutrinos had appreciable mass, then equation [3.103] would have to be amended to account for this. Problem (6) discusses the question of how large a mass is interesting.

Suggested Reading

In the 1960's a national magazine ran a cartoon showing dozens of business- men and women walking the streets of Manhattan looking very important and serious. Thought bubbles over each head revealed their true focus: each was imagining a raucus sex scene. In at least some ways, the Boltzmann equation plays a similar role for physicists and astronomers: noone ever talks about it, but everyone is always thinking about it.

Two excellent astronomy textbooks which do make abundant use of the Boltzmann equation – either explicitly or implicitly – are *Radiative Processes in Astrophysics* (Rybicki and Lightman) and *Galactic Dynamics* (Binney and Tremaine). In the context of cosmology, in addition to the books mentioned in Chapter 1, *The Large Scale Structure of the Universe* (Peebles), written by the field's pioneer, uses the Boltzmann equation extensively, working in synchronous gauge. If you struggled through §3.3, you will be amused (angered?) to see §92 of Peebles' book, where he takes much less space to derive terms due to Compton scattering.

A number of papers deriving the Boltzmann equation for cosmological perturbations are well-worth reading. There is the path-breaking work by Lifshitz (1946), Peebles and Yu (1970), and Bond and Szalay (1983). A nice review was written by Efstathiou (1990). The treatment of Compton scattering presented here is based on Dodelson and Jubas (1995). If you were to read just one paper in this area, I would recommend Ma and Bertschinger (1995) which skips many of the steps presented here but has all the relevant formulae and the added virtue of equations in both conformal Newtonian and synchronous gauges. For derivation of the polarization terms in the Boltzmann equations, see Kosowsky (1996). The first paper to present the Boltzmann equation for tensors was Crittenden et al. (1993).

We will not spend too much time in this book on different gauges or on the decomposition of perturbations into scalar, vector, and tensor parts. Two excellent review articles which discuss both of these topics in detail are Mukhanov, Feldman, and Brandenberger (1992) and Kodama and Sasaki (1984). Both of these are also very good on the subjects of the next two chapters, the perturbed Einstein equations and inflation.

Problems

3.1 Derive the fluid equations for the collisionless, one-dimensional harmonic oscillator by taking the moments of equation [3.6]. The relevant quantities are the number density and the velocity defined as integrals over the distribution function:

$$n \equiv \int_{-\infty}^{\infty} \frac{dp}{2\pi} f \qquad ; \qquad v \equiv \frac{1}{n} \int_{-\infty}^{\infty} \frac{dp}{2\pi} \frac{p}{m} f. \tag{3.106}$$

3.2 The metric in a synchronous gauge is

$$\begin{aligned} g_{00}(\vec{x},t) &= -1 \\ g_{0i}(\vec{x},t) &= 0 \\ g_{ij}(\vec{x},t) &= a^2 \left[\delta_{ij} + h_{ij}\right], \end{aligned} \tag{3.107}$$

with perturbations

$$h = \begin{pmatrix} -2\bar{\eta} & 0 & 0 \\ 0 & -2\bar{\eta} & 0 \\ 0 & 0 & h + 4\bar{\eta} \end{pmatrix} \tag{3.108}$$

where $\bar{\eta}$ has nothing to do with conformal time. Here I have chosen the wave vector $\vec{k}$ to lie in the $\hat{z}$ direction. Derive the equivalent of equation [3.59] in synchronous gauge:

$$\dot{\Theta} + ik\mu\Theta - \frac{\mu^2 \dot{h}}{2} - P_2(\mu)\dot{\bar{\eta}} = -\dot{\tau}\left[\Theta_0 - \Theta + \mu v\right]. \tag{3.109}$$

3.3 Show that the Pauli blocking factor $1 - f_e$ can be set to one for all epochs of interest. Use equation [2.54] to find f_e as a function of temperature and show that as long as the temperature is less than m_e (if not, equation [2.54] does not apply; why not?) f_e is always much less than one.

3.4 Derive equation [3.64], the unintegrated Boltzmann equation for a massive particle.

3.5 Account for the angular dependence of Compton scattering. Start from equation [3.47] but instead of assuming the amplitude is constant, take

$$|\mathcal{M}|^2 = 6\pi\sigma_T m_e^2 (1 + \cos^2[\hat{p} \cdot \hat{p}\,']).$$

Show that correctly accounting for the angular dependence introduces the factor of $(1/2)P_2(\mu)\Theta_2$ presented in equation [3.96].

3.6 Consider effect of a massive neutrino on the evolution equations.

(a) Take the zeroth moment of the unintegrated equation [3.64]. The zeroth moment is important because it feeds into the Einstein equation (acts as a source for metric perturbations) we will derive in the next chapter.

(b) Recent experiments measuring the atmospheric neutrino flux suggest that the mass of the tau neutrino is .07 eV, far larger than either the electron or muon neutrino. Find the contribution of a 0.07 eV neutrino to Ω today. You may assume it is non-relativistic.

(c) Consider the following two scenarios. Each has $\Omega = 1$ divided up among only two components: a cold, dark matter particle and a neutrino. The neutrino in each case has the standard abundance and temperature. The only difference between the two scenarios is in one the neutrino is massless while in the other it has a mass of 0.07 eV. Plot the energy density as a function of scale factor in each of the these scenarios. Note that they should agree very early on (in each case there is only a relativistic neutrino early on) and very late. The only difference comes in the middle.

3.7 Show that ordinary matter is tightly coupled during the relevant epochs in the early universe.

(a) Compute the ratio of the Coulomb scattering rate over the Hubble rate. You may assume that all electrons and protons are ionized.

(b) Show that the rate for neutral hydrogen to scatter off ionized protons is always much larger than the expansion rate even when the ionization fraction is on the order of 10^{-4}.

3.8 Consider tensor perturbations to the metric. These do not perturb $g_{00}(= -1)$ or $g_{0i}(= 0)$. However, the spatial part of the metric is now

$$g_{ij} = a^2 \begin{pmatrix} 1+h_+ & h_\times & 0 \\ h_\times & 1-h_+ & 0 \\ 0 & 0 & 1 \end{pmatrix}.$$

Derive the equation for the photon distribution function in the presence of tensor perturbations. Unlike scalar perturbations, tensor perturbations induce an azimuthal dependence in Θ_l, so decompose the anisotropy due to tensors into

$$\Theta^T(k,\mu,\phi) = \Theta^T_+(k,\mu)(1-\mu^2)\cos(2\phi) + \Theta^T_\times(k,\mu)(1-\mu^2)\sin(2\phi). \tag{3.110}$$

Show that both the $+$ and the $\times$ component satisfy

$$\frac{d\Theta^T_i}{d\eta} + ik\mu\Theta^T_i + \frac{1}{2}\frac{dh_i}{d\eta} = \dot{\tau}\left[\Theta^T_i - \frac{1}{10}\Theta^T_{i,0} - \frac{1}{7}\Theta^T_{i,2} - \frac{3}{70}\Theta^T_{i,4}\right] \tag{3.111}$$

where i stands for either $\times$ or $+$, and the moments are defined as were the scalar moments, in equation [3.104].

Chapter 4

Einstein Equations

The previous chapter set up the formalism to describe how perturbations in the gravitational field affect particle distributions. This formalism led to the set of the equations, [3.96] -[3.103]. We need to supplement these equations with an account of how the perturbations to the particle distributions affect the gravitational field. For this, we need Einstein's equations of general relativity. The calculation detailed in this chapter expands Einstein's equations perturbatively around the zero order homogeneous solution. Far from being subtle or complex as one might expect from general relativity's reputation, this calculation is completely straightforward, although a bit long. Still, working through it is a "must- do-once" exercise, so the steps are presented in some detail.

4.1 The Perturbed Ricci Tensor and Scalar

The fundamental equation of general relativity, [2.17], is a tensor equation, so in principle it represents sixteen separate equations. However, since both sides of the equation are symmetric tensors, only ten of these are distinct. We are interested though in only two of the ten, since the form of the metric we are focusing on has only two independent functions, Φ and Ψ.

Evaluating the left hand side of the Einstein equation requires three pre-steps:

- Compute the Christoffel symbols, $\Gamma^{\mu}_{\alpha\beta}$, for the perturbed metric of equation [3.9].
- From these, form the Ricci tensor, $R_{\mu\nu}$, using equation [2.18].
- Contract the Ricci tensor to form the Ricci scalar, $R \equiv g^{\mu\nu}R_{\mu\nu}$.

Note that, unfortunately, even if we are interested in only several components of the Einstein equation, we need to compute all the elements of the Ricci tensor. For, the Ricci scalar, which depends on all elements of the tensor, enters all components of Einstein's equation.

4.1.1 Christoffel Symbols

We have already computed the zero order Christoffel symbols in equations (2.9) and (2.10). Now we need to look at the first order terms, those that are linear in Φ and/or Ψ. First let us consider $\Gamma^0_{\mu\nu}$, which by definition is

$$\Gamma^0_{\mu\nu} = \frac{1}{2}g^{0\alpha}\left[g_{\alpha\mu,\nu} + g_{\alpha\nu,\mu} - g_{\mu\nu,\alpha}\right] \tag{4.1}$$

where again $_{,\alpha}$ means the derivative with respect to x^{α}. The only non-zero component of $g^{0\alpha}$ is the time component*, which is the inverse of $g_{00} = -1 - 2\Psi$. So, to linear order, $g^{00} = -1 + 2\Psi$. Therefore,

$$\Gamma^0_{\mu\nu} = \frac{-1+2\Psi}{2}\left[g_{0\mu,\nu} + g_{0\nu,\mu} - g_{\mu\nu,0}\right]. \tag{4.2}$$

Take each component in turn; first the $\mu = \nu = 0$ component. Each of the terms in square brackets is identical, so the brackets give $g_{00,0} = -2\Psi_{,0}$. Since we are interested only in first order terms the factor of 2Ψ out in front can be dropped and we are left with

$$\Gamma^0_{00} = \Psi_{,0}. \tag{4.3}$$

The next possibility is that one of the indices μ or ν is spatial and the other time. It doesn't matter which, since the Christoffel symbol is symmetric in its lower indices. In this case, only one of terms in brackets is non-zero, $g_{00,i} = -2\Psi_{,i}$. Once again since this is first order, we can drop the factor of 2Ψ in front, leading to

$$\Gamma^0_{0i} = \Gamma^0_{i0} = \Psi_{,i}. \tag{4.4}$$

Finally, if both lower indices are spatial, only the last term in brackets in equation [4.2] contributes, and

$$\Gamma^0_{ij} = \frac{1-2\Psi}{2}\frac{\partial}{\partial t}\left[\delta_{ij}a^2(1+2\Phi)\right]. \tag{4.5}$$

There is a zero order term here, the one we computed in equation [2.9], and three first order terms:

$$\Gamma^0_{ij} = \delta_{ij}a^2\left[\frac{da/dt}{a} + 2\frac{da/dt}{a}(\Phi - \Psi) + \Phi_{,0}\right]. \tag{4.6}$$

Computing the Christoffel symbols, $\Gamma^i_{\mu\nu}$, will be left as an exercise. They are:

$$\begin{aligned}
\Gamma^i_{00} &= \frac{1}{a^2}\Psi_{,i} \\
\Gamma^i_{j0} = \Gamma^i_{0j} &= \delta_{ij}\left(\frac{da/dt}{a} + \Phi_{,0}\right) \\
\Gamma^i_{jk} &= \delta_{ij}\Phi_{,k} + \delta_{ik}\Phi_{,j} - \delta_{jk}\Phi_{,i}.
\end{aligned} \tag{4.7}$$

Note that the only non-vanishing zero order component is Γ^i_{j0}, in agreement with equation [2.10].

*We will do the calculation with $x^0 = t$, not conformal time. Working with the latter introduces a factor of a^2 into g_{00}.

4.1.2 Ricci Tensor

The Ricci tensor is most easily expressed in terms of the Christoffel symbols, as in equation [2.18]. First, consider the time-time component:

$$R_{00} = \Gamma^{\alpha}_{00,\alpha} - \Gamma^{\alpha}_{0\alpha,0} + \Gamma^{\alpha}_{\beta\alpha}\Gamma^{\beta}_{00} - \Gamma^{\alpha}_{\beta 0}\Gamma^{\beta}_{0\alpha}. \tag{4.8}$$

All of these terms contribute at first order. One simplification comes from considering the $\alpha = 0$ part of all these terms. The first and second terms are equal and opposite as are the last two. So the sum over the index α contributes only when α is spatial. Let's consider each of the terms one by one.

- The first is
$$\Gamma^{i}_{00,i} = \frac{1}{a^2}\Psi_{,ii}, \tag{4.9}$$
using the first of equations [4.7].

- The second term in equation [4.8] is
$$\Gamma^{i}_{0i,0} = 3\left(\frac{d^2a/dt^2}{a} - (\frac{da/dt}{a})^2 + \Phi_{,00}\right) \tag{4.10}$$
using the second of equations [4.7]. The factor of three in front, comes from the implicit sum in δ_{ii}.

- The next term is $\Gamma^{i}_{i\beta}\Gamma^{\beta}_{00}$. Note that Γ^{β}_{00} is first order no matter what β is, so we need keep only the zero order part of $\Gamma^{i}_{i\beta}$. However, the last of equations [4.7] shows that $\Gamma^{i}_{i\beta}$ is first order unless $\beta = 0$. So to first order,
$$\begin{aligned}\Gamma^{i}_{i\beta}\Gamma^{\beta}_{00} &= \Gamma^{i}_{i0}\Gamma^{0}_{00} \\ &= 3\frac{da/dt}{a}\Psi_{,0}\end{aligned} \tag{4.11}$$

- Finally the last term is $\Gamma^{i}_{\beta 0}\Gamma^{\beta}_{0i}$. In this case, if $\beta = 0$ both Γ's are first order, so their product is second order and can be neglected. Therefore, only spatial β need be considered, leading to
$$\begin{aligned}\Gamma^{i}_{\beta 0}\Gamma^{\beta}_{0i} &= \Gamma^{i}_{j0}\Gamma^{j}_{0i} \\ &= 3\left((\frac{da/dt}{a})^2 + 2\frac{da/dt}{a}\Phi_{,0}\right).\end{aligned} \tag{4.12}$$

Collecting these four sets of terms with the correct signs gives

$$R_{00} = -3\frac{d^2a/dt^2}{a} + \frac{1}{a^2}\Psi_{,ii} - 3\Phi_{,00} + 3\frac{da/dt}{a}(\Psi_{,0} - 2\Phi_{,0}). \tag{4.13}$$

Note that the zero order term agrees with equation [2.21].

The space-space part of the Ricci tensor is left as an exercise. It is

$$\begin{aligned}R_{ij} &= \delta_{ij}\Big[\left(2(da/dt)^2 + a\frac{d^2a}{dt^2}\right)(1 + 2\Phi - 2\Psi) \\ &\quad + a(da/dt)\,(6\Phi_{,0} - \Psi_{,0}) + a^2\Phi_{,00} - \Phi_{,kk}\Big] - \Phi_{,ij} - \Psi_{,ij}.\end{aligned} \tag{4.14}$$

We can now contract the indices on the Ricci tensor and find the Ricci scalar.

$$\begin{aligned}R &\equiv g^{\mu\nu}R_{\mu\nu} = g^{00}R_{00} + g^{ij}R_{ij} \\ &= [-1 + 2\Psi]\left[-3\frac{d^2a/dt^2}{a} + \frac{1}{a^2}\Psi_{,ii} - 3\Phi_{,00} + 3\frac{da/dt}{a}(\Psi_{,0} - 2\Phi_{,0})\right] \\ &\quad + \left[\frac{1 - 2\Phi}{a^2}\right]\Big[3\Big\{\left(2(da/dt)^2 + a\frac{d^2a}{dt^2}\right)(1 + 2\Phi - 2\Psi) \\ &\quad + a(da/dt)\,(6\Phi_{,0} - \Psi_{,0}) + a^2\Phi_{,00} - \Phi_{,ii}\Big\} - \Phi_{,ii} - \Psi_{,ii}\Big].\end{aligned} \tag{4.15}$$

First let us check the zero order part of R. Combining terms, we find that it is $6(\frac{da/dt}{a}^2 + \frac{d^2a/dt^2}{a})$, in agreement with equation [2.24]. To get the first order part, $R^{(1)}$, we go through the by-now familiar routine of multiplying terms, keeping only those first order in Φ and Ψ. This gives

$$\begin{aligned}R^{(1)} &= -6\Psi\frac{d^2a/dt^2}{a} - \frac{1}{a^2}\Psi_{,ii} + 3\Phi_{,00} - 3\frac{da/dt}{a}(\Psi_{,0} - 2\Phi_{,0}) \\ &- 6\Psi\left(2(\frac{da/dt}{a})^2 + \frac{d^2a/dt^2}{a}\right) + 3\frac{da/dt}{a}\,(6\Phi_{,0} - \Psi_{,0}) \\ &+ 3\Phi_{,00} - 4\frac{\Phi_{,ii}}{a^2} - \frac{\Psi_{,ii}}{a^2},\end{aligned} \tag{4.16}$$

where the first line contains the terms from R_{00} (the second line in equation [4.15]) and the last two from R_{ij} (the last two lines in equation [4.15]). Combining these leads to

$$\begin{aligned}R^{(1)} &= -12\Psi\left((\frac{da/dt}{a})^2 + \frac{d^2a/dt^2}{a}\right) - \frac{2}{a^2}\Psi_{,ii} + 6\Phi_{,00} \\ &- 6\frac{da/dt}{a}(\Psi_{,0} - 4\Phi_{,0}) - 4\frac{\Phi_{,ii}}{a^2}.\end{aligned} \tag{4.17}$$

4.2 Two Components of Einstein's Equations

We can now derive the evolution equations for Φ and Ψ, the perturbations to the Friedmann–Robertson–Walker metric. There is some freedom here because Einstein's equations have

ten components and we need only two. All the eight other components will either be zero at first order or be redundant.

The first component we will use is the time-time component. Thus we need to evaluate

$$\begin{aligned} G^0_0 &= g^{00}\left[R_{00} - \frac{1}{2}g_{00}R\right] \\ &= (-1+2\Psi)R_{00} - \frac{R}{2}. \end{aligned} \tag{4.18}$$

Here one of the indices has been raised by multiplying G_{00} by g^{00} (recall that g^{0i} vanish). This turns out to simplify the energy-momentum tensor (see Problem (3)) which sources the Einstein tensor. Also note that the second line follows from the first since $g^{00}g_{00} = 1$. We have computed the time-time component of the Ricci tensor and the Ricci scalar, so we can write down

$$\begin{aligned} G^0_0 &= -3(\frac{da/dt}{a})^2 - 6\Psi\frac{d^2a/dt^2}{a} - \frac{1}{a^2}\Psi_{,ii} + 3\Phi_{,00} - 3\frac{da/dt}{a}(\Psi_{,0} - 2\Phi_{,0}) \\ &\quad +6\Psi\left((\frac{da/dt}{a})^2 + \frac{d^2a/dt^2}{a}\right) + \frac{1}{a^2}\Psi_{,ii} - 3\Phi_{,00} \\ &\quad +3\frac{da/dt}{a}(\Psi_{,0} - 4\Phi_{,0}) + 2\frac{\Phi_{,ii}}{a^2}. \end{aligned} \tag{4.19}$$

We will very shortly see many cancellations, so the number of terms will drop dramatically, but already we notice that the zero order term is correct. Combining other terms leads to

$$G^0_0 = -3(\frac{da/dt}{a})^2 - 6\frac{da/dt}{a}\Phi_{,0} + 6\Psi(\frac{da/dt}{a})^2 + 2\frac{\Phi_{,ii}}{a^2}. \tag{4.20}$$

Einstein's equation equates G^0_0 with $8\pi G T^0_0$ where $T_{\mu\nu}$ is the stress energy tensor. To complete our derivation of the first evolution equation for Φ and Ψ, therefore, we need to compute the source term, T^0_0. Roughly speaking, T^0_0 is the energy density of all the particles in the universe, so schematically it is

$$T^0_0(\vec{x},t) = \sum_{\vec{p}} E(\vec{p})n(\vec{p},\vec{x},t) \tag{4.21}$$

where n is the number density of particles with momentum $\vec{p}$. We need to perform this sum over all species (photons, neutrinos, etc.) weighting by the number of spin states for each species g_α. Then turning the sum over momentum into an integral, we have

$$T^0_0(\vec{p},\vec{x},t) = -\sum_{\text{all species }\alpha} g_\alpha \int \frac{d^3p}{(2\pi)^3}\, E_\alpha(p) f_\alpha(\vec{p},\vec{x},t). \tag{4.22}$$

The minus sign here arises from the metric component g^{00} which multiplies T_{00} to give T^0_0. It is important to relate this expression to the perturbation variables we defined in Chapter 3 for the photons, neutrinos, dark matter, and baryons, since each contributes to the energy

momentum tensor. This is easiest for the dark matter and baryons. For we defined the right hand side as $-\rho_i^{(0)}(1+\delta_i)$ where i labels either dark matter or baryons. For photons and neutrinos, a little more care is required. Using the definition of Θ in equation [3.33], we have

$$T^0_{0,\gamma} = -2\int \frac{d^3p}{(2\pi)^3}\, p\left[f^{(0)} - p\frac{\partial f^{(0)}}{\partial p}\Theta\right]. \tag{4.23}$$

The first term here is just the zero order photon energy density, ρ_γ. To reduce the second term, we first do the angular integral, which picks out the monopole Θ_0 from Θ. Then, we integrate by parts. This changes the sign and introduces a factor of 4 since $\partial p^4/\partial p = 4p^3$. Then,

$$T^0_0 = -\rho_\gamma\left[1 + 4\Theta_0\right]. \tag{4.24}$$

The factor of four here is obvious in retrospect. The perturbation variable Θ is the fractional temperature change, while the energy momentum tensor is interested in the fractional energy change, $\delta\rho/\rho$. We should have expected that since $\rho \propto T^4$, $\delta\rho/\rho = 4\delta T/T$. In any event, it falls out of the algebra. I harp on it only to warn those who turn to the literature that authors are virtually split between those who define Θ as $\delta\rho/\rho$ and those who opt for the convention we use here.

Returning to Einstein's equation, we collect first order terms, divide both sides by two, and find that the time-time component is

$$-3\frac{da/dt}{a}\Phi_{,0} + 3\Psi(\frac{da/dt}{a})^2 + \frac{\Phi_{,ii}}{a^2} = -4\pi G\left[\rho_{\rm DM}\delta_{\rm DM} + \rho_B\delta_{\rm B} + 4\rho_\gamma\Theta_0 + 4\rho_\nu\Theta_{\nu,0}\right]. \tag{4.25}$$

It is again useful to write the equation in terms of conformal time and Fourier transformed variables. This introduces an extra factor of $1/a$ every time a time derivative appears and turns the spatial derivative $\Phi_{,ii}$ to $-k^2\Phi$, so

$$k^2\Phi + 3\frac{\dot{a}}{a}\left(\dot{\Phi} - \Psi\frac{\dot{a}}{a}\right) = 4\pi G a^2\left[\rho_{\rm DM}\delta_{\rm DM} + \rho_B\delta_{\rm B} + 4\rho_\gamma\Theta_0 + 4\rho_\nu\Theta_{\nu,0}\right]. \tag{4.26}$$

This is our first evolution equation for Φ and Ψ.

We now obtain a second evolution equation for Φ and Ψ. Rather than choosing just one of the spatial components of G_{ij}, we will consider the *longitudinal, traceless* part of G. This turns out to be the most illuminative combination. Traceless means we subtract out the sum G_{ii}, while the longitudinal component satisfies

$$\epsilon_{ijk}G^{||}_{kl,jl} = 0, \tag{4.27}$$

where ϵ_{ijk} is an anti-symmetric tensor with properties $\epsilon_{123} = +1, \epsilon_{132} = -1, \epsilon_{112} = 0, \ldots$

To extract the longitudinal, traceless component of G, it is simplest to move to Fourier space and multiply G by $\hat{k}_i\hat{k}_j - (1/3)\delta_{ij}$, which is a *projection* operator. That is, it picks out the piece which is longitudinal, traceless and only that part. Note that any part of G which is proportional to δ_{ij} vanishes when acted upon in this fashion since

$$\delta_{ij}\left(\hat{k}_i\hat{k}_j - (1/3)\delta_{ij}\right) = 0. \tag{4.28}$$

This greatly simplifies our work, since we see from equation [4.14] that the only part of G which is not proportional to δ_{ij} is $-\Phi_{,ij} - \Psi_{,ij} = k_i k_j(\Phi + \Psi)$. Thus,

$$\left(\hat{k}_i\hat{k}_j - (1/3)\delta_{ij}\right) G_{ij} = \frac{2}{3}k^2(\Phi + \Psi). \tag{4.29}$$

The longitudinal, traceless part of the energy-momentum tensor is extracted in the same fashion

$$\left(\hat{k}_i\hat{k}_j - (1/3)\delta_{ij}\right) T_{ij} = g_\alpha a^2 \int \frac{d^3p}{(2\pi)^3}\frac{p^2\mu^2 - (1/3)p^2}{E(p)} f. \tag{4.30}$$

The factor of a^2 here enters because T_{ij} is proportional to the integral of $P_i P_j/P^0$, and $P_i = g_{ij}P^j \propto a$ using equation [3.17]. We can immediately recognize the combination $\mu^2 - 1/3$ as proportional to the second Legendre polynomial, more precisely equal to $(2/3)P_2(\mu)$. Therefore, the integral picks out the quadrupole part of the distribution. Of course the zero order part of the distribution function has no quadrupole, so the source term is first order, proportional to Θ_2, which is non-zero only for neutrinos and photons. The integral in equation [4.30] for massless particles is

$$\begin{aligned} -2\int \frac{dpp^2}{2\pi^2}p^2\frac{\partial f^{(0)}}{\partial p}\int_{-1}^{1}\frac{d\mu}{2}\frac{2P_2(\mu)}{3}\Theta(\mu) &= 2\frac{2\Theta_2}{3}\int\frac{dpp^2}{2\pi^2}p^2\frac{\partial f^{(0)}}{\partial p} \\ &= -\frac{8\rho^{(0)}\Theta_2}{3} \end{aligned} \tag{4.31}$$

where the first equality follows from the definition of the quadrupole and second from an integration by parts.

As the second evolution equation, we therefore have

$$k^2(\Phi + \Psi) = -32\pi G a^2\left[\rho_\gamma\Theta_2 + \rho_\nu\Theta_{\nu,2}\right]. \tag{4.32}$$

That is, the two gravitational potentials are equal and opposite unless the photon or neutrino have appreciable quadrupole moments. In practice, the photon's quadrupole contributes little to this sum, because its moment is very small during the time when it has appreciable energy density. (Recall the argument after equation [3.52].) Only the collisionless neutrino has an appreciable moment early on when radiation dominates the universe.

4.3 Tensor Perturbations

Until now, we have focused almost exclusively on scalar perturbations to the homogeneous FRW universe. Practically this means we have taken the form of the metric to be that of equation [3.9], with only two functions, Φ and Ψ, specifying the perturbations. This choice is reasonable: as we have seen, scalar perturbations to the metric are sourced by density fluctuations and vise versa. For the most part, the density fluctuations that led to the structure of the universe are our primary interest.

Nonetheless, many theories of structure formation produce, in addition to scalar flutuations, *tensor* perturbations to the metric. These are potentially detectable because they

produce observable distortions in the CMB, especially on large scales. Sprinkled throughout the book, therefore, are exercises (with hints) relating to tensor perturbations. The tools needed to study these are precisely those we crafted when studying scalar perturbations. For the most part, therefore, I regard tensors as one rather large homework problem, which introduces no new physics.

One question which naturally arises when working out these exercises, though, is why consider scalar and tensor perturbations separately? To answer this question (and to alleviate the homework load) this section derives Einstein's equations for tensor perturbations. We will see that scalar and tensor perturbations *decouple*; that is, they evolve completely independently. So the presence of tensor perturbations does not affect the scalars and vise versa. Contrast this with Φ and Ψ. We have just shown that they are quite tightly coupled to each other. It is impossible to learn about Φ without also solving for Ψ. The decoupling of scalars and tensors is a manifestation of a fancy theorem, the *Decomposition Theorem*. Needless to say, it is much more instructive to work out an example of this theorem than to prove it abstractly. Also, as you would expect, the same theorem can be applied to *vector* perturbations. These too are produced by some early universe models (but not as ubiquitously as tensors) and can be treated completely independently.

Tensor perturbations can be characterized by a metric with $g_{00} = -1$, zero space-time components $g_{0i} = 0$, and spatial elements

$$g_{ij} = a^2\begin{pmatrix} 1 + h_+ & h_\times & 0 \\ h_\times & 1 - h_+ & 0 \\ 0 & 0 & 1 \end{pmatrix}. \tag{4.33}$$

That is, the perturbations to the metric are described by two functions, $h_+(\vec{x}, t)$ and $h_\times(\vec{x}, t)$, assumed small. For definiteness, I have chosen the perturbations to be in the $x - y$ axis. This corresponds to an implicit choice of axes; in particular, it corresponds to choosing the $z-$ axis to be in the direction of the wavevector, $\hat{k}$. Often this convention is not used, and the functions h_+ and $h_\times$ are taken as vectors in the two-dimensional subspace perpendicular to $\hat{k}$, spanned by two polarization vectors.

Once the metric in equation [4.33] has been written down, we can blast away and derive the Einstein equations. Once again the derivation proceeds in three steps: (i) Christoffel symobls. (ii) Ricci tensor, and (iii) Ricci scalar.

4.3.1 Christoffel Symbols for Tensor Perturbations

First consider $\Gamma^0_{\alpha\beta}$. The metric we are considering in equation [4.33] has constant g_{00} and vanishing g_{0i}. Recall that the Christoffel symbol is a sum of derivatives of the metric. The only terms that will be non-zero are those which involve derivatives of the spatial part of the metric, $g_{ij,\alpha}$. Therefore, we can immediately argue that

$$\Gamma^0_{00} = \Gamma^0_{i0} = 0. \tag{4.34}$$

The term with two lower spatial indices is

$$\Gamma^0_{ij} = -\frac{g^{00}}{2}g_{ij,0}$$

$$= \frac{1}{2} g_{ij,0}. \tag{4.35}$$

To simplify notation, let's introduce the first order matrix $\mathcal{H}$, defined via

$$\mathcal{H} \equiv \begin{pmatrix} h_+ & h_\times & 0 \\ h_\times & -h_+ & 0 \\ 0 & 0 & 0 \end{pmatrix} \tag{4.36}$$

so that $g_{ij} = a^2(\delta_{ij} + \mathcal{H}_{ij})$. Therefore,

$$g_{ij,0} = 2\frac{da/dt}{a} g_{ij} + a^2 \mathcal{H}_{ij,0} \tag{4.37}$$

and

$$\Gamma^0_{ij} = \frac{da/dt}{a} g_{ij} + \frac{a^2 \mathcal{H}_{ij,0}}{2}. \tag{4.38}$$

When both lower indices on Γ are 0, the Christoffel symbol vanishes. The two remaining components are Γ^i_{0j} and Γ^i_{jk}. The former is

$$\Gamma^i_{0j} = \frac{g^{ik}}{2} g_{jk,0}. \tag{4.39}$$

The time derivative of g_{jk} acts on both the scale factor and on the perturbations h, as in equation [4.37], so

$$\Gamma^i_{0j} = \frac{g^{ik}}{2}\left[2\frac{da/dt}{a} g_{jk} + a^2 \mathcal{H}_{jk,0}\right]. \tag{4.40}$$

But $g^{ik} g_{jk} = \delta_{ij}$, so the first term here is simply $(da/dt)\delta_{ij}/a$. To get the second, we can set $g^{jk} = \delta_{jk}/a^2$ (i.e. neglect first order terms) since it multplies the first order $\mathcal{H}$. So,

$$\Gamma^i_{0j} = \frac{da/dt}{a}\delta_{ij} + \frac{1}{2}\mathcal{H}_{ij,0}, \tag{4.41}$$

where I have used the fact that $\mathcal{H}_{ij}$ is symmetric.

The last Christoffel symbol we need is Γ^i_{jk}. In Problem (6) you will show that

$$\Gamma^i_{jk} = \frac{1}{2}\left[\mathcal{H}_{ij,k} + \mathcal{H}_{ik,j} - \mathcal{H}_{jk,i}\right]. \tag{4.42}$$

4.3.2 Ricci Tensor for Tensor Perturbations

Following the same steps as in the scalar perturbation case, we now combine these Christoffel symbols to form the Ricci tensor. First we compute the time-time component of the Ricci tensor.

$$R_{00} = \Gamma^\alpha_{00,\alpha} - \Gamma^\alpha_{0\alpha,0} + \Gamma^\alpha_{\beta\alpha}\Gamma^\beta_{00} - \Gamma^\alpha_{\beta 0}\Gamma^\beta_{0\alpha}. \tag{4.43}$$

We have shown that the Christoffel symbol vanishes for tensor perturbations when the two lower indices are time-time. Therefore, the first and third terms here are zero. Using the same argument, the indices α and β in the second and fourth terms must be spatial, so

$$R_{00} = -\Gamma^i_{0i,0} - \Gamma^i_{j0}\Gamma^j_{0i}. \tag{4.44}$$

Using equation [4.41] for Γ^i_{j0} which is the only element appearing, we find that

$$\begin{aligned} R_{00} &= -3\frac{\partial}{\partial t}\frac{da/dt}{a} - \frac{1}{2}\mathcal{H}_{ii,0} \\ &\quad - \left(\frac{da/dt}{a}\delta_{ij} + \frac{1}{2}\mathcal{H}_{ij,0}\right)\left(\frac{da/dt}{a}\delta_{ij} + \frac{1}{2}\mathcal{H}_{ij,0}\right). \end{aligned} \tag{4.45}$$

On the first line, the trace $\mathcal{H}_{ii}$ vanishes since h_+ appears in the metric with opposite signs along the diagonal. Expanding the second line out to first order leads to a similar cancellation: $\mathcal{H}_{ij}$ is multiplied by δ_{ij}, so there are no first order terms. The zero order terms combine to form

$$R_{00} = -3\frac{d^2a/dt^2}{a}, \tag{4.46}$$

an equation in which we are by now quite confident since this is the third time we have derived it (see equations [2.21] and [4.13]). Of course the big news here is not that we have correctly derived the zero order term, but rather that tensor perturbations do not appear at first order in R_{00}. Looking ahead, we will soon see that the Ricci scalar also has no tensor contribution (even though R_{ij} does). Therefore, we can anticipate that the time-time component of Einstein's equations contains no tensor perturbations. This is important for it tells us that density perturbations – which form the right hand side of the time-time component as shown in equation [4.25] – do *not* induce any tensor perturbations. We are beginning therefore to get a glimmer of the decomposition theorem. Density perturbations and scalar perturbations to the metric are coupled; indeed their names are often used as synonyms. Tensor perturbations however are decoupled from these and evolve on their own.

The spatial components of the Ricci tensor do depend on the tensor perturbation variables. We now turn to

$$R_{ij} = \Gamma^\alpha_{ij,\alpha} - \Gamma^\alpha_{i\alpha,j} + \Gamma^\alpha_{\alpha\beta}\Gamma^\beta_{ij} - \Gamma^\alpha_{\beta j}\Gamma^\beta_{i\alpha}. \tag{4.47}$$

Let's consider the first two terms together. Expanding out leads to

$$\Gamma^\alpha_{ij,\alpha} - \Gamma^\alpha_{i\alpha,j} = \Gamma^0_{ij,0} + \Gamma^k_{ij,k} - \Gamma^k_{ik,j} \tag{4.48}$$

since $\alpha = 0$ does not contribute in $\Gamma^\alpha_{i\alpha,j}$. The hardest (i.e. longest) term here is the first which involves multiple time derivatives. Let's postphone its calculation by recalling that $\Gamma^0_{ij} = g_{ij,0}/2$ so that the first term can be written in shorthand as $g_{ij,00}/2$. The last two terms in equation [4.48] partially cancel each other so that

$$\Gamma^\alpha_{ij,\alpha} - \Gamma^\alpha_{i\alpha,j} = \frac{g_{ij,00}}{2} + \frac{1}{2}\left[\mathcal{H}_{jl,il} + \mathcal{H}_{il,jl} - \mathcal{H}_{ji,ll}\right]. \tag{4.49}$$

In Fourier space, derivatives simply bring down factors of $\vec{k}$. Therefore, in Fourier space $\mathcal{H}_{il,jl} \to -k_j k_l \mathcal{H}_{il}$, the minus sign coming from the pair of i's. Recall that we chose $\hat{k}$ to be along the $z-$ axis. Therefore, the derivative indices j and l in $\mathcal{H}_{il,jl}$ must be equal to 3. But $\mathcal{H}_{i3} = 0$ so this term and its cousin $\mathcal{H}_{jl,il}$ must both vanish. Therefore,

$$\Gamma^\alpha_{ij,\alpha} - \Gamma^\alpha_{i\alpha,j} = \frac{g_{ij,00}}{2} - \frac{1}{2}\mathcal{H}_{ji,ll}. \tag{4.50}$$

The third term in equation [4.47], $\Gamma^\alpha_{\alpha\beta}\Gamma^\beta_{ij}$, is non-zero only when the index α is spatial, so

$$\Gamma^\alpha_{\alpha\beta}\Gamma^\beta_{ij} = \Gamma^k_{k0}\Gamma^0_{ij} + \Gamma^k_{kl}\Gamma^l_{ij}. \tag{4.51}$$

But each of the Christoffel symbols in the second term here are first order, so their product vanishes. In the first term, the sum over k makes the first order terms go away, so Γ^k_{k0} is purely zeroth order, $3\frac{da/dt}{a}$. Therefore,

$$\Gamma^\alpha_{\alpha\beta}\Gamma^\beta_{ij} = \frac{3}{2}\frac{da/dt}{a}g_{ij,0}. \tag{4.52}$$

The final term in equation [4.47] will be left as an exercise; it is

$$\Gamma^\alpha_{\beta j}\Gamma^\beta_{i\alpha} = 2\left(\frac{da/dt}{a}\right)^2 g_{ij} + 2a\frac{da}{dt}\mathcal{H}_{ij,0}. \tag{4.53}$$

We can now combine all four terms in equation [4.47] to get

$$\begin{aligned} R_{ij} &= \frac{g_{ij,00}}{2} - \frac{1}{2}\mathcal{H}_{ji,kk} + \frac{3}{2}\frac{da/dt}{a}g_{ij,0} \\ &- 2\left(\frac{da/dt}{a}\right)^2 g_{ij} - 2a\frac{da}{dt}\mathcal{H}_{ij,0}. \end{aligned} \tag{4.54}$$

We now need to expand out the time derivatives of the metric. Using equation [4.37], one finds

$$g_{ij,00} = 2g_{ij}\left(\frac{d^2a/dt^2}{a} + (\frac{da/dt}{a})^2\right) + 4a\frac{da}{dt}\mathcal{H}_{ij,0} + a^2\mathcal{H}_{ij,00}. \tag{4.55}$$

Therefore the Ricci tensor is

$$\begin{aligned} R_{ij} &= g_{ij}\left(\frac{d^2a/dt^2}{a} + 2(\frac{da/dt}{a})^2\right) + \frac{3}{2}a\frac{da}{dt}\mathcal{H}_{ij,0} \\ &+ a^2\frac{\mathcal{H}_{ij,00}}{2} - \frac{1}{2}\mathcal{H}_{ij,kk}. \end{aligned} \tag{4.56}$$

Again we see that we have successfully recaptured the zero order part of the Ricci tensor. Remarkably, we will see that the first order parts – when used in Einstein's equations – do not couple to the scalar perturbations.

First, though, we must compute the Ricci scalar:

$$R = g^{00}R_{00} + g^{ij}R_{ij}. \tag{4.57}$$

The time-time product is all zero order, so we can neglect it when considering the first order piece $R^{(1)}$. The space-space contraction has two types of terms. First, there are the terms in equation [4.56] proportional to the metric g_{ij}. These contract with g^{ij} to give a factor of four, leading to no first order terms. All the other terms in equation [4.56] are first order, so when contracting them we can set g^{ij} to its zero order value, δ_{ij}/a^2. This corresponds to taking the trace of the first order terms in equation [4.56]. Since all first order terms are proportional to $\mathcal{H}_{ij}$, the trace vanishes. Therefore, tensor perturbations do not affect (at first order) the Ricci scalar.

4.3.3 Einstein's Equations for Tensor Perturbations

We can now read off the perturbations to the Einstein Tensor induced by tensor modes. Since the Ricci scalar is unperturbed by tensor perturbations, the first order Einstein tensor is simply

$$G^{(1)}_{ij} = R^{(1)}_{ij}. \tag{4.58}$$

From equation [4.56], we have

$$G^{(1)}_{ij} = \frac{3}{2}a\frac{da}{dt}\mathcal{H}_{ij,0} + a^2\frac{\mathcal{H}_{ij,00}}{2} - \frac{1}{2}\mathcal{H}_{ij,kk}. \tag{4.59}$$

We can now derive a set of equations governing the evolution of the tensor variables, h_+ and $h_\times$.

To derive an equation for h_+, let us consider the difference between the $_{11}$ and $_{22}$ components of the Einstein tensor. Now $G^{(1)}_{ij}$ in equation [4.59] is proportional to $\mathcal{H}_{ij}$ and its derivatives. Since $\mathcal{H}_{11} = -\mathcal{H}_{22} = h_+$, $G^{(1)}_{11}$ is equal and opposite to $G^{(1)}_{22}$. Therefore,

$$G^{(1)}_{11} - G^{(1)}_{22} = 3a\frac{da}{dt}h_{+,0} + a^2h_{+,00} - h_{+,kk}. \tag{4.60}$$

Fourier transform this and change to conformal time (so that $h_{+,0} = \dot{h}_+/a$ and $h_{+,00} = \ddot{h}_+/a^2 - (\dot{a}/a^2)\dot{h}_+$). Then,

$$G^{(1)}_{11} - G^{(1)}_{22} = \ddot{h}_+ + 2\frac{\dot{a}}{a}\dot{h}_+ + k^2h_+. \tag{4.61}$$

The right hand side of this component of Einstein's equations is zero (see Problem 3.6), and $h_\times$ obeys the same equation (see Problem 3.7) so the tensor modes are governed by

$$\ddot{h}_\alpha + 2\frac{\dot{a}}{a}\dot{h}_\alpha + k^2h_\alpha = 0. \tag{4.62}$$

where $\alpha = +, \times$. Equation [4.62] is a wave equation. For example, if we neglect the expansion of the universe so that the damping term in equation [4.62] vanishes, we immediately see that the most general solution

$$h_\alpha(\vec{x},\eta) = \int d^3k e^{i\vec{k}\cdot\vec{x}}\left[Ae^{ik\eta} + Be^{-ik\eta}\right]. \tag{4.63}$$

The two modes here corresponds to waves travelling in the $\pm z-$ direction at the speed of light.

Equation [4.62] is a generalization of the wave equation to an expanding universe. Problem (11) illustrates that if the universe is purely radiation or matter dominated, exact analytic solutions can be obtained. These are oscillatory, like the simple ones in equation [4.63], but also damp out. Figure 4.1 shows the evolution of h_α for three different wavelength modes. The large scale mode (with $k\eta_0 = 10$ remains constant at early times when its wavelength is larger than the horizon $k\eta < 1$. Once its wavelength becomes comparable to the horizon, the amplitude begins to die off, oscillating several times until the present. The small scale

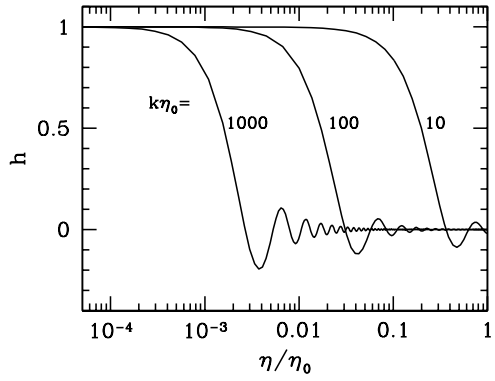


Figure 4.1: Evolution of gravity waves. Three different modes are shown, labeled by their wavenumbers. Smaller scale modes decay earlier.

mode $k\eta_0 = 1000$ shown in figure 4.1 also begins to decay when its wavelength becomes comparable to the horizon. Since its entry into the horizon ccurs much earlier though, the decay is much more efficient. By today, the amplitude is completely neglibile.

An important point about the effect of gravity waves on the CMB anisotropy spectrum can be gleaned from figure 4.1. Because smalle scale modes decay earlier than large scale modes, at decoupling (at $\eta/\eta_0 \simeq 0.02$) only modes with $k\eta_0$ greater than about 100 persist. All smaller scale modes can be neglected. Therefore, anisotropies on small angular scales will not be affected by gravity waves. Only the large scale anisotropies are impacted by gravity waves.

4.4 The Decomposition Theorem

The Decomposition Theorem states that perturbations to the metric can be divided up into three type: *scalar, vector,* and *tensor*. Each of these types perturbations evolves independently. That is, if some physical process in the early universe sets up tensor perturbations, say, these do not induce scalar perturbations. To determine the evolution of scalar perturbations, say, we will not have to worry about possible vector perturbations.

Now that we have computed the contributions to the Einstein tensor $G_{\mu\nu}$ from scalars and tensors, we can demonstrate the decomposition of these two types of perturbations. To do this, remember that we obtained the scalar equations by considering the two components of Einstein's tensor:

$$G^0_0 \quad ; \quad \left(\hat{k}_i\hat{k}_j - (1/3)\delta_{ij}\right) G_{ij}. \tag{4.64}$$

Inserting these components into Einstein's equations led to equations [4.26] and [4.32]. If

we can show that tensor perturbations do not contribute to these two components, then we will have convinced ourselves of at least part of the Decomposition Theorem, namely that the equations governing scalar equations are not affected by tensors.

It is easy to see that tensor perturbations do not contribute to G^0_0. For G^0_0 depends on R_{00} and R. But we have seen that both of these do not depend on h_+ or $h_\times$.

To get the second relevant component of G, multiply equation [4.47] by the projection operator:

$$\begin{aligned}\left(\hat{k}_i\hat{k}_j - (1/3)\delta_{ij}\right) G^{(1)}_{ij} &= (\delta_{i3}\delta_{j3} - (1/3)\delta_{ij}) \\ &\times \left[\frac{3}{2}a\frac{da}{dt}\mathcal{H}_{ij,0} + a^2\frac{\mathcal{H}_{ij,00}}{2} - \frac{1}{2}\mathcal{H}_{ij,kk}\right].\end{aligned} \tag{4.65}$$

where the equality holds since we have chosen $\hat{k}$ to lie in the $z-$ direction. The terms in which indices i and j are set to 3 vanish since $\mathcal{H}_{33} = 0$. The only remaining terms are those proportional to δ_{ij}. But the Kronicker delta instructs us to take the trace of $\mathcal{H}$. This too vanishes. The scalar equations we derived in the previous section are therefore unchanged by the presence of tensor modes. This is a manifestation of the Decomposition Theorem.

4.5 Summary

The Einstein equation relates perturbations in the metric to perturbations in the matter and radiation. Taking two components of the Einstein equation $G_{\mu\nu} = 8\pi G T_{\mu\nu}$, we found that the equations governing scalar perturbations Φ and Ψ are

$$\begin{aligned} k^2\Phi + 3\frac{\dot{a}}{a}\left(\dot{\Phi} - \Psi\frac{\dot{a}}{a}\right) &= 4\pi G a^2\left[\rho_{\rm DM}\delta_{\rm DM} + \rho_B\delta_{\rm B} + 4\rho_\gamma\Theta_0 + 4\rho_\nu\Theta_{\nu,0}\right] \quad ([4.26]) \\ k^2(\Phi + \Psi) &= -32\pi G a^2\left[\rho_\gamma\Theta_2 + \rho_\nu\Theta_{\nu,2}\right]. \quad ([4.32]) \end{aligned}$$

Some of the other components of Einstein's equation are redundant; they add no new information about the evolution of Φ and Ψ. An example is the time-space component, which you can derive in Problem (4). At times, though, one form of the evolution equation will be more useful than another. For example, one combination (Problem (5)) of these equations leads to an algebraic equation for the potential,

$$k^2\Phi = 4\pi G a^2\left[\rho_M\delta_M + 4\rho_R\Theta_{R,0} + \frac{3aH}{k}\left(i\rho_M v_M + 4\rho_R\Theta_{R,1}\right)\right] \tag{4.66}$$

where subscript M includes all matter such as baryons and dark matter and subscript R all radiation such as neutrinos and photons. More precisely

$$\rho_M v_M \equiv \rho_{\rm DM} v_{\rm DM} + \rho_B v_{\rm B} \tag{4.67}$$

and similarly for the other matter and radiation quantities.

Other components of Einstein's equation contain information not about the scalar perturbations Φ and Ψ, but vector and tensor perturbations. Scalar, vector, and tensor perturbations are decoupled: each evolves indepedently of the others. We will see in Chapter 5 that inflation can produce tensor perturbations, so it is important to know what the Einstein equation says about their evolution. We showed that there are two functions which can characterize tensor perturbations, h_+ and $h_\times$; each of these evolves independently and satisfies

$$\ddot{h}_\alpha + 2\frac{\dot{a}}{a}\dot{h}_\alpha + k^2 h_\alpha \quad = 0. \qquad ([4.62])$$

where α denotes $+, \times$. In an expanding universe, the amplitude of the gravity waves described by equation [4.62] falls off a given mode enters the horizon.

Suggested Reading

Most cosmology books offer some treatment of the perturbed Einstein equations in cosmology (although few give them the emphasis I think they are due). Again *The Large Scale Structure of the Universe* (Peebles) is a useful reference, especially for synchronous gauge. *Cosmological Inflation and Large Scale Structure* (Liddle and Lyth) has a very nice treatment which, among other virtues, explains the physics of gauge choices. Probably the two most comprehensive works are the review articles by Mukhanov, Feldman, and Brandenberger (1992) and Kodama and Sasaki (1984), with the former slghtly more accessible and the latter more general.

The general relativity books mentioned in Chapter 2 all have good discussion of gravity waves. Before turning to any of the technical literature, though, you must read *Black Holes and Time Warps* (Thorne), a wonderful mixture of the history, science, and personalities associated with 20th century general relativity. It is the best popular science book I have ever read.

Problems

4.1 Derive the Christoffel symbols, $\Gamma^i_{\mu\nu}$, given in equation [4.7]. When doing this, you will need g^{ij}. Show that it is equal to $\delta_{ij}(1-2\Phi)/a^2$.

4.2 Show that R_{ij} is given by equation [4.14].

4.3 The general relativistic expression for the stress energy tensor in terms of the distribution functions is given by

$$T_{\mu\nu}(\vec{x},t) = g_\alpha \int \frac{dP_1 dP_2 dP_3}{(2\pi)^3}(-g)^{-1/2}\frac{P_\mu P_\nu}{P_0} f(\vec{x},\vec{p},t) \qquad (4.68)$$

where the P_μ were defined back in equation [3.10], g_α is the number of spin states for species α, and g is the determinant of $g_{\mu\nu}$. Show that, with scalar perturbations to the metric, the phase space integral for the time-time component reduces to that in equation [4.22]. Show that the contribution from species α to T^0_i is

$$T^0_i = g_\alpha a \int \frac{d^3p}{(2\pi)^3} p_i f_\alpha(\vec{p},\vec{x},t). \qquad (4.69)$$

Note the extra factor of a.

4.4 Compute the time-space component of the Einstein tensor. Show that, in Fourier space,

$$G^0_i = 2ik_i\left(\frac{\dot{\Phi}}{a} - H\Psi\right). \qquad (4.70)$$

Combine with the stress-energy tensor derived in Problem (3) to show that

$$\dot{\Phi} - aH\Psi = \frac{4\pi G a^2}{ik}\left[\rho_{\rm DM} v_{\rm DM} + \rho_B v_{\rm B} - 4i\rho_\gamma \Theta_1 - 4i\rho_\nu \Theta_{\nu,1}\right]. \qquad (4.71)$$

The time-space component of Einstein's equations adds no new information once we already have the two equations derived in the text. Deciding which two to use is a matter of choice (although the best programmers use the superfluous equations as a check on their codes). In the next chapter, we will find the time-space component to be the most convenient when describing the scalar perturbations generated during inflation.

4.5 Take the Newtonian limit of Einstein's equations. Combine the time-time equation [4.26] with the time-space equation of Problem (4) to obtain the algebraic (i.e. no time derivatives) equation for the potential given in equation [4.66]. Show that this reduces to Poisson's equation (with the appropriate factors of a) when the wavelength is much smaller than the horizon ($k\eta >> 1$).

4.6 Fill in the blanks in the derivation of the tensor equation.

(a) Show that Γ^i_{jk} is given by equation [4.42] in the presence of tensor perturbations.

(b) Show that the last term in equation [4.47] is given by equation [4.53].

4.7 We have defined the perturbations to the photon distribution function via equation [3.32]. Show that, if Θ depends only on μ the cos of the angle between $\hat{k}(\equiv \hat{z}$ here) and $\hat{p}$, then $T_{11} - T_{22}$ vanishes. This is indeed the dependence we have been dealing with so far. This is yet another aspect of the Decomposition Theorem: the terms Θ that source the scalar perturbations (and are sourced by them) do not affect tensor perturbations. Anisotropies induced by tensor perturbations will have Θ of the form

$$\Theta(\mu, \phi) = (1 - \mu^2)\cos(2\phi)\Theta_+(\mu) \tag{4.72}$$

for those perturbations generated by h_+ and a similar expression for $h_\times$ with the cos replaced by a sin.

4.8 By taking the $_{12}$ component of Einstein's equation, show that $h_\times$ obeys the same equation as does h_+.

4.9 Show that scalars do not contribute to either $G_{11} - G_{22}$ or to G_{12}. This completes the demonstration of the Decomposition Theorem for scalars and tensors.

4.10 Consider vector perturbations to the metric. These can be described by two function h_{xz} and h_{yz} where again only the spatial part of the metric is perturbed. The perturbative part of g_{ij} is

$$h^V_{ij} = \begin{pmatrix} 0 & 0 & h_{xz} \\ 0 & 0 & h_{yz} \\ h_{xz} & h_{yz} & 0 \end{pmatrix}. \tag{4.73}$$

Show that h_{xz} and h_{yz} do not affect any of the equations we have derived so far for scalar or tensor evolution. Yet another aspect of the Decomposition Theorem.

4.11 Solve the wave equation [4.62] if the universe is purely matter dominated. Do the same for the radiation dominated case.

4.12 Define the *transfer function* for gravity wave evolution as

$$T(k) \equiv \frac{h_\alpha(k,\eta)}{h_\alpha(k,\eta=0)}\left(\frac{k\eta}{3j_1(k\eta)}\right). \tag{4.74}$$

You should recognize the term in parentheses as the inverse of the matter-dominated solution you derived in Problem (11). Solve equation [4.62] numerically and compute the transfer function. Compare your solution with the fit of Turner, White, & Lidsey,

$$T(y) = \left[1 + 1.34y + 2.5y^2\right]^{1/2} \tag{4.75}$$

where $y \equiv (k\eta_0/370h)$ (with h parametrizing the Hubble constant). Assume the universe today is flat and matter dominated, but account for transition from matter to radiation.

Chapter 5

Initial Conditions

We have derived the equations governing perturbations in the universe. Before we start solving the equations, we need to know the initial conditions. This quest for initial conditions will lead to a whole new field of physics, the theory of inflation. Inflation was introduced partly to explain how regions which could not have been in causal contact with each other have the same temperature. It was soon realized that the very mechanism that explains the uniformity of the temperature in the universe can also account for the origin of perturbations in the universe. Therefore, in order to produce a set of initial conditions, we will need to detour into the world of inflation. One warning: we are not sure that inflation is the mechanism that generated the initial perturbations. It is very difficult to test a theory based on energy scales well beyond the reach of accelerators. Nonetheless, it is by far the most plausible explanation. Indeed, one of current problems in cosmology is that there is really no viable alternative to inflation. Also, the next generation of CMB and large scale structure observations will put inflation to some sringent tests.

5.1 The Einstein-Boltzmann Equations at Early Times

Let us consider first the Boltzmann equations ([3.96] - [3.103]) at very early times. In particular, we want to consider times so early that for any $k-$ mode of interest, $k\eta << 1$. This inequality immediately leads to several important simplifications. Consider the terms $\dot{\Theta}$ and $ik\mu\Theta$ in equation [3.96]. The first term is of order Θ/η, while the second is of order $k\Theta$. Therefore, the first is larger than the second by a factor of order $1/(k\eta)$, which, by assumption, is much greater than one. In a similar way, we can argue that all terms in the Boltzmann equations multiplied by k can be neglected at early times. Physically, what this means is that at early times, all perturbations of interest have wavelengths ($\sim 1/k$) much larger than the distance over which causal physics operates. Therefore, for the purpose of evolution, the temperature field can be considered uniform: its spatial derivatives vanish (recall that the factors of k arise because in Fourier space a derivative gets turned into a factor of k). This physical argument leads to the second observation, that higher multipoles $(\Theta_1, \Theta_2, \ldots)$ are much smaller than the monopole, Θ_0. Therefore, the perturbations to the

photon and neutrino temperatures evolve according to

$$\begin{aligned} \dot{\Theta}_0 + \dot{\Phi} &= 0 \\ \dot{\Theta}_{\nu,0} + \dot{\Phi} &= 0 \end{aligned} \tag{5.1}$$

The same principles can be applied to the matter distributions. The overdensity equations reduce to

$$\begin{aligned} \dot{\delta}_{\rm DM} &= -3\dot{\Phi} \\ \dot{\delta}_B &= -3\dot{\Phi}. \end{aligned} \tag{5.2}$$

The velocities are comparable to the first moments of the radiation distributions, so they are smaller than the overdensities by a factor of order $k\eta$ and may be set to zero initially. In fact, the baryon velocity is not only comparable to the photon first moment, Θ_1: it is equal to it by virtue of the strength of Compton scattering. That is, the largeness of $\dot{\tau}$ in equation [3.102] ensures that $v_{\rm B} = -3i\Theta_1$. We will use this later when re-examining the Boltzmann equations closer to decoupling. For now, we are interested in times so early that the only relevant fact is that higher moments are all negligibly small.

Now let us turn to the Einstein equations at early times. First consider equation [4.26]. The first term there contains a factor of k^2 so may be neglected. Also the two matter terms on the right are negligible at early times since radiation dominates. Therefore, we have

$$3\frac{\dot{a}}{a}\left(\dot{\Phi} - \frac{\dot{a}}{a}\Psi\right) = 16\pi G a^2 \left(\rho_\gamma \Theta_0 + \rho_\nu \Theta_{\nu,0}\right). \tag{5.3}$$

But since radiation dominates, $a \propto \eta$ (recall equation [2.29] and the discussion immediately afterwards) so $\dot{a}/a = 1/\eta$. Therefore,

$$\begin{aligned} \frac{\dot{\Phi}}{\eta} - \frac{\Psi}{\eta^2} &= \frac{16\pi G\rho a^2}{3}\left(\frac{\rho_\gamma}{\rho}\Theta_0 + \frac{\rho_\nu}{\rho}\Theta_{\nu,0}\right) \\ &= \frac{2}{\eta^2}\left(\frac{\rho_\gamma}{\rho}\Theta_0 + \frac{\rho_\nu}{\rho}\Theta_{\nu,0}\right) \end{aligned} \tag{5.4}$$

where the last equality follows by virtue of the zero order Einstein equation.

To simplify further, we can define the ratio of neutrino energy density to the total radiation density as

$$f_\nu \equiv \frac{\rho_\nu}{\rho_\gamma + \rho_\nu}. \tag{5.5}$$

Then equation [5.4] can be rewritten as

$$\dot{\Phi}\eta - \Psi = 2\left((1 - f_\nu)\Theta_0 + f_\nu \Theta_{\nu,0}\right). \tag{5.6}$$

Recall that equation [5.1] relates the derivative of the monopoles to the derivative of the potential. We can therefore eliminate both monopoles from equation [5.6] by differentiating both right and left hand sides.

First, however, I will appeal to an incorrect version of the second Einstein equation. So far we have used only one Einstein equation. The second, equation [4.32], describes how the higher moments of the photon and neutrino distributions cause $\Psi+\Phi$ to be non-zero. Let us here neglect these higher order moments, which cause the sum of the gravitational potentials to be slightly non-zero*. Under this approximation, we can eliminate Ψ everywhere by simply setting it to $-\Phi$.

Now we can differentiate equation [5.6] to find

$$2\dot{\Phi}+\ddot{\Phi}\eta=-2\dot{\Phi} \tag{5.7}$$

where the right hand side follows since both $\dot{\Theta}_0$ and $\dot{\Theta}_{\nu,0}$ are equal to $-\dot{\Phi}$ for these large scale modes. Setting $\Phi=\eta^p$ leads to the algebraic equation:

$$4p+p(p-1)=0 \tag{5.8}$$

which allows two solution: $p=0,-3$. The $p=-3$ mode is the decaying mode. If it is excited very early on, it will quickly die out and have no impact on the universe. The $p=0$ mode, on the other hand, does not decay if excited. It is the mode we are interested in. If some mechanism can be found which excites this mode, this mechanism may well be responsible for the perturbations in the universe.

Focusing therefore on only the $p=0$ mode, we see that equation [5.6] relates the gravitational potential to the neutrino and photon overdensities:

$$\Phi=2\left((1-f_\nu)\Theta_0+f_\nu\Theta_{\nu,0}\right). \tag{5.9}$$

Both Θ_0 and $\Theta_{\nu,0}$ are also constant in time. In most models of structure formation, they are equal since whatever causes the perturbations tends not to distinguish between photons and neutrinos. Therefore, we will set

$$\Theta_0(k,\eta_i)=\Theta_{\nu,0}(k,\eta_i) \tag{5.10}$$

which leads to

$$\Phi(k,\eta_i)=2\Theta_0(k,\eta_i) \tag{5.11}$$

where I have explicitly written the $k-$ dependence of all these variables and the fact that we are setting up the initial conditions at some early time η_i.

The initial conditions for matter, both $\delta_{\rm DM}$ and $\delta_{\rm B}$, depend upon the nature of the primordial perturbations. Combining the first of equations [5.1] and [5.2] leads to

$$\delta_{\rm DM}=3\Theta_0+\text{constant} \tag{5.12}$$

with an identical equation for the baryon overdensity. Primordial perturbations are often divided into those for which the constant in equation [5.12] is zero (*adiabatic* perturbations)

*See Problem (2) for a careful accounting of the effect of the neutrino quadrupole; the photon quadrupole is kept miniscule by Compton scattering, so it really does not contribute to equation [4.32].

and those for which the constant is non-zero (*isocurvature* perturbations). Adiabatic perturbations have a constant matter to radiation ratio everywhere since

$$\frac{n_{\rm DM}}{n_\gamma}=\frac{n_{\rm DM}^{(0)}}{n_\gamma^{(0)}}\left[\frac{1+\delta_{\rm DM}}{1+3\Theta_0}\right]. \tag{5.13}$$

The prefactor, the ratio of zero order number densities, is a constant in both space and time. For the ratio of matter to radiation number density to be uniform, therefore, the combination inside the brackets which linearizes to $1+\delta_{\rm DM}-3\Theta_0$ must be independent of space. So the perturbations must sum to zero,

$$\delta_{\rm DM}=\delta_{\rm B}=3\Theta_0 \tag{5.14}$$

for adiabatic perturbations. There are models based on isocurvature perturbations, but these have not been very successful to date, so for the most part we will stick with adiabatic initial conditions.

For the most part, velocities and dipole moments are negligibly small in the very early universe. However, we will encounter situations where we need to know the initial conditions for these as well. You will show in Problem (3) that the appropriate initial conditions are

$$\begin{aligned}\Theta_1 &= \Theta_{\nu,1}=\frac{iv_{\rm B}}{3}=\frac{iv_{\rm DM}}{3}\\ &=-\frac{k\Phi}{6aH}.\end{aligned} \tag{5.15}$$

5.2 The Horizon

If this book was a novel or a biography, a better title for this section might be *Midlife Crisis*. The main character would have attended a good high school, studied hard and gotten good SAT scores, and gone on to an Ivy League college. There he fell in love with an exciting, but sensible, woman; upon graduating, he set up some interviews, and got a good job downtown. He married his college girlfriend, and after several years in the city, they moved to the suburbs, and had three kids. Our hero contributed to the community and was recognized all over town as a solid citizen. He was moving up fast in his company and there was talk about a political position. Just when he was about to declare his candidacy, he began to have doubts. "What have I been doing with my life? What is really important? Were all those years of study and work simply a 'track'? Did I go through these just because everyone else was moving in the same direction? Where is the innovation and the signature that my life is mine?" And worse, he has a secret, an underlying feeling that everything he has built is based on a fallacy.

OK, maybe it wouldn't be a best-seller but it does serve as a useful metaphor for our study of perturbations in the universe. Until now, we have done everything in a systematic, proper way. We reviewed the standard Big Bang cosmology. We expanded about this zero order smooth universe, getting evolution equations for the perturbations to the particle distributions and to the gravitational fields. We realized that these coupled differential equations needed initial conditions so in the last section we set those up. However, now we

must ask, What caused those initial perturbations? It is one thing to say that $\Phi = 2\Theta_0$ initially. It is quite another to explain what caused Φ to be non-zero in the first place.

And it is worse than that. To understand why let us recall the physical meaning of the conformal time η: it is the maximum comoving distance travelled by light since the beginning of the universe. Equivalently, objects separated by comoving distances larger than η today were not ever in causal contact: there is simply no way information could have propagated over distances larger than η. For this reason, η is called the comoving horizon.

With this in mind, we can now revisit the condition used in the previous section that $k\eta << 1$. The wavenumber k is roughly equal to the inverse of the wavelength of the mode in question (give or take a factor of 2π). Therefore $k\eta$ is the ratio of the comoving horizon to the comoving wavelength. If this ratio is much smaller than one, then the mode in question has a wavelength so large that no causal physics could possibly have affected it. A picture worth remembering is shown in Figure 5.1. The horizon grows as the scale factor increases. On the other hand, comoving wavelengths remain constant. All modes of cosmological interest therefore had wavelengths much larger than the horizon early on. Eventually these cosmological modes *enter* the horizon; after that, causal physics begins to operate on them.

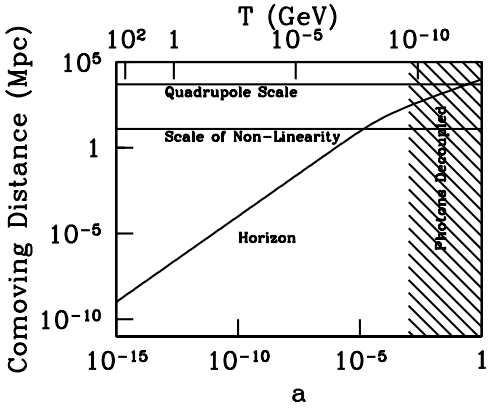


Figure 5.1: The comoving horizon as a function of the scale factor. Also shown are two comoving wavelengths, which remain constant with time. Early in the history of the universe, both of these modes – as well as all other modes of cosmological interest – had wavelengths much larger than the horizon.

The truly disturbing feature of this realization is most apparent when considering the microwave background today. On all scales observed the CMB is very close to isotropic. How can this be? The largest scales observed have entered the horizon just recently, long after decoupling. (An example is the scale corresponding to the quadrupole moment of

the CMB, shown in figure 5.1.) Before decoupling, the wavelengths of these modes are so large that no causal physics could force deviations from smooothness to go away. After decoupling, the photons do not interact at all; they simply freestream. So even though it is technically possible that photons reaching us today from opposite directions had a chance to communicate with each other and equilibrate to the same temperature, practically this could not have happened. Why then is the CMB temperature so uniform? This is a profound problem that we have glossed over by simply assuming that the temperature is uniform and that perturbations about the zero order temperature are small.

A more intuitive picture of the horizon problem is shown in figure 5.2. At any given time, the region within the cone is connected to us (at the center). Photons that we observe today from the last scattering surface were well outside our horizon when they were first emitted. The most disturbing aspect of this is the observation of large angle isotropy, an indication that photons apparently separated by many horizons at the last scattering surface nonetheless shared the same temperature (to a part in 10^5).

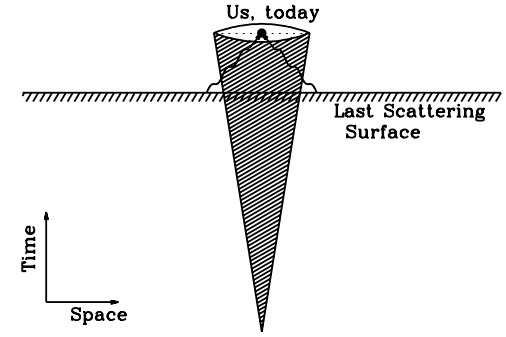


Figure 5.2: The horizon problem. The region inside the cone at any time is causally connected to us (at the center). Photons emitted from the last scattering surface (at redshift ~ 1100) started outside of this region. Therefore, at the LSS they were not in causal contact with us and therefore certainly not with each other. Yet their temperatures are almost identical.

5.3 Inflation

This section describes a beautiful solution to the horizon problem outlined in the previous section. First, we explore a logical way out of the previous argument by realizing that an early epoch of rapid expansion solves the horizon problem. Then we consider Einstein's equation to tell us what type of energy is needed in order to produce this rapid expansion, showing that negative pressure is required. Finally, we consider a scalar field theory and show that negative pressure is easy to accomodate in such a theory.

Two comments about the field theory implementation. First, field theory has a reputation as a difficult subject. It is, but the part we will need for inflation is decidedly simple. Indeed, almost all we will need to know about field theory we've already used in the previous chapter on general relativity. The second point is that there is no known scalar field which can drive inflation. (A skeptic might point out that there is no known fundamental scalar field at all!) Therefore, it may well be true that the idea of inflation is correct but it is driven by something other than a scalar field. Having said that, there are a number of reasons to work with scalar fields, as we will do whenever we need to specify the source of inflation. First, almost all fundamental particle physics theories contain scalar fields. In fact, historically it was particle physicists studying high energy extensions of the Standard Model (in particular Grand Unified Theories) who proposed the idea of inflation driven by a scalar field as a natural bi-product of some of these extensions. Second, almost all current work on inflation is based on a scalar field (or sometimes two). Finally, scalar field theory is the easiest to deal with. Including fermions or vectors doesn't seem to help much but it does complicate things severely.

5.3.1 A Solution to the Horizon Problem

Look again at figure 5.1. On top of the figure I have drawn an axis which depicts the temperature of the cosmic plasma for the given value of the scale factor. We know quite a bit about physics going up to the limits on the plot, several hundred GeV. Beneath these energies, the standard model of particle physics works very well. Beyond those energies, although we have ideas, there is no experimental reason to prefer one theory over another. Since the energy content of the universe determines $a(t)$, when you mentally extrapolate the horizon in figure 5.1 back to $a = 0$, or equivalently to infinitely high temperatures, you are really making an assumption. You are assuming that nothing stange happened early on, in particular that the universe was always radiation dominated at early times. If this were true, then it does indeed follow that the comoving horizon received a negligible contribution from the very early universe, that photons can travel only very small distances in the first fraction of a second after the Big Bang.

This suggests a solution to the horizon problem. Perhaps early in the history of the universe the scale factor did not grow as $t^{1/2}$. If it didn't, then our extrapolation back to very early times in figure 5.1 has be amended. How would a have to vary with t in order to generate a huge horizon early on? Let's play with this idea. Break up the total comoving horizon into two parts, one primordial and the standard one we have been computing until now:

$$\eta_{\rm total} = \eta_{\rm prim} + \int_{t_e}^{t} \frac{dt'}{a(t')} \tag{5.16}$$

where t_e denotes the time at the end of the period of *non-* $t^{1/2}$ expansion. If this epoch began at t_b, then

$$\eta_{\rm prim} = \int_{t_b}^{t_e} \frac{dt}{a(t)}. \tag{5.17}$$

Suppose that during this period $a \propto t^p$, again with p not necessarily equal to $1/2$ (as it would

be if the radiation dominated) or $2/3$ (matter dominated). Then

$$\begin{aligned} \eta_{\rm prim} &= \frac{1}{a_e}\int_{t_b}^{t_e} \frac{dt}{(t/t_e)^p} \\ &= \left(\frac{1}{1-p}\right)\frac{t_e}{a_e}\left[1-\left(t_e/t_b\right)^{p-1}\right] \end{aligned} \tag{5.18}$$

where a_e and a_b denote the scale factor at the beginning and end of this non-standard epoch. Since the universe becomes radiation dominated afterwards, (in which case $(da/dt)/a = 1/2t$) the time at the end of the epoch can be expressed as $t_e = 1/(2H_e)$. Also the ratio of beginning and ending times can be written in terms of the ratio of scale factors, so

$$\eta_{\rm prim} = \frac{1}{p-1}\frac{1}{2H_e a_e}\left(\frac{a_e}{a_b}^{1-1/p} - 1\right). \tag{5.19}$$

We would like $\eta_{\rm prim}$ to be larger than any scales of interest today, so that equilibration could have occurred very early. The largest comoving scales of interest today are those comparable to the Hubble radius, $1/(a_0 H_0)$,† that so we require $a_0 H_0 \eta_{\rm prim} >> 1$. Therefore, neglecting factors of order unity, we need

$$\frac{a_0 H_0}{a_e H_e}\left|\left(\frac{a_e}{a_b}^{1-1/p} - 1\right)\right| >> 1. \tag{5.20}$$

If $p < 1$ then the term in parentheses is of order one since (a_e/a_b) (which is greater than one) raised to a negative power is a small number. Then the primordial horizon is of order the prefactor, which is very small. To see that it is small, let's ignore the relatively brief epoch of recent matter domination and assume that the universe has been radiation dominated since a_e (you can correct this assumption in problem 4.5). Then H scales as a^{-2} so $a_0 H_0/a_e H_e = a_e$. If a_e corresponds to a time at which the temperature was 10^{15} GeV (more on this choice later), then the left hand side in equation [5.20] is of order $10^{-4}\text{eV}/10^{15}\text{Gev} = 10^{-28}$. It is not bigger than one. So epochs in which $a \propto t^p$ with $p < 1$ do not solve the horizon problem. This is already a fascinating conclusion for it tells us that epochs in which the universe is decelerating, i.e.

$$\frac{d^2 a}{dt^2} < 0, \tag{5.21}$$

do not solve the horizon problem.

Can an accelerating universe solve the horizon problem? Suppose $p > 1$; to simplify matters, suppose it is large enough that the power $1 - 1/p$ in equation [5.20] can be safely set to one. Then the condition for solving the horizon problem reduces to

$$\frac{a_0 H_0}{a_e H_e}\frac{a_e}{a_b} >> 1. \tag{5.22}$$

Since the first ratio has already been determined to be of order 10^{-28} (but again don't forget the corrections in Problem (6)), the horizon problem is solved if $a_e/a_b >> 10^{28}$. Perhaps

†Even though we are using the convention that $a_0 = 1$, it is useful to include it here to remind us that we will be considering ratios of scale factors at different times.

the most common way to arrange this is to construct a model wherein the scale factor grows exponentially with time during this brief epoch, which I'll now start calling *inflation*. In that case, $a_e/a_b = e^{\lambda(t_e - t_b)}$. For the scale factor to increase by a factor of 10^{28}, the argument of the exponential must be of order $ln(10^{28}) \sim 64$, so inflation can solve the horizon problem if the universe expands exponentially for about sixty e-folds.

The idea that the horizon blows up early on is depicted in figure 5.3. The physical size (a

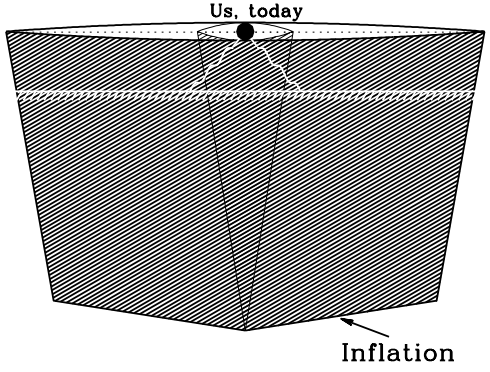


Figure 5.3: Inflationary solution to the horizon problem. Larger cone shows the horizon in an inflationary model; smaller inner cone shows the horizon without inflation. During inflation, the physical horizon blows up very rapidly.

times the comoving size) of a casually connected region blows up exponentially quickly during inflation. So regions that we observe to be astronomical today were actually microscopically small before inflation, and they were in causal contact with each other.

While the physical picture of figure 5.3 has permeated the popular and much of the technical literature about inflation, for our purposes, it will be important to think in comoving space. Towards that end, consider figure 5.4 which shows the comoving Hubble radius as a function of the scale factor. The right side of this plot is simply figure 5.1, which tells us that the comoving scales of interest to us were far larger than the Hubble radius in the standard cosmology. The left hand side of the plot though shows that an inflationary epoch reduces the comoving Hubble radius dramatically. This makes sense: since the scale factor is inflating very rapidly, it becomes increasingly difficult for photons to move along the comoving grid (which is itself expanding with a). Before inflation started, the comoving Hubble radius was very large, larger than any scale of cosmological interest today, so all such scales were well within the horizon.

Note the symmetry in figure 5.4. Scales just entering the horizon today – roughly sixty e-folds after the end of inflation – left the horizon sixty e-folds before the end of inflation. The amplitude of the perturbations on these scales remained constant as long as they were

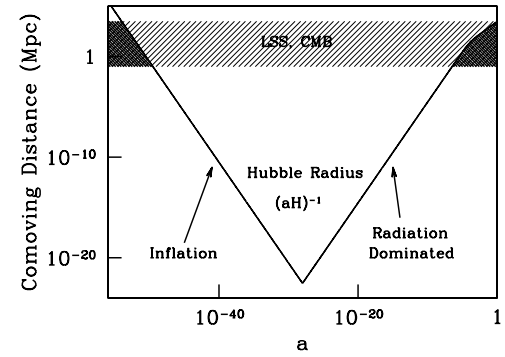


Figure 5.4: The comoving Hubble radius as a function of scale factor. Scales of cosmological interest (shaded band) were larger than the Hubble radius until $a \sim 10^{-5}$. Dark shaded regions show when these scales were smaller than the Hubble radius, and therefore susceptible to microphysical processes.

super-horizon. So, when we measure them today, we are actually seeing them as they were when they first left the horizon during the inflationary era (modulo whatever processing has taken place since they re-entered the horizon, processing we will study in great detail in Chapters 6 and 7). It is clearly important to understand the generation of perturbations during inflation.

Another feature of figure 5.4 worth noting is that the comoving Hubble radius is plotted, not the comoving horizon. The total comoving horizon defined in equation [5.16] is not a very useful time-parameter. It asymptotes to a very large value at the outset of inflation and chages little after that. All scales of interest today are smaller than the comoving horizon, but this does not tell us whether physics can operate on these scales today: it simply says that at one point in the distant past, such scales were connected. If the scales are smaller than the Hubble radius, though, which is the distance light can travel in one expansion time, then physics can operate on these scales at present. In such a case, for example, we can no longer simply assume that perturbations remain constant. An alternative, which is also common in the literature on inflation, is to subtract off the primordial part of the comoving horizon, and re-define η as

$$\eta = \int_{t_e}^{t} \frac{dt'}{a(t')}, \tag{5.23}$$

so that the total comoving horizon is $\eta_{\rm prim} + \eta$. This is the convention we will follow; note that this means that during inflation, η is negative, but as always monotically increasing. A scale leaves the horizon in the sense of figure 5.4 when $k|\eta|$ becomes less than one, and returns at late times when $k\eta$ becomes larger than one.

To sum up, inflation – an epoch in which the universe accelerates – solves the horizon

problem. During the accelerated expansion the horizon gets extremely large, so that it is much larger than we would otherwise expect. Regions which are separated by vast distances today were actually in causal contact during this very brief, primordial period of inflation. At that time, these regions were given the necessary initial conditions, the smoothness we observe today, but also, as we will soon see, the small perturbations about smoothness that eventually grew into galaxies and other structure in the universe.

5.3.2 Negative Pressure

We have shown that an accelerating universe can solve the horizon problem. Since general relativity ties the expansion of the universe to the energy in it, we now need to ask what type of energy can produce acceleration. We can get an immediate answer if we appeal to the time-time and space-space components of the zero order Einstein equations. They are (equations [2.26] and [2.62])

$$\begin{aligned}\left(\frac{da/dt}{a}\right)^2 &= \frac{8\pi G}{3}\rho \\ \frac{d^2a/dt^2}{a}+\frac{1}{2}\left(\frac{da/dt}{a}\right)^2 &= -4\pi G P. \end{aligned} \tag{5.24}$$

Multiplying the first of these by a half and then subtracting one from the other eliminates the first derivative of a leaving

$$\frac{d^2a/dt^2}{a} = -\frac{4\pi G}{3}\left(\rho+3P\right). \tag{5.25}$$

Acceleration is defined to mean that d^2a/dt^2 is positive. For this to happen, the terms in parentheses on the right must be negative. So inflation requires

$$P < -\frac{\rho}{3}. \tag{5.26}$$

Since the energy density is always positive, the pressure must be negative.

Negative pressure is not something with which we have any familiarity. Non-relativistic matter has small positive pressure proportional to temperature divided by mass, while a relativistic gas has $P = +\rho/3$, again positive. So whatever it is that drives inflation is not ordinary matter or radiation (we suspected as much since matter and radiation lead to $p = 2/3, 1/2$ in the language of the previous section, values of p which were shown not to solve the horizon problem).

5.3.3 Implementation with a Scalar Field

We have become familiar with the fields $\Psi(\vec{x},t)$ and $\Phi(\vec{x},t)$, deriving equations for them which govern their evolution (Einstein's equations) and the evolution of particles which are affected by them. These two fields are parts of the multi-component field, the metric $g_{\mu\nu}$. The metric is one of the fundamental fields in physics, but there are others. Every elementary

particle – the electron, neutrino, quarks, photon, etc. – is associated with its own field. It would be wonderful if one of these fields, the electromagnetic potential associated with photons say, could serve as the source for an inflationary model. Unfortunately we do not yet have such a concrete model. Instead, I will discuss inflation in terms of a generic scalar field (not a fermion like the quarks and leptons or a vector like the electromagnetic field). The simplest version of the standard model does indeed have within it one such field, the Higgs field. But, again unfortunately, we know too much about the Higgs of the standard model. Its interactions and properties are constrained enough for us to know that it cannot serve as the source for inflation. So we will drop any pretensions of connecting the generic scalar field which drives inflation to known physics. Making this connection is left as a homework problem for a future Nobel laureate.

We want to know if a scalar field – which I will call $\phi(\vec{x},t)$, not to be confused with the perturbation $\Phi(\vec{x},t)$ – can have negative $P+3\rho$. So our first task is to write down the energy-momentum tensor for ϕ. This is

$$T^\alpha_\beta = g^{\alpha\nu}\frac{\partial\phi}{\partial x^\nu}\frac{\partial\phi}{\partial x^\beta} - g^\alpha_\beta\left[\frac{1}{2}g^{\mu\nu}\frac{\partial\phi}{\partial x^\mu}\frac{\partial\phi}{\partial x^\nu}+V(\phi)\right]. \tag{5.27}$$

Here $V(\phi)$ is the potential for the field. For example a free field with mass m has a potential $V(\phi) = m^2\phi^2/2$. A warning about signs: if you delve into the literature you will invariably find different signs than those in equation [5.27]. These are dictated by the choice of metric. While our metric signature $(-,+,+,+)$ is probably most common in the context of cosmology, it is probably not as common in particle physics. Beware. Following most work on the subject we will assume that ϕ is mostly homogeneous, consisting of a zero order part, $\phi^{(0)}(t)$, and a first order perturbation, $\delta\phi(\vec{x},t)$. In this section we will derive information about the zero order homogeneous part, its energy density and pressure and its time evolution. Later we will consider its perturbations and how they are generated.

For the homogeneous part of the field, only time derivatives of ϕ are relevant so the energy-momentum tensor in equation [5.27] reduces to

$$T^{(0)\alpha}_\beta = -g^\alpha_0 g^0_\beta\left(\frac{d\phi^{(0)}}{dt}\right)^2 + g^\alpha_\beta\left[\frac{1}{2}\left(\frac{d\phi^{(0)}}{dt}\right)^2 - V(\phi^{(0)})\right]. \tag{5.28}$$

The time-time component of T^0_0 is equal to $-\rho$, so the energy density is

$$\rho = \frac{1}{2}\left(\frac{d\phi^{(0)}}{dt}\right)^2 + V(\phi^{(0)}). \tag{5.29}$$

The first term here is the kinetic energy density of the field, the second its potential energy density. A homogeneous scalar field therefore has much the same dynamics as a single particle moving in a potential (think of $\phi^{(0)}(t)$ as the position of the particle $x(t)$). In fact this analogy dominates even the language used to describe inflation. The pressure for the homogeneous field is $P = T^{(0)i}_i$ (no sum over spatial index i), so

$$P = \frac{1}{2}\left(\frac{d\phi^{(0)}}{dt}\right)^2 - V(\phi^{(0)}). \tag{5.30}$$

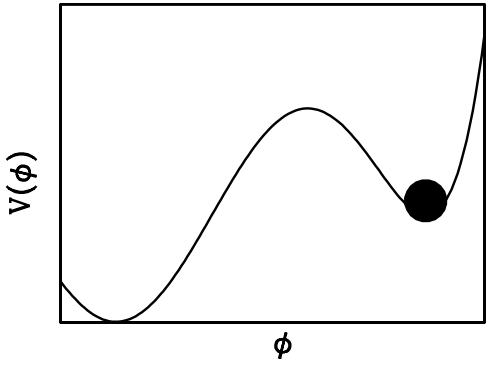


Figure 5.5: A scalar field trapped in a false vacuum. Since it is trapped, it has little kinetic energy. The potential energy is non-zero, however, so the pressure is negative.

A field configuration with negative pressure is therefore one with more potential energy than kinetic. An example is shown in figure 5.5, in which a field is trapped in a *false vacuum*, i.e. a local, but not the global, minimum of the potential.

There is something important to notice about a field trapped in a false vacuum. Since $\phi^{(0)}$ is constant, its energy density (which is all potential) remains constant with time, at least until it tunnels out of the false vacuum. Constant energy density is much different than anything with which we are familiar. The densities of both matter and radiation, for example, fall off very rapidly as the universe expands. Therefore, even if the universe initially contains a mixture of matter, radiation, and false vacuum energy, it will quickly become dominated by the vacuum energy. In fact this phenomenon, that the energy density of a scalar field away from its true minimum dominates over other constituents in the universe, is more general than this. Most models of inflation are driven by a field not trapped in a false vacuum, but rather slowly rolling towards the true minimum. The energy density of such a slow rolling field is also very close to constant (if the potential is not too steep) so it quickly comes to dominate.

For a trapped field, it is trivial to determine the evolution of the scale factor. Since the energy density is constant, Einstein's equation for the evolution of a is

$$\frac{da/dt}{a} = \sqrt{\frac{8\pi G\rho}{3}} = \text{constant}. \tag{5.31}$$

We immdiately see that a field trapped in a false vacuum produces exponential expansion,

$$a(t) = a_e e^{H(t-t_e)}, \tag{5.32}$$

where again a_e and t_e are the scale factor and time at the end of inflation. The primordial

comoving horizon defined in equation [5.17] is, in this case,

$$\eta_{\text{prim}} = \frac{1}{H_e a_e}\left(e^{H(t_e-t_b)} - 1\right). \tag{5.33}$$

So if the field is trapped for at least sixty e-foldings ($H(t_e - t_b) > 60$), the horizon problem is solved.

To determine the evolution of $\phi^{(0)}$ in general when the field is not trapped, we return to the Einstein equations as given in [5.24]. Consider the first of these. If the dominant component in the universe is ϕ, then the energy density on the right hand side becomes $(d\phi^{(0)}/dt)^2/2 + V$. Differentiating this first equation therefore leads to

$$2\frac{da/dt}{a}\left[\frac{d^2a/dt^2}{a} - \left(\frac{da/dt}{a}\right)^2\right] = \frac{8\pi G}{3}\left[\left(\frac{d\phi^{(0)}}{dt}\right)\left(\frac{d^2\phi^{(0)}}{dt^2}\right) + V'\frac{d\phi^{(0)}}{dt}\right] \tag{5.34}$$

where V' is defined as the derivative of V with respect to the field $\phi^{(0)}$. We can replace the first term in brackets on the left by $-4\pi G(\rho/3 + P)$ as in equation [5.25]. Similarly the second term on the left is $8\pi G\rho/3$. The left hand side therefore becomes

$$\frac{da/dt}{a}8\pi G\left[-(\rho/3) - P - 2\rho/3\right] = -8\pi GH\left(\frac{d\phi^{(0)}}{dt}\right)^2. \tag{5.35}$$

Equating this to the right side of equation [5.34] leads to the evolution equation for a homogeneous scalar field in an expanding universe

$$\frac{d^2\phi^{(0)}}{dt^2} + 3H\frac{d\phi^{(0)}}{dt} + V' = 0. \tag{5.36}$$

Most models of inflation are *slow roll* models, in which the zero order field, and hence the Hubble rate, vary slowly. Therefore, it is conventional to define

$$\epsilon \equiv \frac{-\dot{H}}{aH^2} = -\frac{a}{H}\frac{dH}{da} \tag{5.37}$$

where overdots still denote derivatives with respect to conformal time η. Since H is always decreasing, ϵ is always positive. During inflation, it is typically small, while it is equal to two during a radiation era. In fact, one definition of an inflationary epoch is one in which $\epsilon < 1$. To express a in terms of η, we use the approximation that ϵ is constant. In that case, η can be integrated by parts

$$\begin{aligned}\eta &= \int_{a_e}^{a}\frac{da}{Ha^2} \\ &= \frac{-1}{aH}\Big|_{a_e}^{a} - \int_{a_e}^{a}\frac{da}{a}\frac{dH/da}{H^2} \\ &= \frac{-1}{aH}\Big|_{a_e}^{a} + \epsilon\int_{a_e}^{a}\frac{da}{Ha^2}.\end{aligned} \tag{5.38}$$

Fortuitously, the last term on the bottom line here is equal to ϵ times the integral we need to evaluate. So,

$$\eta(a) = \frac{1}{1-\epsilon}\left(\frac{1}{a_e H_e} - \frac{1}{aH}\right). \tag{5.39}$$

The comoving Hubble radius at the end of inflation, $(a_e H_e)^{-1}$ is very small (see figure 5.4), so during inflation

$$\eta \simeq -\frac{1}{aH}\,\frac{1}{1-\epsilon}. \tag{5.40}$$

During inflation, then, η is very nearly equal to the minus the Hubble radius.

5.4 Gravity Wave Production

Inflation does more than solve the horizon problem. The power of inflation is its ability to correlate scales that would otherwise be disconnected. The zero order scheme outlined in the previous section ensures that the universe will be uniform on all scales of interest today. There are perturbations about this zero order scheme though, and these perturbations – again acting early on when the scales are causally connected – persist long after inflation has terminated.

We are most interested in perturbations to the energy density since these are the seeds for galaxy formation and show up as anisotropies in the microwave background today. In addition to perturbations in the energy density, though, inflation generates fluctuations in the gravitational metric. This is not surprising since metric perturbations are coupled to density perturbations. Treating the combined fluctuations in the density and the scalar part of the metric is algebraically tedious (as we will see in the next section) though, and the underlying physics is not very transparent. Therefore, in this section we will do a simpler problem and examine the tensor fluctuations in the metric. As we saw in Chapter 3, these are not coupled to the density so are much easier to compute. In addition, these fluctuations turn out to be a unique signature of inflation and offer the best window on the physics driving inflation.

During inflation, the universe consists primarily of a uniform scalar field and a uniform background metric. Against this background, the fields fluctuate quantum mechanically. At any given time, the average fluctuation is zero, because there are regions in which the field is slightly larger than its average value and regions in which it is smaller. The average of the square of the fluctuations (the variance), however, is not zero. Our goal is to compute this variance and see how it evolves as inflation progresses. Looking ahead, once we know this variance, we can draw from a distribution with this variance to set the initial conditions.

5.4.1 Quantizing the Harmonic Oscillator

In order to compute the quantum fluctuations in the metric, we need to quantize the field. The way to do this, in the case of both tensor and scalar perturbations, is to rewrite the problem so that it looks like a simple harmonic oscillator. Once that is done, we will appeal to our knowledge of this simple system. It is worthwhile therefore recording some basic facts about the quantization of the harmonic oscillator.

- A simple harmonic oscillator with unit mass and frequency ω is governed by the equation

$$\frac{d^2x}{dt^2} + \omega^2 x = 0. \tag{5.41}$$

- Upon quantization, x becomes a quantum operator

$$\hat{x} = v(\omega,t)\hat{a} + v^*(\omega,t)\hat{a}^\dagger \tag{5.42}$$

  where v is a normalized, solution to equation [5.41],

$$v(\omega,t) = \frac{e^{-i\omega t}}{\sqrt{2\omega}} \tag{5.43}$$

  and $\hat{a}$ is an operator which acts on the state of the system.

- In particular, $\hat{a}$ annihilates the *vacuum* state $|0>$, in which there are no particles. It also satisfies the commutation relation

$$[\hat{a},\hat{a}^\dagger] \equiv \hat{a}\hat{a}^\dagger - \hat{a}^\dagger\hat{a} = 1. \tag{5.44}$$

  Other commutators vanish: $[\hat{a},\hat{a}] = [\hat{a}^\dagger,\hat{a}^\dagger] = 0$. It is straightforward to show (Problem (8)) that these commutation relations are equivalent to the (perhaps more familiar) relations between $\hat{x}$ and its momentum $\hat{p}$:

$$[\hat{x},\hat{p}] = i. \tag{5.45}$$

These facts enable us to compute the quantum fluctuations of the operator $\hat{x}$ in the ground state:

$$\begin{aligned} <|\hat{x}|^2> &\equiv <0|\hat{x}^\dagger\hat{x}|0> \\ &= <0|\left(v^*\hat{a}^\dagger + ve^{-i\omega t}\hat{a}\right)\left(v\hat{a} + v^*\hat{a}^\dagger\right)|0>. \end{aligned} \tag{5.46}$$

Since $\hat{a}|0>=0$, the first term in the second set of parentheses vanishes. Similarly, $<0|\hat{a}^\dagger = 0$, so we are left with

$$\begin{aligned} <|\hat{x}|^2> &= |v(\omega,t)|^2 <0|\hat{a}\hat{a}^\dagger|0> \\ &= |v(\omega,t)|^2 <0|[\hat{a},\hat{a}^\dagger] + \hat{a}^\dagger\hat{a}|0>. \end{aligned} \tag{5.47}$$

The second term again vanishes since $\hat{a}$ annihilates the vacuum, while the first is unity, so the variance in $\hat{x}$ is

$$<|\hat{x}|^2> = |v(\omega,t)|^2, \tag{5.48}$$

in this case $1/2\omega$. This is (almost) all we need to know about quantum fluctuations in order to compute the fluctuations in the early universe generated by inflation.

5.4.2 Tensor Perturbations

Recall that tensor perturbations to the metric are described by two function h_+ and $h_\times$, each of which obeys the equation ([4.62]),

$$\ddot{h} + 2\frac{\dot{a}}{a}\dot{h} + k^2 h = 0. \tag{5.49}$$

We would like to massage this equation into the form of a harmonic oscillator, so that h can be easily quantized. To do this, define[‡]

$$v \equiv \frac{ah}{\sqrt{16\pi G}}. \tag{5.50}$$

Derivatives of h with respect to conformal time can be rewritten as

$$\frac{\dot{h}}{\sqrt{16\pi G}} = \frac{\dot{v}}{a} - \frac{\dot{a}}{a^2}v \tag{5.51}$$

and

$$\frac{\ddot{h}}{\sqrt{16\pi G}} = \frac{\ddot{v}}{a} - 2\frac{\dot{a}}{a^2}\dot{v} - \frac{\ddot{a}}{a^2}v + 2\frac{(\dot{a})^2}{a^3}v. \tag{5.52}$$

Inserting these into equation [5.49], and multiplying by $\sqrt{32\pi G}$, gives

$$\begin{aligned} \frac{\ddot{v}}{a} - 2\frac{\dot{a}}{a^2}\dot{v} - \frac{\ddot{a}}{a^2}v + 2\frac{(\dot{a})^2}{a^3}v &+ 2\frac{\dot{a}}{a}\left(\frac{\dot{v}}{a} - \frac{\dot{a}}{a^2}v\right) + k^2\frac{v}{a} \\ &= \frac{1}{a}\left[\ddot{v} + \left(k^2 - \frac{\ddot{a}}{a}\right)v\right] = 0. \end{aligned} \tag{5.53}$$

This is precisely the form we know how to use. It has no damping term ($\propto \dot{v}$) so we can immediately write down an expression for the quantum operator

$$\hat{v}(\vec{k},\eta) = v(k,\eta)\hat{a}_{\vec{k}} + v^*(k,\eta)a^\dagger_{\vec{k}}, \tag{5.54}$$

where the coefficients of the creation and annihilation operators satisfy the equation

$$\ddot{v} + \left(k^2 - \frac{\ddot{a}}{a}\right)v = 0. \tag{5.55}$$

[‡]Regarding the factor of $16\pi G$ here, the only way I know of deriving this is to write down the action for the fields $h_{+,\times}$. The kinetic term is then multiplied by a factor of $1/32\pi G$. A canonical scalar field has prefactor equal to a half. So the additional $16\pi G$ must be absorbed into a redefinition of the field. The hard part of this is writing down the action to second order in perturbation variables. We have seen that even first order perturbations are cumbersome to track. On the other hand, by dimensional analysis – the fact that $h(\vec{x})$ is dimensionless while a canonical scalar field has dimensions equal to mass – we could have guessed that the factor of $m_{\rm Pl} = G^{-1/2}$ is required. Finally note that this prefactor does not affect the equation for v; it simply provides the normalization that becomes important when trying to determine the amplitude of the gravity wave spectrum.

Using our harmonic oscillator analogy, we can write the variance of perturbations in the v field as

$$< \hat{v}^\dagger(\vec{k},\eta)\hat{v}(\vec{k}',\eta) >= |v(\vec{k},\eta)|^2(2\pi)^3\delta^3(\vec{k}-\vec{k}'). \tag{5.56}$$

There is one difference between this expression and the analagous expression for the one-dimensional harmonic oscillator in equation [5.48]. A quantum field is defined in all space, so can be considered as a collection, an infinite collection, of oscillator, each at a different spatial position (or, in Fourier space, at different values of $\vec{k}$). The qunatum fluctuations in each of these oscillators are independent (as long as the equations are linear) so $\hat{v}(\vec{k})$ is completely uncorrelated with $\hat{v}(\vec{k}')$ if $\vec{k} \neq \vec{k}'$. The Dirac delta function in equation [5.56] enforces this independence. Recalling that $v = ah/\sqrt{16\pi G}$, we see that

$$\begin{aligned} < \hat{h}^\dagger(\vec{k},\eta)\hat{h}(\vec{k}',\eta) > &= \frac{16\pi}{m_{\rm Pl}^2}\frac{|v(k,\eta)|^2}{a^2}(2\pi)^3\delta^3(\vec{k}-\vec{k}') \\ &\equiv (2\pi)^3 P_h(k)\delta^3(\vec{k}-\vec{k}') \end{aligned} \tag{5.57}$$

where the seond line defines the *power spectrum* of the primordial perturbations to the metric. Conventions for the power spectrum abound in the literature; the one I've chosen in equation [5.57] is not the most popular in the early universe community. Often a factor of k^{-3} is added so that the power spectrum is dimensionless. I prefer to omit this factor to be consistent with the large scale structure community which likes its power spectra to have dimensions of k^{-3}. In any event, with this definition,

$$P_h(k) = \frac{16\pi}{m_{\rm Pl}^2}\frac{|v(k,\eta)|^2}{a^2}. \tag{5.58}$$

We have now reduced the problem of determining the spectrum of tensor perturbations produced during inflation to one of solving a second order differential equation for $v(k,\eta)$, equation [5.55]. To solve this equation, we first need to evaluate $\ddot{a}/a$ during inflation. Recall that $H = \dot{a}/a^2$, so

$$\begin{aligned} \frac{\ddot{a}}{a} &= \frac{1}{a}\frac{d}{d\eta}\left(a^2 H\right) \\ &= 2a^2H^2 + a\dot{H}. \end{aligned} \tag{5.59}$$

Rewriting the second derivative in terms of ϵ leads to

$$\frac{\ddot{a}}{a} = 2a^2H^2\left(1 - \frac{\epsilon}{2}\right). \tag{5.60}$$

Using the solution for η in equation [5.40] leads to

$$\frac{\ddot{a}}{a} = \frac{2-\epsilon}{\eta^2(1-\epsilon)^2}. \tag{5.61}$$

So the equation for v is

$$\ddot{v} + \left(k^2 - \frac{2-\epsilon}{\eta^2(1-\epsilon)^2}\right)v = 0. \tag{5.62}$$

Finally we introduce one more piece of notation. Define a new index

$$\nu \equiv \frac{3-\epsilon}{2(1-\epsilon)} \leftrightarrow \epsilon = 1 + \frac{2}{1-2\nu}. \tag{5.63}$$

The coefficient of $1/\eta^2$ can be rewritten in terms of ν, so that

$$\ddot{v} + \left(k^2 + \frac{(1/4)-\nu^2}{\eta^2}\right) v = 0. \tag{5.64}$$

The solution to this second order differential equation is proportional to the Hankel function of the first kind of order ν, $(-k\eta)^{1/2} H_\nu^{(1)}(-k\eta)$. To determine the coefficient, we consider v at very early times before inflation has done most of its work. At that time, $-\eta$ is large, of order $\eta_{\rm prim}$, so the k^2 term dominates, and the equation reduces precisely to that of the simple harmonic oscillator. In that case, we know (equation [5.43]) that the properly normalized solution is $e^{-ik\eta}/\sqrt{2k}$. So we have

$$\begin{aligned} \lim_{-k\eta\to\infty} v(k,\eta) &= \frac{e^{-ik\eta}}{\sqrt{2k}} \\ &= C \lim_{-k\eta\to\infty} \left[(-k\eta)^{1/2} H_\nu^{(1)}(-k\eta)\right]. \end{aligned} \tag{5.65}$$

The asymptotic limit of the Hankel function is

$$\lim_{-k\eta\to\infty} H_\nu^{(1)}(-k\eta) = e^{-(2\nu+1)i\pi/4} \sqrt{\frac{2}{-\pi k\eta}} \exp\left[-ik\eta\right], \tag{5.66}$$

so

$$v(k,\eta) = e^{(2\nu+1)i\pi/4} \sqrt{\frac{-\pi\eta}{4}} H_\nu^{(1)}(-k\eta). \tag{5.67}$$

After inflation has worked for many e-folds $k|\eta|$ becomes very small. Now that v has been normalized, we can determine the amplitude of v, and hence the variance of the super-horizon gravitational wave amplitude, by taking the small argument limit of the Hankel function.

$$\lim_{-k\eta\to 0} v(k,\eta) = e^{(2\nu+1)i\pi/4} \sqrt{\frac{-\eta\pi}{4}} \frac{-i\Gamma(\nu)}{\pi} \left(\frac{2}{-k\eta}\right)^\nu. \tag{5.68}$$

Figure 5.6 shows the evolution of $h \propto v/a$ during inflation. At early times h falls as $1/a$ as inflation reduces the amplitude of the modes. Once $-k\eta$ becomes smaller than unity, the mode leaves the horizon, after which h remains constant. We could have anticipated this without resorting to Hankel functions: when k can be neglected, equation [5.55] reduces to $\ddot{v} - (\ddot{a}/a)v = 0$, with the obvious solution, $v \propto a$. It is a useful exercise (Problem (9)) to show that the limit for v in equation [5.68] does indeed scale as a.

The primordial power spectrum for tensor modes, which scales as $|v|^2/a^2$, is therefore constant in time after inflation has stretched the mode to be larger than the horizon. This constant determines the initial conditions for the gravity waves, those with which to start

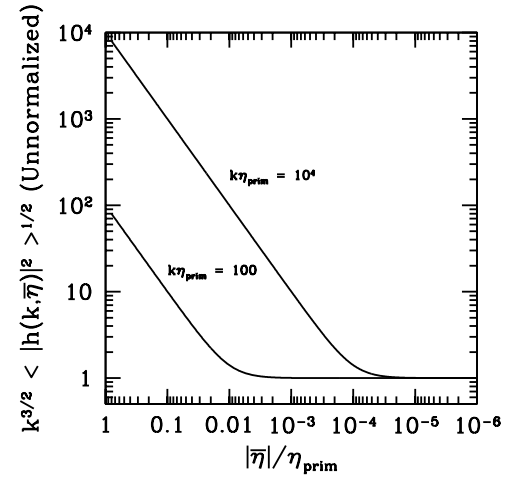


Figure 5.6: The root mean square fluctuations in the tensor field during inflation for two different $k-$ modes. Once a mode "leaves the horizon" ($-k\eta \sim 1$), its RMS amplitude remains constant. Note that, after a mode has left the horizon, its RMS amplitude times $k^{3/2}$ is the same for all modes. This is called a scale-free spectrum, strictly true only if $\epsilon = 0$ (the choice here).

off $h_{+,\times}$ at early times (where in this context "early" means well after inflation has ended but before decoupling). Equations [5.58] and [5.68] show that this constant is

$$P_h(k) = \frac{16\Gamma^2(\nu)}{k^3 m_{\rm Pl}^2} \frac{2^{2\nu-2}}{(-\eta a)^2(-k\eta)^{2\nu-3}}. \tag{5.69}$$

Since we know the time dependence here is illusory, it is convenient – or at least customary – to evaluate a and η when $-k\eta = 1$. Then, since (equation [5.40]) $-\eta a = [H(1-\epsilon)]^{-1}$ and $\Gamma(3/2) = \sqrt{\pi}/2$,

$$P_h(k) = \frac{2^{2\nu}\pi\Gamma^2(\nu)}{k^3\Gamma^2(3/2)}(1-\epsilon)^2 \frac{H^2}{m_{\rm Pl}^2}\Big|_{-k\eta=1}. \tag{5.70}$$

This is our final expression for the primordial power spectrum of gravity waves. Detection of these waves would, quite remarkably, measure the Hubble rate during inflation. Since potential energy usually dominates kinetic energy in inflationary models, a measure of H would be tantamount to measuring the potential V, again quite remarkable in view of the likelihood that inflation was generated by physics at energy scales above 10^{15} GeV, twelve order of magnitude beyond the capacity of present day accelerators. There is no guarantee that gravity waves produced during inflation will be detectable. Indeed, since $H^2 = \rho/m_{\rm Pl}$,

we see that the power spectrum is proportional to $\rho/m_{\rm Pl}^4$, the energy density at the time of inflation in units of the Planck mass. If inflation takes place at scales sufficiently smaller than the Planck scale, then primordial gravity waves will not be detected. Later in the book, we will develop the machinery necessary to answer the question, How small can the gravity wave component be and still be detected?

Two final technical points are in order regarding equation [5.70]. First, this is the power spectrum for h_+ and $h_\times$ separately; these are uncorrelated, so the power spectrum for all modes must be multiplied by a factor of 2. Second, I have chosen to evaluate P_h at $-k\eta = 1$; in the literature, you will often find it evaluated at $k/aH = 1$. Since P_h is constant, of course, it doesn't matter when it is evaluated. I mention it only because the final expression for P_h looks slightly different in the two cases (see Problem (10)).

5.5 Density Perturbations

The goal of this chapter is to find the initial conditions for the perturbations about a smooth FRW universe. In §5.1, we related all the perturbations $(\Theta_0, \delta_{\rm DM}, \ldots)$ to the scalar perturbations to the metric, Ψ. Once we find initial conditions for Ψ, we therefore will have initial conditions for all other perturbative quantities as well. In principle, we are almost there: we now know how to compute the spectrum of these primordial scalar – or density – perturbations. Following the previous section, we should systematically

- Get an evolution equation for Ψ, akin to equation [5.49] for tensors.
- Put this equation into the form of a simple harmonic oscillator, similar to equation [5.53].
- Quantize Ψ and determine the power spectrum of the perturbations by solving the simple harmonic oscillator-like equation, normalizing when the mode is within the horizon. This will give the primordial power spectrum, the equivalent of equation [5.70].

In principle this is exactly what we are going to do in this section. Unfortunately, in practice, these steps – especially the first two – will require quite a bit more work than they did in the tensor case. The main complication in the scalar case is that Ψ couples to perturbations in the scalar field ϕ driving inflation[§]. Contrast this with the simplicity of the tensor perturbations, $h_{+,\times}$, which do not couple to the scalar field at all (see Problem (11) for a proof of this). We had therefore just one second order equation for h which followed immediately from Einstein's equations. This was pretty easy to turn into the form of a harmonic oscillator. In the scalar case, we will need two equations since we have two perturbation variables Ψ and ϕ. These will be coupled equations, and finding the linear combination of the two which behaves like a harmonic oscilator is not trivial. Nonetheless, the problem is of sufficient importance that it is worth incurring the algebraic expense.

[§]Again, I repeat the warning that the scalar field driving inflation, ϕ, is not to be confused with Φ, the other field (in addition to Ψ) describing scalar perturbations to the metric. Throughout this discussion, Φ will set to $-\Psi$ since higher order moments are negligible during inflation.

5.5.1 Evolution equations for Ψ and ϕ

There are a number of ways of obtaining equations for the perturbations to the metric and the scalar field. The simplest is to use what we have done so far, and choose two components of the Einstein equations. The two most convenient are the time-time (equation [4.20]) and time-space (Problem (4) of Chapter 4):

$$\begin{aligned} k^2\Psi + 3aH\left(\dot\Psi + aH\Psi\right) &= 4\pi G a^2 \delta\mathcal{T}^0_0 \\ ik_i(\dot\Psi + aH\Psi) &= -4\pi G a \delta\mathcal{T}^0_i. \end{aligned} \tag{5.71}$$

Here I have simply copied the results from Chapter 4, replacing Φ with $-\Psi$.

To get the two equations we need, therefore, we must expand the energy-momentum tensor defined in equation [5.27] about the zero order solution $\phi^{(0)}$, thereby obtaining the first order piece, $\delta\mathcal{T}$. First let's compute $\delta\mathcal{T}^0_i$. Since the time-space components of the scalar metric are zero, the second set of terms in equation [5.27], those with prefactor g^α_β vanish. Therefore,

$$\mathcal{T}^0_i = g^{0\nu}\phi_{,\nu}\phi_{,i} \tag{5.72}$$

where I have returned to using $_{,\nu}$ to denote the derivative with respect to x^ν. Since $g^{0j} = 0$, the index ν must be equal to 0. Recall that the zero order field $\phi^{(0)}$ is homogeneous, so $\phi^{(0)}_{,i} = 0$. The space-time component of the energy-momentum tensor therefore has no zero order piece. To extract the first order piece, we can set $\phi_{,i}$ to $\delta\phi_{,i} = ik_i\delta\phi$, where $\delta\phi$ is the perturbation to the homogeneous scalar field. Then, setting all other factors to their zero order values leads to

$$\delta\mathcal{T}^0_i = -\frac{ik_i}{a}\dot\phi^{(0)}\delta\phi. \tag{5.73}$$

The factor of a enters the denominator here because $\phi^{(0)}_{,0} = \dot\phi^{(0)}/a$ (recall that $\dot{}$ is derivative with respect to conformal time). The time-space component of Einstein's equations is then

$$\dot\Psi + aH\Psi = 4\pi G\dot\phi^{(0)}\delta\phi. \tag{5.74}$$

The time-time component of the energy-momentum tensor is a little more difficult.

$$\mathcal{T}^0_0 = g^{00}(\phi_{,0})^2 - \frac{1}{2}g^{\alpha\beta}\phi_{,\alpha}\phi_{,\beta} - V. \tag{5.75}$$

Setting $\phi = \phi^{(0)} + \delta\phi$ leads to

$$\mathcal{T}^0_0 = \frac{1}{2}(-1+2\Psi)\left(\phi^{(0)}_{,0} + \delta\phi_{,0}\right)^2 - \frac{1}{2a^2}(1+2\Psi)\delta\phi_{,i}\delta\phi_{,i} - V(\phi^{(0)} + \delta\phi). \tag{5.76}$$

The spatial derivatives come in pairs, and pairs of first order variables $(\delta\phi_{,i})$ lead to second order terms. These may therefore be neglected. The potential may be expanded as a zero order term, $V(\phi^{(0)})$ plus a first order correction, $V'\delta\phi$, so the first order correction to the energy-momentum tensor is

$$\begin{aligned} \delta\mathcal{T}^0_0 &= \Psi(\phi^{(0)}_{,0})^2 - \phi^{(0)}_{,0}\delta\phi_{,0} - V'\delta\phi \\ &= \frac{\Psi(\dot\phi^{(0)})^2 - \dot\phi^{(0)}\dot\delta\phi}{a^2} - V'\delta\phi. \end{aligned} \tag{5.77}$$

With this expression for the first order energy-momentum tensor, we can now write Einstein's time-time equation as

$$k^2\Psi + 3aH\left(\dot{\Psi} + aH\Psi\right) = 4\pi G\left[\Psi(\dot{\phi}^{(0)})^2 - \dot{\phi}^{(0)}\dot{\delta\phi} - a^2V'\delta\phi\right]. \tag{5.78}$$

We can get this in a slightly simpler form by using the time-space component equation [5.74] to write the term in parentheses on the left as $4\pi G\dot{\phi}^{(0)}\delta\phi$. Then, moving the Ψ term on the right over to the left leads to

$$\left(k^2 - 4\pi G(\dot{\phi}^{(0)})^2\right)\Psi = -4\pi G\left[\dot{\phi}^{(0)}\dot{\delta\phi} + (3aH\dot{\phi}^{(0)} + a^2V')\delta\phi\right]. \tag{5.79}$$

5.5.2 Back to the Harmonic Oscillator

We now have two first order equations, [5.74] and [5.79], for Ψ and $\delta\phi$. We want to turn these two first order equations into one second order equation, one resembling the harmonic oscillator equation. This requires quite a bit of algebra. It proves convenient to define a second "slow-roll" variable to complement ϵ:

$$\begin{aligned}\delta &\equiv \frac{-1}{aH\dot{\phi}^{(0)}}\left[aH\dot{\phi}^{(0)} - \ddot{\phi}^{(0)}\right] \\ &= \frac{-1}{aH\dot{\phi}^{(0)}}\left[3aH\dot{\phi}^{(0)} + a^2V'\right].\end{aligned} \tag{5.80}$$

Here the paucity of Greek letters becomes a hindrance. The second slow roll parameter is more conventionally defined as η, but we obviously cannot follow that convention as η is our conformal time. (Early universe cosmologists use τ for conformal time, freeing up η, but we do not have that luxury since we need τ for optical depth.) My choice of δ is also fairly common, but we need to bear in mind that this has nothing to do with the overdensities introduced in Chapter 3. The second line here follows from equation [5.36], but with a twist. The derivatives there are with respect to time t, while dots here denote derivatives with respect to conformal time η. You should convince yourself that, when written in terms of conformal time derivatives, equation [5.36] maintains its form, but with the factor of three in the second term replaced by a two. In what follows, we will also find useful the relation

$$4\pi G(\dot{\phi}^{(0)})^2 = \epsilon a^2H^2 \tag{5.81}$$

easily obtained with the definition of ϵ in equation [5.37] by differentiating the Hubble rate as given by Einstein's equation ($H = [8\pi G\rho/3]^{1/2}$).

With this definition and expression, the two evolution equations – [5.74] and [5.79] – become

$$\begin{aligned}\dot{\Psi} + aH\Psi &= 4\pi G\dot{\phi}^{(0)}\delta\phi \\ \left(1 - \frac{k^2}{4\pi G(\dot{\phi}^{(0)})^2}\right)\Psi &= \frac{\dot{\delta\phi}}{\dot{\phi}^{(0)}} + \frac{aH\delta}{\dot{\phi}^{(0)}}\delta\phi.\end{aligned} \tag{5.82}$$

We now eliminate Ψ in favor of the variable u defined as

$$u \equiv a\left(\delta\phi + \frac{\dot{\phi}^{(0)}}{aH}\Psi\right). \tag{5.83}$$

This implies that

$$\Psi = \frac{aH}{\dot{\phi}^{(0)}}\left(\frac{u}{a} - \delta\phi\right) \tag{5.84}$$

and its derivative with respect to conformal time is

$$\begin{aligned}\dot{\Psi} &= -[\delta + \epsilon]\frac{a^2H^2}{\dot{\phi}^{(0)}}\left(\frac{u}{a} - \delta\phi\right) \\ &+ \frac{aH}{\dot{\phi}^{(0)}}\left(\frac{\dot{u}}{a} - Hu - \dot{\delta\phi}\right).\end{aligned} \tag{5.85}$$

In terms of u and $\delta\phi$ the two evolution equations become

$$\dot{u} - (aH[\delta + \epsilon])\,u = (1-\delta)a^2H\delta\phi + a\dot{\delta\phi} \tag{5.86}$$

$$\left(1 - \frac{k^2}{a^2H^2\epsilon}\right)u = \frac{\dot{\delta\phi}}{H} + a\delta\phi\left[1 - \frac{k^2}{a^2H^2\epsilon} - \delta\right]. \tag{5.87}$$

Equations [5.86] and [5.87] are two first order evolution equations which we can rewrite as one second order equation for u. Here is the strategy: (i) eliminate $\dot{\delta\phi}$ in equation [5.86] by using equation [5.87]; (ii) differentiate the resulting equation [5.86] (which now has no $\dot{\delta\phi}$ terms) with respect to conformal time; then (iii) again eliminate $\dot{\delta\phi}$ everywhere it now occurs using equation [5.87]; and finally (iv) eliminate $\delta\phi$ everywhere using (the undifferentiated, $\dot{\delta\phi}$-less) equation [5.86].

Carrying out step (i) and dividing by aH leads to

$$\frac{1}{aH}\dot{u} - \left(\delta + \epsilon + 1 - \frac{k^2}{a^2H^2\epsilon}\right)u = \frac{k^2}{aH^2\epsilon}\delta\phi. \tag{5.88}$$

We now differentiate this with respect to η, using the fact that $\dot{\epsilon} = 2aH\epsilon(\epsilon + \delta)$ (also see the relation in Problem 4.11). Thus,

$$\begin{aligned}\frac{1}{aH}\ddot{u} - \left(2 + \delta - \frac{k^2}{a^2H^2\epsilon}\right)\dot{u} &- \left(\dot{\delta} + \dot{\epsilon} + \frac{2k^2(1+\delta)}{aH\epsilon}\right)u \\ &= \dot{\delta\phi}\left(\frac{k^2}{aH^2\epsilon}\right) - \frac{k^2}{H\epsilon}\delta\phi\,(1+2\delta).\end{aligned} \tag{5.89}$$

Eliminating $\dot{\delta\phi}$ by using equation [5.87] leads to

$$\frac{1}{aH}\ddot{u} - \left(2 + \delta - \frac{k^2}{a^2H^2\epsilon}\right)\dot{u} - \left(\dot{\delta} + \dot{\epsilon} + \frac{k^2(3+2\delta)}{aH\epsilon} - \frac{k^4}{a^3H^3\epsilon^2}\right)u$$

$$= \delta\phi\left(\frac{k^2}{H\epsilon}\right)\left(-2-\delta+\frac{k^2}{a^2H^2\epsilon}\right)$$

$$= -\left(2+\delta-\frac{k^2}{a^2H^2\epsilon}\right)\left(\dot{u}-aH\left(\delta+\epsilon+1-\frac{k^2}{a^2H^2\epsilon}\right)u\right) \tag{5.90}$$

where the last line uses equation [5.88], thereby completing the final step. The coefficients of $\dot{u}$ cancel on the left and right as do the coefficients of the k^4u terms. A little more work is needed to see that the coefficient of the k^2u term is $1/aH$, just like the $\ddot{u}$ term. Collecting terms, and neglecting the time derivatives of the slow roll parameters, we see that the second order equation governing u is

$$\ddot{u}+\left(k^2-\frac{2+3\delta+6\epsilon]}{\eta^2}\right)v=0. \tag{5.91}$$

Here I have dropped terms higher order in ϵ, δ.

We can now estimate the power spectrum of fluctuations in u, the linear combination of ϕ and Ψ defined in equation [5.83]. The equation governing u is identical to the tensor evolution equation ([5.64]), the only difference being a different index ν. Here, to first order in the slow roll parameters the index ν is equal to

$$\nu=\frac{3}{2}+\delta+2\epsilon. \tag{5.92}$$

Therefore, in complete analogy with tensor fluctuations, by squaring equation [5.68], we find

$$P_u(k,\eta)=\left(\frac{2^\nu\Gamma(\nu)}{4\Gamma(3/2)}\right)^2\frac{1}{k^{2\nu}(-\eta)^{2\nu-1}}, \tag{5.93}$$

where I have taken the $-k\eta << 1$ limit of $|u|^2$. We will shortly relate the scalar power spectrum, which characterizes fluctuations in Ψ, to this power spectrum. For now, note that for the same reasons that $v \propto a$ in the super-horizon limit, here too $u \propto a$ once a mode leaves the horizon.

5.5.3 Scalar Power Spectrum

There is one final hurdle to cross before we write down the power spectrum for fluctuations of the gravitational potential Φ. We know what the fluctuation spectrum looks like for u, which is a linear combination of Ψ (or equivalently Φ since they are equal and opposite during inflation) and $\delta\phi$. The problem is that long after the modes of interest leave the horizon, inflation ends. Typically, the end of inflation occurs when the energy density in ϕ transforms into ordinary matter and radiation. The way that this happens is still a topic of active research. However, most likely the transition from ϕ-energy to ordinary matter-radiation energy (or *reheating*) does not affect the perturbation spectrum, so we will not discuss it. What we must discuss, though, is what happens to u once the energy density in ϕ goes away.

The way to resolve this issue is to recognize that the $\delta\phi$ piece that appears in u is related to a component of the stress-energy tensor. Let us therefore, re-express u in terms of Ψ and the stress energy tensor. This new expression then will be valid even after the vacuum energy in ϕ transforms into matter and radiation. Using equation [5.73], we can write

$$ik_iT_i^0=\frac{k^2a(\rho+P)\delta\phi}{\dot{\phi}^{(0)}} \tag{5.94}$$

where I have used the fact that $\rho+P=(\dot{\phi}^{(0)})^2/a^2$ for the scalar field. Therefore, in terms of the stress-energy tensor, our harmonic oscillator-like variable u can be re-expressed as

$$u=\frac{\dot{\phi}^{(0)}}{H}\left(\frac{ik_iT_i^0H}{k^2(\rho+P)}+\Psi\right). \tag{5.95}$$

Recall that outside the horizon, $u \propto a$; but equation [5.81] tells us that $\phi^{(0)}/H$ also scales as a. We might expect therefore the term in parentheses in equation [5.95] to remain constant outside the horizon. In fact, a careful analysis first performed by Bardeen (1980) shows that this is true, that the combination,

$$-\zeta\equiv\frac{ik_iT_i^0H}{k^2(\rho+P)}+\Psi \tag{5.96}$$

remains contant outside the horizon. This means that $P_\zeta=(H/\dot{\phi}^{(0)})^2P_u$ remains constant outside the horizon.

The final step in connecting P_u with P_Φ is to find ζ in the post-inflation, radiation dominated era. To do this, we must compute $k_iT_0^i$ for radiation. You already did this in Chapter 4 in Problem (4). The result is

$$k_iT_i^0=-4iak\rho_R\Theta_1 \tag{5.97}$$

where ρ_R is the energy density of all radiation and Θ_1, the dipole of the photon distribution, is the same as all the other dipoles (of e.g. neutrinos and all other massless particles). Also, the pressure for radiation is equal to a third of the energy density, so

$$\zeta = -\frac{3aH\Theta_1}{k}-\Psi$$

$$=\frac{3}{2}\Phi. \tag{5.98}$$

The second equality follows from the initial conditions relating the dipole to the potential (equation [5.15]) and the fact that $\Psi=-\Phi$.

We can now relate the power spectrum of Φ to that of u. Combining equation [5.98] with the definition of ζ gives

$$P_\Phi=\frac{4}{9}P_\zeta=\frac{4}{9}\left(\frac{H}{\dot{\phi}^{(0)}}\right)^2P_u. \tag{5.99}$$

Just as with tensor modes, P_Φ remains constant after the mode leaves the horizon. Evaluating P_u in equation [5.93] at $-k\eta = 1$ leads to

$$P_\Phi(k) = \frac{2}{9}\left(\frac{2^{\nu-3/2}\Gamma(\nu)}{\Gamma(3/2)}\right)^2 \frac{H^2}{k(\dot{\phi}^{(0)})^2}\bigg|_{-k\eta=1}. \tag{5.100}$$

We see that the amplitude of scalar perturbations is largest if the field driving inflation is rolling very slowly down its potential. Note that in the slow roll approximation, the index $\nu \simeq 3/2$ so the ratio in the parentheses in equation [5.100] can be set to one. Another way to write the power spectrum of scalar perturbations is to eliminate $\dot{\phi}^{(0)}$ here by using equation [5.81]. In fact, since we want to evaluate all terms at $-k\eta = 1$, $(\dot{\phi}^{(0)}/H)^2$ can be set to $\epsilon a^2/4\pi G = \epsilon k^2/4\pi G H^2$. So

$$P_\Phi(k) = \frac{8\pi}{9k^3}\frac{H^2}{\epsilon m_{\rm Pl}^2}|_{-k\eta=1}. \tag{5.101}$$

Comparing to equation [5.70], we see that the ratio of scalar to tensor modes is of order $1/\epsilon$; that is, we expect scalar modes to dominate. Finally, another way of writing the scalar power spectrum is to eliminate ϵ in favor of the potential and its derivative, using the result of Problem (13),

$$P_\Phi(k) = \frac{128\pi^2}{9k^3}\left(\frac{H^2V^2}{m_{\rm Pl}^4 V'^2}\right)\bigg|_{-k\eta=1}. \tag{5.102}$$

5.6 Summary and Spectral Indices

In order to understand how scales which should be uncorrelated today are observed to have almost identical temperatures, we are virtually forced into the theory of inflation. In addition to explaining away the nagging fine-tuning problems of the standard cosmology, inflation is also a mechanism for generating primordial perturbations over the smooth universe.

Inflation predicts that quantum mechanical perturbations in the very early universe are first produced when the relevant scales are causally connected. Then these scales are wisked outside the horizon by inflation, only to re-enter much later to serve as initial conditions for the growth of structure and anisotropy in the universe. The perturbations are drawn from a Gaussian distribution with mean zero and variance given by the power spectrum. In the case of scalar perturbations, the ones of most importance for us, the power spectrum is given by equation [5.101]. Perturbations to the tensor part of the metric are also produced, are also Gaussian with mean zero; the power spectrum of tensor modes is given by equation [5.70]. In both of these expressions the slow roll parameter ϵ is, as defined in equation [5.37], proportional to the derivative of the Hubble rate. Since the Hubble rate is close to constant during inflation – due to the dominance of potential energy – ϵ is typically small. As a consistency check, we showed that its time derivative is quadratic in the slow roll parameters. In the two formulae for the power spectra, the index ν is defined slightly differently. For tensors it depends only on ϵ (equation [5.63]). In the scalar case, it depends also on a second slow roll parameter, δ defined in equation [5.80]; the dependence is given in equation [5.92].

A spectrum in which $k^3P(k)$ is constant (i.e. does not depend on k) is called a *scale-invariant* or *scale-free* spectrum. Apart from small deviations encoded in the slow roll parameters, both the scalar and the tensor perturbations are scale-free. This is both a blessing and a curse. It is good because it is a fairly definite prediction, easy to test. It is unfortunate because a scale-free spectrum is what one might have exected even without the complex machinery of inflation. Indeed, a scale-free spectrum is also referred to as a Harrison-Zel'dovich-Peebles spectrum, crediting the smart people who first proposed it as the appropriate distribution for the initial conditions, a proposal that pre-dates inflation by many years. This really is too bad, because if we observe a scale-free spectrum, and most present observations are consistent with this, then inflation cannot fairly claim all the credit. However, if we observe a small mixture of tensor modes and/or a small deviation from a scale-free spectrum, then this will go a long way towards convincing skeptics that inflation is responsible for the primordial perturbations.

To quantify the deviations from scale invariance, it is conventional to write the primordial power spectra as

$$\begin{aligned} P_\Phi(k) &= \frac{8\pi}{9k^3}\frac{H^2}{\epsilon m_{\rm Pl}^2}|_{-k\eta=1} \equiv \frac{50\pi^2}{9k^3}\left(\frac{k}{H_0}\right)^{n-1}\delta_H^2 \\ P_h(k) &= \frac{2^{2\nu}\pi\Gamma^2(\nu)}{k^3\Gamma^2(3/2)}(1-\epsilon)^2\frac{H^2}{m_{\rm Pl}^2}|_{-k\eta=1} \equiv A_T k^{n_T-3}. \end{aligned} \tag{5.103}$$

A scale-free scalar spectrum therefore corresponds to $n = 1$, while $n_T = 0$ for a scale-free tensor spectrum. The amplitudes of the perturbations are determined by δ_H (subscript $_H$ for amplitude at horizon crossing) and A_T. The simplest way to compute the index n in terms of the slow roll parameters is realize that

$$\frac{d\ln[P_\Psi]}{d\ln k} = n - 4. \tag{5.104}$$

The derivatives with respect to k are slightly tricky because the $k-$ dependence is hidden in the fact that ϵ and δ are to be evaluated at $-k\eta = 1$. As an example,

$$\begin{aligned} \frac{d\ln H|_{-k\eta=1}}{d\ln k} &= \frac{k}{H}\frac{dH}{d\eta}|_{-k\eta=1} \times \frac{d\eta|_{-k\eta=1}}{dk} \\ &= \frac{k}{H}\left(-aH^2\epsilon\right) \times \frac{1}{k^2}. \end{aligned} \tag{5.105}$$

To lowest order in slow-roll parameters, aH can be set to $-1/\eta = k$, so the logarithmic derivative of H is $-\epsilon$. The only other part of equation [5.101] that contributes at first order in the slow roll parameters is $d\ln(\epsilon^{-1})/d\ln k$, which by the same technique can be shown to be $-2(\epsilon+\delta)$. So, to lowest order,

$$n = 1 - 4\epsilon - 2\delta. \tag{5.106}$$

The tensor case is even easier since only the derivative of H breaks scale invariance at lowest order. Thus,

$$n_T = -2\epsilon. \tag{5.107}$$

The fact that the tensor index n_T is proportional to ϵ leads to one of the robust predictions of inflation. Many inflationary models have been proposed which offer different predictions for ϵ and δ. Almost all of these however maintain the feature that the ratio of tensor to scalar modes (which we saw above was proportional to ϵ) is directly related to the tensor spectral index (here also seen to be directly proportional to ϵ). As you progress through this book, moving from the evolution of anisotropies to their analyses, try to bear in mind the crucial question of whether this prediction can be put to the observational test.

The slow roll parameters are a convenient way to summarize the predictions of an inflationary model. However, ultimately we are interested in the physics, so we are interested in how these parameters relate back to the fundamental entity, the potential V of the scalar field responsible for inflation. You will show in Problem (4.12) that these parameters can be expressed in terms of the potential and its derivatives. Therefore, extracting the values of ϵ and δ from the data is tantamount to probing the potential of the field driving inflation. Given that the expected scale of this potential is on the order of 10^{16} GeV (Problem 4.13), this is quite an impressive probe!

Suggested Reading

The thirty or so pages on inflation in this chapter, which were heavily slanted towards production of perturbations, offer but a glimpse into the many facets of this remarkable theory. Recently, Guth wrote a popular account of his discovery of inflation, *The Inflationary Universe*. One of the other originators of the theory, Linde, has a more technical book, *Inflation and Quantum Cosmology*, which emphasizes model building much more than I have here. As I mentioned earlier, *The Early Universe* (Kolb and Turner) has an excellent chapter on inflation. The recent *Cosmological Inflation and Large Scale Structure* (Liddle and Lyth) is most similar in spirit to this book, with a heavy emphasis on perturbations. The discussion there of the perturbation spectrum is laden with less algebra than the one in §5.5 so is worth reading. (Beware that their Planck mass is our $m_{\rm Pl}/\sqrt{8\pi}$.)

An extremely clear and deep look into inflation is given in *300 Years of Gravitation* (ed. Hawking and Israel) in the article by Blau and Guth. Many other articles in that thick compilation volume are also fascinating. The initial article by Guth (1981) is completely accessible and as clear a statement possible of the problems that led to inflation and the initial attempt (old inflation) to solve them. Much of the treatment in §5.4 and §5.5 is based on the eight page paper of Stewart and Lyth (1993), a remarkably concise treatment of the perturbation spectrum.

The initial conditions relating the various perturbations described in §5.1 are perhaps most clearly discussed in the review article by Efstathiou (1990). Isocurvature perturbations, for the most part ignored here, are treated in detail there.

I will no doubt be lambasted by some for ignoring the possibility of perturbations produced by topological defects. At the risk of alienating these folks further, I claim that to the extent that these theories make predictions, the predictions do not seem in accord with observations. A more diplomatic response is that there exist many books with comprehensive discussions of topological defects. Among them are *Cosmic Strings and Other Topological Defects* (Vilenkin and Shellard) and *The Formation and Evolution of Cosmic Strings* (ed. Gibbons, Hawking, and Vachaspati).

Problems

5.1 Find the ratio of neutrino to radiation energy density, f_ν. Assume that there are three species of massless neutrinos.

5.2 Account for the neutrino quadrupole moment when setting up initial conditions.

(a) Start with equation [3.103]. This is an equation for $\Theta_\nu(\mu)$. Turn this into a hierarchy of equations for the neutrino moments:

$$\begin{aligned}
\dot{\Theta}_{\nu,0} + k\Theta_{\nu,1} &= -\dot{\Phi} \\
\dot{\Theta}_{\nu,1} - \frac{k}{3}\left(\Theta_{\nu,0} - 2\Theta_{\nu,2}\right) &= \frac{k}{3}\Psi \\
\dot{\Theta}_{\nu,2} - \frac{2}{5}k\Theta_{\nu,1} &= 0.
\end{aligned} \tag{5.108}$$

To do this, you need to recall the definition of these moments, which is equivalent to that for photons, equation [3.104]. A good way to reduce equation [3.103] into this hierarchy is to multiply it first by P_0 and then integrate over $\int_{-1}^{1} d\mu$. This leads to the first equation above. Then multiply equation [3.103] by P_1 to get the second and P_2 to get the third. More details are shown in §7.3, where we go through the same exercise for the photon moments. In the third equation you may neglect $\Theta_{\nu,3}$ because it is smaller than $\Theta_{\nu,2}$ by a factor of order $k\eta$ (prove this!).

(b) Eliminate $\Theta_{\nu,1}$ from these equations and show that

$$\ddot{\Theta}_{\nu,2} = \frac{2k^2}{15}\left(\Psi + \Theta_{\nu,0} - 2\Theta_{\nu,2}\right). \tag{5.109}$$

Drop $\Theta_{\nu,2}$ on the right hand side because it is much smaller than $\Psi + \Theta_{\nu,0}$.

(c) Rewrite Einstein's equation, [4.32], as

$$\Theta_{\nu,2} = -(k\eta)^2 \frac{\Phi + \Psi}{12 f_\nu}. \tag{5.110}$$

This neglects the photon quadrupole. Argue that Compton scattering sets $\Theta_2 << \Theta_{\nu,2}$ so this is a reasonable assumption.

(d) Now differentiate this form of Einstein's equation twice to get an expression for $\ddot{\Theta}_{\nu,2}$. Equate this to the expression for $\ddot{\Theta}_{\nu,2}$ derived in part (b). (You may drop all derivatives of Φ and Ψ when doing this since the mode of interest is the $p = 0$ constant mode.) Use this equation to express $\Theta_{\nu,0}$ in terms of Φ and Ψ.

(e) Finally assume that $\Theta_0 = \Theta_{\nu,0}$ and use your expression for $\Theta_{\nu,0}$ to rewrite equation [5.11] as a relation between the two gravitational potentials. Show that this relation is

$$\Phi = -\Psi\left(1 + \frac{2f_\nu}{5}\right). \tag{5.111}$$

5.3 Show that the initial conditions for the velocities and dipoles of matter and radiation are as given in equation [5.15].

5.4 Inflation also solves the *flatness* problem. This is the question of why the energy density today is so close to critical.

(a) Suppose that

$$\Omega(t) \equiv \frac{8\pi G\rho(t)}{3H^2(t)} \tag{5.112}$$

is equal to 0.3 today, where ρ counts the energy density in matter and radiation (assume zero cosmological constant). From equation [1.3], find $\Omega(t) - 1$ in terms of the present curvature density $\Omega_K \equiv 1 - \Omega_m - \Omega_R$. How close to one would $\Omega(t)$ have been back at the Planck epoch (assuming no inflation took place so that the scale factor at the Planck epoch was of order 10^{-32})? This fine-tuning of the initial conditions is the

flatness problem. If not for the fine tuning, an open universe would be *obviously* open (i.e. Ω would be almost exactly zero) today.

(b) Now show that inflation solve the flatness problem. Extraplotate $\Omega(t) - 1$ back to the end of inflation, and then through sixty e-folds of inflation. What is $\Omega(t) - 1$ right before these sixty e-folds of inflation?

5.5 Another way of looking at the problems that inflation solves is to consider the enotropy within our Hubble volume. This is proportional to the total number of particles in the volume, with a proportionality constant of order unity. How many photons are there within our Hubble volume today? Explain how inflation produces entropy this large.

5.6 We showed that, if the universe was always dominated by ordinary matter or radiation early on, then the comoving horizon when the scale factor was a_e (very small) was $a_0 H_0 / a_e H_e$ times the comoving Hubble radius today. Compute this ratio assuming that the temperature was equal to 10^{15} GeV at a_e. Account for the radiation to matter transition at $a \sim 10^{-4}$.

5.7 Consider a free, homogeneous scalar field with mass m. The potential for this field is $V = m^2\phi^2/2$. Show that, if $m >> H$, the scalar field oscillates around zero with frequency equal to its mass. Also show that its energy density falls off as a^{-3}, so it behaves exactly like ordinary non-relativistic matter. To do this, use equation [5.36], take an ensatz of the form $\phi = A(t)\cos(\omega t)$, and assume that $\omega >> (dA/dt)/A$. Show at the end that this assumption is valid. This process is equivalent to the WKB approximation introduced in quantum mechanics.

5.8 Compute some well-known properties of the quantized harmonic oscillator.

(a) The momentum of the harmonic oscillator with unit mass is $p = dx/dt$. Compute

$$[\hat{x}, \hat{p}]$$

and show that it is equal to i. You can obtain the operator $\hat{p}$ by differentiating $\hat{x}$ (equation [5.42]) with respect to time.

(b) Compute the zero point energy of the harmonic oscillator with unit mass. Do this by quantizing the energy

$$E = \frac{p^2}{2} + \frac{\omega^2 x^2}{2}$$

and then computing its expectation value in the ground state: $< 0|\hat{E}|0 >$.

5.9 Show that $(-\eta)^{1/2-\nu}$ scales as a, proving that the limit in equation [5.68] is correct. A simple way to do this is to evaluate the derivative

$$\frac{d}{d\eta}\frac{(-\eta)^{1/2-\nu}}{a}$$

and show that it is equal to zero.

5.10 Evaluate P_h when $k/aH = 1$ instead of when $-k\eta = 1$ as we did in the text. Show that this changes the factor of $(1-\epsilon)^2$ in equation [5.70] to $(1-\epsilon)^{2/(1-\epsilon)}$. Of course this change is illusory since it is compensated by evaluating H at a slightly different time.

5.11 Show that gravity waves are not sourced by the scalar field during inflation. To do this, recall that the right hand side of equation [5.49] is

$$\delta\mathcal{T}_{11} - \delta\mathcal{T}_{22}$$

where $\delta\mathcal{T}$ is the perturbation to the stress-energy tensor (assumed to be dominated by ϕ) and, as in the derivation of equation [4.62], I have chosen $\vec{k}$ to be in the $\hat{z}$ direction. Show that this right hand side is indeed zero for the scalar field.

5.12 Show that

$$\frac{d}{d\eta}\left(\frac{1}{aH}\right) = \epsilon - 1$$

5.13 Express the slow roll parameters ϵ and η in terms of the potential V and its derivatives with respect to ϕ. Show that, to lowest order,

$$\epsilon = \frac{1}{16\pi G}\left(\frac{V'}{V}\right)^2$$

and

$$\delta = \epsilon - \frac{1}{8\pi G}\frac{V''}{V}$$

where primes denote derivatives with respect to $\phi^{(0)}$.

5.14 Determine the predictions of an inflationary model with a quartic potential,

$$V(\phi) = \lambda\phi^4.$$

(a) Compute the slow roll parameters ϵ and δ in terms of ϕ.

(b) Determine ϕ_e, the value of the field at which inflation ends by setting $\epsilon = 1$ at the end of inflation.

(c) To determine the spectrum, you will need to evaluate ϵ and δ at $-k\eta = 1$. Choose the wavenumber k to be equal to $a_0 H_0$, roughly the horizon today. Show that the requirement $-k\eta = 1$ then corresponds to

$$e^{60} = \int_0^N dN' \frac{e^{N'}}{(H(N')/H_e}$$

where H_e is the Hubble rate at the end of inflation, and N is defined to be the number of e-folds before the end of inflation:

$$N \equiv \ln\left(\frac{a_e}{a}\right).$$

(d) Take the Hubble rate to be a constant in the above with H/H_e equal to one. This imples that $N \simeq 60$. Turn this into an expression for ϕ. The simplest way to do this is to note that $N = \int_t^{T_e} dt' H(t')$ and assume that H is dominated by potential energy. Show that this mode leaves the horizon when $\phi^2 = 60 m_{\rm Pl}^2/\pi$

(e) Determine the predicted values of n and n_T.

(f) Estimate the scalar amplitude in terms of λ. As a rough estimate, assume that $k^3 P(k)$ for this mode is equal to 10^{-9} (very roughly the large scale anisotropy). What value does this imply for λ.

This model illustrates many of the features of contemporary models. In it, (i) the field is of order – even greater than – the Planck scale, but (ii) the energy scale is much smaller due to (iii) the very small coupling constant.

Chapter 6

Inhomogeneities

Having set up the system of equations to be solved and the initial conditions for the perturbations, we can now calculate the inhomogeneities and anisotropies in the universe. In this first solutions chapter, we start with the perturbations to the dark matter. In principle these are coupled to all other perturbations. In practice, though, perturbations to the dark matter depend very little on the details of the radiation perturbations. Dark matter, by definition, is affected by radiation only indirectly, through the gravitational potentials. At late times, when the universe is dominated by matter, these potentials are independent of the radiation. At early times, while it is true that the potentials are determined by the radiation, it is also true that the radiation perturbations are relatively simple, so that all moments beyond the monopole and dipole can be neglected. The converse is not true, as we will see in the next chapter: To treat the anisotropies properly we will need to know how the matter perturbations behave.

The ultimate goal of this exercise is to compare theory with observations. We will solve for the evolution of each Fourier mode, $\delta_{\rm DM}(k,\eta)$. Given this solution, and the initial power spectrum generated by inflation, we can construct the power spectrum of matter today. At least on large scales, this is the most important observable. On small scales, comparison with observation today is more difficult: one must worry about non-linearities and gas dynamics when comparing with the galaxy distribution. Nonetheless, even on small scales, the linear power spectrum, which we compute in this chapter, is often the starting point for any quantitative statement about the distribution of matter.

6.1 Prelude

Gravitational instability is a powerful idea, easy to understand, and most likely reposnsible for the structure in our universe. As time evolves, matter accumulates in initially overdense regions. It doesn't matter how small the initial overdensity was (e.g. in typical cosmological scenarios, the overdensity was of order a part in a hundred thousand); eventually enough matter will be attracted to the region to form structure.

The $F = ma$ of gravitational instability is the equation governing overdensities δ. Schematically, it reads

$$\ddot{\delta} + [\text{Pressure} - \text{ Gravity}]\,\delta = 0. \tag{6.1}$$

These basic forces, depicted in figure 6.1 act in opposite directions. For, gravity acts to increase overdensities, grabbing more matter into the region. Since there are more particles in an overdense region, random thermal motion causes a net loss of mass in an overdense region. Therefore, if pressure is strong, inhomogeneities do not grow. As indicated by the cartoon equation [6.1], if pressure is low, δ grows exponentially; if it is large, δ oscillates with time.

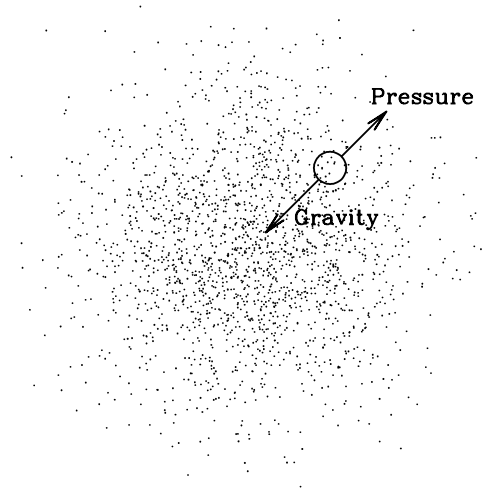


Figure 6.1: Gravitational instability. Mass near an overdense region is attracted to the center by gravity but repelled by pressure. If the region is dense enough, gravity wins and the overdensity grow with time.

We will see many manifestations of the simple form of gravitational instability depicted in equation [6.1]. Different ambient cosmological conditions alter the growth rate. For example, in a matter dominated universe, δ grows only as a power of time, not exponentially, while in a radiation dominated universe, the growth is but logarithmic. We will treat super-horizon versions of this equation as well as the more familiar sub-horizon version. When going though the math, though, it is useful to bear in mind the dueling concepts of gravity and pressure.

6.1.1 Three Stages of Evolution

The evolution of cosmological perturbations breaks up naturally into three stages. To see this, let's cheat and look at the solutions for several different modes. Figure 6.2 shows the gravitational potential as a function of scale factor for a long, medium, and short wavelength

mode. Early on, all of the modes are outside the horizon ($k\eta << 1$) and the potential is constant. At intermediate times (shaded in the figure), two things happen: the wavelengths fall within the horizon and the universe evolves from radiation domination ($a << a_{EQ}$) to matter domination ($a >> a_{EQ}$). Without getting into the details, we see that the order of these epochs (a_{EQ} and the epoch of horizon crossing) greatly affects the potential. The large scale mode, which enters the horizon well after a_{EQ}, evolves much differently than the small scale mode, which enters the horizon before equality. Finally, at late times, all the modes evolve identically again, in this case (where $\Omega_m = 1$) remaining constant.

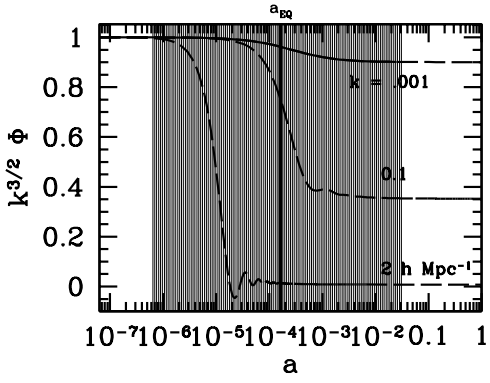


Figure 6.2: The linear evolution of the gravitional potential Φ. Dashed line denotes that the mode has entered the horizon. Evolution through the shaded region is described by the transfer function. The potential is unnormalized, but the relative normalization of the three modes is as it would be for scale invariant perturbations. Here baryons have been neglected, $\Omega_m = 1$, and $h = 1/2$.

We are able to observe the distribution of matter predominantly at late epochs, in the third stage of evolution, when all modes are evolving identically. If we wish to relate the potential during these times to the primordial potential set up during inflation, and we do, we can write schematically

$$\Phi(k,a) = \Phi(k, a_{\rm early}) \times \Big\{\text{Transfer Function}(k)\Big\} \times \Big\{\text{GrowthFunction}(a)\Big\}. \qquad (6.2)$$

The transfer function describes the evolution of perturbations through the epochs of horizon cossing and radiation/matter transition (the shaded region in figure 6.2), while the growth factor describes the wavelength-independent growth at late times. This schematic equation is indeed roughly how the growth factor and the transfer function are defined, with two caveats, both concerning convention. Notice from figure 6.2 that even the largest wavelength perturbations decline slightly as the universe passes through the epoch of equality. This

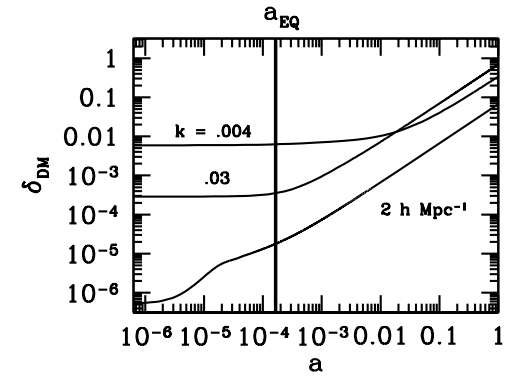


Figure 6.3: The evolution of perturbations to the Dark Matter.

decline is conventionally removed so that the transfer function on large scales is equal to one. Therefore, the transfer function is defined as

$$T(k) \equiv \frac{\Phi(k, a_{\rm late})}{\Phi_{\rm Large\ Scale}(k, a_{\rm late})} \qquad (6.3)$$

where $a_{\rm late}$ denotes an epoch well after the transfer function regime and the *Large Scale* solution is the primordial Φ decreased by a small amount. We will derive in §6.2 that – neglecting anisotropic stresses– this factor is equal to (9/10). The second caveat concerns the growth function. The ratio of the potential to its value right after the transfer function regime is defined to be

$$\frac{\Phi(a)}{\Phi(a_{\rm late})} \equiv \frac{D_1(a)}{a} \qquad (a > a_{\rm late}), \qquad (6.4)$$

where D_1 is called the growth function. In the flat, matter dominated case depicted in figure 6.2, then, the potential is constant so $D_1(a) = a$. With these conventions, we have

$$\Phi(k,a) = \frac{9}{10}\Phi(k, a_{\rm early})T(k)\frac{D_1(a)}{a} \qquad (a > a_{\rm late}). \qquad (6.5)$$

The easiest way to probe the potential is to measure the matter distribution. Figure 6.3 shows the evolution of the matter overdensity for three different modes. Notice that at late times – when the potential is constant and all the modes within the horizon – the overdensity grows with the scale factor ($\delta_{\rm DM} \propto a$). This explains the seemingly odd nomenclature above (Why is it called a *growth* function if the potential remains constant?): D_1 describes the growth of the matter perturbations at late times. This growth is completely consistent with

our intuition that as time evolves, overdense regions attract more and more matter, thereby becoming more overdense.

We can now express the power spectrum of the matter distribution in terms of the primordial power spectrum generated during inflation, the transfer function, and the growth function. The simplest way to relate the matter overdensity to the potential at late times is to use Poisson's equation (the large k, no-radiation limit of equation [4.66])

$$\Phi = \frac{4\pi G\rho_{\rm DM} a^2 \delta_{\rm DM}}{k^2} \qquad (a > a_{\rm late}). \tag{6.6}$$

The density of dark matter is $\rho_{\rm DM} = \Omega_m \rho_{\rm cr}/a^3$, and $4\pi G\rho_{\rm cr} = (3/2)H_0^2$, so

$$\delta_{\rm DM}(k,a) = \frac{k^2 \Phi(k,a) a}{(3/2)\Omega_m H_0^2} \qquad (a > a_{\rm late}). \tag{6.7}$$

This, together with equation [6.5] allows us to relate the overdensity today to the primoridal potential

$$\delta_{\rm DM}(k,a) = \frac{3}{5}\frac{k^2}{\Omega_m H_0^2}\Phi(k, a_{\rm early}) T(k) D_1(a) a_{\rm late} \qquad (a > a_{\rm late}). \tag{6.8}$$

The primordial potential at a given $\vec{k}-$ mode is drawn from a Gaussian distribution with mean zero and variance (equation [5.103]) $P_\Phi = (50\pi^2/9k^3)(k/H_0)^{n-1}\delta_H^2$. So the power spectrum at late times is

$$P(k,a) = 2\pi^2 \delta_H^2 \frac{k^n}{\Omega_m^2 H_0^{n+3}} T^2(k) D_1^2(a) \qquad (a > a_{\rm late}). \tag{6.9}$$

Figure 6.4 shows the power spectrum today for two different models. Note that in both of the models $P \propto k$ on large scales, where the transfer function is unity. This behavior is apparent from equation [6.9] and corresponds to the simplest inflationary model, wherein $n = 1$. On small scales the power spectrum turns over. To understand this, look back at figure 6.2. The small scale mode there ($k = 2$ h Mpc^{-1}) enters the horizon well before matter/radiation equality. During the radiation epoch the potential decays, so the transfer function is much smaller than unity. The effect of this on matter perturbations can be seen in figure 6.3, where the growth of $\delta_{\rm DM}$ is retarded starting at $a \simeq 10^{-5}$ after the mode has entered the horizon and ending at $a \simeq 10^{-4}$ when the universe becomes matter dominated. Modes that enter the horizon even earlier undergo more suppression. Thus, the power spectrum is a decreasing function of k on small scales.

This leads to the realization that there will be a turnover in the power spectrum at a scale corresponding to the one which enters the horizon at matter/radiation equality. The power of this realization is apparent in figure 6.4, which shows two different models: one corresponding to a flat, matter dominated universe today (often called standard Cold Dark Matter or sCDM) and the other a universe with a cosmological constant today (Lambda Cold Dark Matter or ΛCDM). The major difference between the two models is that sCDM has more matter ($\Omega_m = 1$) and hence an earlier $a_{\rm EQ}$. An earlier $a_{\rm EQ}$ means fewer scales

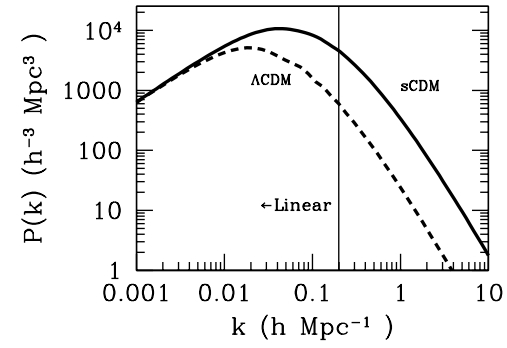


Figure 6.4: The power spectrum in two Cold Dark Matter models, with (ΛCDM) and without (sCDM) a cosmological constant. The spectra have been normalized to agree on large scales. The spectrum in the cosmological constant model turns over on larger scales because of a later $a_{\rm EQ}$. Scales to the left of the vertical line are still evolving linearly.

enter the horizon during the radiation dominated epoch, and therefore the turnover occurs on smaller scales. Finally, an important scale to keep in mind is the scale above which nonlinearities cannot be ignored. This is roughly $k_{\rm NL} \simeq 0.2$ h Mpc^{-1}. The power spectra shown in figure 6.4 are the linear power spectra today. On scales smaller than $k_{\rm NL}$, one cannot blindly compare the spectra from figure 6.4 with the matter distribution today.

6.1.2 Method

What are the evolution equations for the dark matter overdensity? In principle, these are the full set of Boltzmann equations derived in Chapter 2 and the pair of Einstein equations from Chapter 3. In practice, though, the full set of equations is not needed. To understand why, recall that early on ($a < a_*$), the photon distribution can be characterized by only two moments, the monopole Θ_0 and the dipole Θ_1. All other moments are suppressed because the photons are tightly coupled to the electron/proton gas. After decoupling this ceases to be true and, to completely characterize the photon distribution we will need to follow high moments. However, for the purposes of the matter distribution, what the photons are doing after a_* is irrelevant. For, by that time, which is typically well into the matter era, the potential is dominated by the dark matter itself. To sum up then, we can neglect all photon moments except for the monopole and dipole when we are considering the evolution of the matter distribution.

Neglecting the higher radiation moments, the four relevant Boltzmann equations become

$$\dot{\Theta}_{R,0} + k\Theta_{R,1} = -\dot{\Phi} \tag{6.10}$$

$$\dot{\Theta}_{R,1} - \frac{k}{3}\Theta_{R,0} = \frac{-k}{3}\Phi \tag{6.11}$$

$$\dot{\delta}_{\rm DM} + ikv_{\rm DM} = -3\dot{\Phi} \tag{6.12}$$

$$\dot{v}_{\rm DM} + \frac{\dot{a}}{a}v_{\rm DM} = ik\Phi \tag{6.13}$$

Even with the assumption that only the monopole and dipole are retained, getting from equation [3.96] to equations [6.10] and [6.11] requires some explanation and work. First, the explanation: The subscript $_R$ here refers to radiation, both neutrinos and photons. Both species contribute to the gravitational potential (which is our interest in this chapter) and both start out with the same initial conditions. It is not quite as obvious that both follow the same evolution equations (the $\dot{\tau}$ terms can be neglected in equation [3.96]) or that these evolution equations are the ones given in [6.10] and [6.11]. But it is true, at least in the limit of small baryon density, and again only for the purposes of following the matter evolution. You can work out the details in Problem (1), and we will explore the full photon evolution equation in the next chapter.

To close the set of equations for the dark matter density, we need an equation for the gravtitational potential Φ. You may have noticed that in equations [6.10]-[6.13], I set $\Psi \to -\Phi$, anticipating the fact that one of Einstein equations ([4.32]) tells us that the sum of the two potentials vanishes if there are no quadrupole moments. Since some of the Einstein equations are redundant, we have several choices now that we need one last equation relating Φ to the radiation and matter overdensities. We can use either the time-time component ,equation [4.26],

$$k^2\Phi + 3\frac{\dot{a}}{a}\left(\dot{\Phi} + \frac{\dot{a}}{a}\Phi\right) = 4\pi G a^2\left[\rho_{\rm DM}\delta_{\rm DM} + 4\rho_R\Theta_{R,0}\right]. \tag{6.14}$$

Here, again I have set Ψ to $-\Phi$, neglected the baryons*, and merged the neutrino and photon contributions to the potential. The alternative is to use the algebraic (no time derivatives) equation ([4.66])

$$k^2\Phi = 4\pi G a^2\left[\rho_{\rm DM}\delta_{\rm DM} + 4\rho_R\Theta_{R,0} + \frac{3aH}{k}\left(i\rho_{\rm DM}v_{\rm DM} + 4\rho_R\Theta_{R,1}\right)\right]. \tag{6.15}$$

Both of these equations will be useful to us at various times, although only one is necessary to close the set of equations for the five variables $\rho_{\rm DM}, v_{\rm DM}, \Theta_{R,0}, \Theta_{R,1}$, and Φ.

At this stage, the simplest thing to do is solve the set of five coupled equations numerically. If equation [6.14] is used, there are no numerical difficulties†, and with very little work, you

*This is a fairly good approximation since in most models, the baryon density is much smaller than the dark matter density. We will explore the effects of baryons in section 5.6

†This is not completely true. The main point is that equation [6.15] is very difficult to use because the initial conditions have to be set extremely accurately. The one numerical problem I found using equation [6.14] occurs on small scales when I tried to evolve all the way to the present. The photon moments then become difficult to track, and any reasonable differential equation solver will balk at late times. However, there are several simple solutions to this: (i) by the late times in question, the potential is constant so there is no need to evolve all the way to the present or (ii) stop following the photon moments after a certain time; they don't have any effect on the matter distribution at late times anyway.

can have a code which computes the transfer function (in the absence of baryons) in about a second.

Analytic solutions for the dark matter density are harder to come by. I know of no analytic solution valid on all scales at all times. To make progress, we will have to take some limits which reduce the full set of five equations to a more managable two or three. The cost is that these limits will be valid only for certain scales at certain times. Patching these analytic solutions together to obtain a reasonable transfer function is as much art as science.

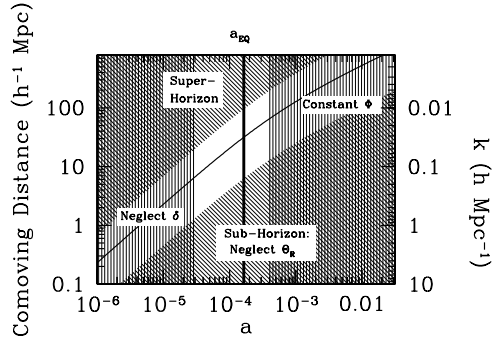


Figure 6.5: Physics of the transfer function. Hatched regions show where analytic expressions exist. The gaps in the center show that no analytic solutions exist to capture the full evolution of intermediate scale modes. The curve monotonically increasing from bottom left to top right is the comoving horizon.

As a guide to this analytic work which will occupy us much of the rest of this chapter, consider figure 6.5. The solid curve is the comoving horizon (conformal time), which increases with time, equal to about 30 h^{-1} Mpc at the epoch of equality‡. A given comoving scale remains constant with time. Take for example, a comoving distance of 10 h^{-1} Mpc, corresponding to wavenumber $k = 0.1$ h Mpc^{-1}. At early times ($a < 10^{-5}$) this distance is larger than the horizon, so $k\eta << 1$. We can then drop all terms proportional to k in the evolution equations. In section 5.2.1, we will derive an exact solution for the potential in this super-horizon limit. Unfortunately, figure 6.5 indicates that, for the mode in question, this super-horizon solution is valid only until $a \simeq 10^{-5}$. At much later times ($a > 10^{-3}$) the mode is well within the horizon and the radiation perturbations have become irrelevant (since the universe is matter dominated). We will see in section 5.3.2 that, under these conditions, another analytic solution can be found. The difficulty is matching the super-horizon solution to the sub-horizon solution.

The problem of matching the super-horizon solution to the sub-horizon solution can be solved for very large scale ($k < 0.01$ h Mpc^{-1}) and very small scale ($k > 0.5$ h Mpc^{-1}) modes.

‡This is model dependent; the plot shows sCDM, with $h = 0.5$.

In the large scale case, we will see in section 5.2.2 that once the universe becomes matter dominated, $\Phi =$ constant is a solution to the evolution equations even as the mode crosses the horizon. This fact serves as a bridge between the super- and sub-horizon solutions, both of which have constant Φ in the matter dominated regime. In the small scale case, we can neglect matter perturbations as the mode crosses the horizon, since these modes cross the horizon when the universe is deep in the radiation era. Then, once the mode is sufficiently within the horizon, radiation perturbations decay away, and we can match on to the sub-horizon, no radiation perturbation solution of section 5.3.2.

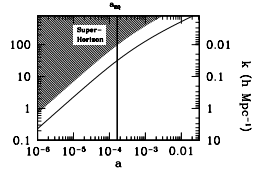
With analytic expressions on both large and small scales, we can obtain a good fit to the transfer function by splining the two solutions together. We will see in section 5.4 that this works, primarily because the transfer function is so smooth, monotonically decreasing from unity on large scales.

6.2 Large Scales

On very large scales, we can get analytic solutions for the potential first through the matter-radiation transition and then through horizon crossing. We start with the super-horizon solution valid through the matter-radiation transition. The results of §6.2.1 will be that the potential drops by a factor of 9/10 as the universe goes from radiation to matter domination.

6.2.1 Super-Horizon Solution

For modes that are far outside the horizon, $k\eta << 1$ and we can drop all terms in the evolution equations dependent on k. From equations [6.10] and [6.12], we see that, in this limit, the velocity decouples from the evolution equations. This immediately reduce the number of equations to solve from five to three. For the third equation, we notice that equation [6.15] has terms inversely proportional to k. These will be difficult to deal with, so let us choose equation [6.14] instead. We are left with



$$\dot{\Theta}_{R,0} = -\dot{\Phi} \tag{6.16}$$

$$\dot{\delta}_{\rm DM} = -3\dot{\Phi} \tag{6.17}$$

$$3\frac{\dot{a}}{a}\left(\dot{\Phi} + \frac{\dot{a}}{a}\Phi\right) = 4\pi G a^2 \left[\rho_{\rm DM}\delta_{\rm DM} + 4\rho_R\Theta_{R,0}\right]. \tag{6.18}$$

We can go a step further by realizing that the first two equations require $\delta_{\rm DM} - 3\Theta_{R,0}$ to be constant. Further, we know that this constant is zero (these are the initial conditions). So let us use the dark matter equation ([6.17]) and the Einstein equation with $\Theta_{R,0}$ set to $\delta_{\rm DM}/3$. The Einstein equation is then

$$3\frac{\dot{a}}{a}\left(\dot{\Phi} + \frac{\dot{a}}{a}\Phi\right) = 4\pi G a^2 \rho_{\rm DM}\delta_{\rm DM}\left[1 + \frac{4}{3y}\right]. \tag{6.19}$$

Here I have introduced

$$y \equiv \frac{a}{a_{\rm EQ}} = \frac{\rho_{\rm DM}}{\rho_R} \tag{6.20}$$

which we will use as an evolution variable instead of η or a.

Equations [6.17] and [6.19] are two first order equations for the two variables $\delta_{\rm DM}$ and Φ. The stategy will be to turn these two first order equations into one second order equation and then solve. First, though, let us rewrite the equations in terms of the new variable y. The derivative with respect to y is related to that with respect to η via the Jacobian,

$$\begin{aligned} \frac{d}{d\eta} &= \frac{dy}{d\eta}\frac{d}{dy} \\ &= aHy\frac{d}{dy} \end{aligned} \tag{6.21}$$

where the second line follows from the definition of y and the fact that $\dot{a} = a^2 H$. In terms of y then, the Einstein equation becomes

$$\begin{aligned} y\Phi' + \Phi &= \frac{y}{2(y+1)}\delta_{\rm DM}\left[1 + \frac{4}{3y}\right] \\ &= \frac{3y+4}{6(y+1)}\delta_{\rm DM} \end{aligned} \tag{6.22}$$

where prime denotes derivatives with respect to y and the right side of the first line follows since $8\pi G\rho_{\rm DM}/3 = (8\pi G\rho/3)y/(y+1) = H^2 y/(y+1)$. In general, to turn two first order equations into one second order equation, the trick is to differentiate one of them. Here, to simplify the algebra, we first rewrite equation [6.22] as an expression for $\delta_{\rm DM}$; then differentiate with respect to y; and finally set $\delta'_{\rm DM}$ to $-3\Phi'$ thanks to the dark matter equation, [6.17]. This leads to

$$-3\Phi' = \frac{d}{dy}\left\{\frac{6(y+1)}{3y+4}\left[y\Phi' + \Phi\right]\right\}. \tag{6.23}$$

Carrying out the derivative is tedious but straightforward. We are left with

$$\Phi'' + \frac{21y^2 + 54y + 32}{2y(y+1)(3y+4)}\Phi' + \frac{\Phi}{y(y+1)(3y+4)} = 0. \tag{6.24}$$

Remarkably, Kodama and Sasaki (1986) found an analytic solution to equation [6.24]. They introduced a new variable

$$u \equiv \frac{y^3}{\sqrt{1+y}}\Phi. \tag{6.25}$$

In terms of this variable, you will show (Problem 5.4) that equation [6.24] becomes

$$u'' + u'\left[\frac{-2}{y} + \frac{3/2}{1+y} - \frac{3}{3y+4}\right] = 0. \tag{6.26}$$

That is, there is no term proportional to u. Instead of a second order equation for Φ, then, we have a first order equation for u'. Fortunately, this first order equation is integrable. Starting from

$$\frac{du'}{u'} = dy\left[\frac{2}{y} - \frac{3/2}{1+y} + \frac{3}{3y+4}\right], \tag{6.27}$$

we can integrate to get

$$\ln(u') = \text{constant} + 2\ln(y) - (3/2)\ln(1+y) + \ln(3y+4). \tag{6.28}$$

Then exponentiating gives

$$u' = A\frac{y^2(3y+4)}{(1+y)^{3/2}} \tag{6.29}$$

where A is a constant to be determined.

We are one integral away from an analytic expression for the gravitational potential. Remembering the definition of u, we can integrate equation [6.29] to obtain

$$\frac{y^3}{\sqrt{1+y}}\Phi = A\int_0^y \frac{y^2(3y+4)}{(1+y)^{3/2}}. \tag{6.30}$$

Note that there should be another constant, $u(0)$, here. However, since $y^3\Phi \to 0$ early on, this constant is vanishes. By similar logic, we can determine the constant A even before performing the integral. For small y, the integrand becomes $4y^2$, so for small y equation [6.30] becomes $\Phi = 4A/3$. Therefore, $A = 3\Phi(0)/4$. The integral can be done analytically (Problem (3) again) leaving

$$\Phi = \frac{\Phi(0)}{10}\frac{1}{y^3}\left[16\sqrt{1+y} + 9y^3 + 2y^2 - 8y - 16\right]. \tag{6.31}$$

Equation [6.31] is our final expression for the potential on super-horizon scales. Although it is not obvious, at small y this expression sets $\Phi = \Phi(0)$, a constant. This must be so, since we chose the two constants of integration with precisely this condition. At large y, once the universe has become matter-dominated, the y^3 term in the brackets dominates, so $\Phi \to (9/10)\Phi(0)$. This is precisely the result we were after: the potential on even the largest scales drops by ten percent as the universe passes through the epoch of equality.

Let us compare this analytic result, valid only when modes are super-horizon, with the numerical results. Figure 6.6 shows that the solution works perfectly on the largest scales and even tolerably well (better than ten percent) for scales as small as $k = .01\text{h Mpc}^{-1}$. This is slightly better than we had anticipated from a crude estimate of where the super-horizon solution is valid (figure 6.5) and will be important for us later on when spline together the large and small scales solutions. A feature of the analytic solution which may be surprising to you is that, although it is true that the (large scale) potentials are constant in both the matter and radiation epochs, the transition between the pure matter and pure radiation eras is quite long. For example, and this is an important example for the purposes of the CMB as we will see in the next chapter, the potentials, even for the largest scale modes, are still decaying as late as $a \simeq 10^{-3}$, significantly after $a_{\rm EQ}$. In models with less matter, $a_{\rm EQ}$ is pushed even closer to 10^{-3} so the decay of the potentials becomes even more apparent at the time of decoupling.

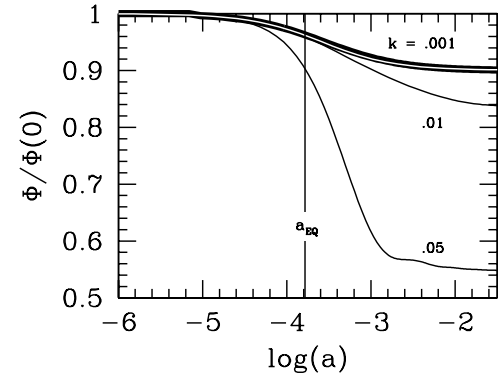
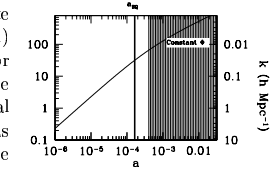


Figure 6.6: Super-horizon evolution of the potential in a CDM model with no baryons, $h = 0.5$ and $\Omega_m = 1$. Thick solid line shows the analytic result of equation [6.31], valid only on large scales. White curve within is for the mode $k = .001\text{h Mpc}^{-1}$. Two other smaller scale modes are shown.

6.2.2 Through horizon crossing



One interesting feature of figure 6.6 which you should take note of is that large scale potential (the numerical solution) remains constant even at very late times ($a > 10^{-2}$). For $k = .01\text{h Mpc}^{-1}$, the potential is constant long after the mode has entered the horizon. This result, that the potential remains constant as it crosses the horizon is valid as long as the universe is matter dominated. We now set out to prove it.

We are interested then in our set of five equations in the limit that radiation is not important. The potential depends only on the matter inhomogeneities, so we can neglect the two radiation equations, [6.10] and [6.11]. In addition to the two matter equations, we now keep the second of Einstein's equations, [6.15]. This is an algebraic equation, meaning that we could in principle eliminate Φ in the two matter equations and be left with a system of two first order differential equations. These two first order equations in general have two solutions. Instead of solving them directly, though, we can cheat using our knowledge of the initial conditions. Here is the idea: we just learned that, deep in the matter epoch, super-horizon potentials are constant. Therefore, the initial conditions for our problem are that the potential is constant ($\dot{\Phi} = 0$). If we can show that constant Φ is one of the two general solutions to the set of matter-dominated equations, then we don't care what the other solution is. For, the initial conditions ensure that the constant Φ solution will be *the* solution.

We want to see, then, if the set of equations

$$\dot{\delta}_{\rm DM} + ikv_{\rm DM} = 0 \tag{6.32}$$

$$\dot{v}_{\rm DM} + aHv_{\rm DM} = ik\Phi \tag{6.33}$$

$$k^2\Phi = \frac{3}{2}a^2H^2\left[\delta_{\rm DM} + \frac{3aHiv_{\rm DM}}{k}\right] \tag{6.34}$$

admit a solution with Φ a constant in time. We can use the algebraic equation, [6.34] to eliminate $\delta_{\rm DM}$ from the other two equations. Using the fact that $d(aH)/d\eta = -a^2H^2/2$, we see that equation [6.32] becomes

$$\frac{2k^2\dot{\Phi}}{3a^2H^2} + \frac{2k^2\Phi}{3aH} - \frac{3aHi\dot{v}_{\rm DM}}{k} + \frac{3a^2H^2iv_{\rm DM}}{2k} + ikv_{\rm DM} = 0. \tag{6.35}$$

We now have two first order equations for Φ and $v_{\rm DM}$. The strategy is to turn these two equations into one second order equation for Φ. First eliminate $\dot{v}_{\rm DM}$ from equation [6.35] by using the velocity equation. This leaves

$$\frac{2k^2\dot{\Phi}}{3a^2H^2} + \left[\frac{iv_{\rm DM}}{k} + \frac{2\Phi}{3aH}\right]\left(\frac{9a^2H^2}{2} + k^2\right) = 0. \tag{6.36}$$

If the second order equation is of the form $\alpha\ddot{\Phi} + \beta\dot{\Phi} = 0$, that is, if it has no terms proportional to Φ, then $\Phi =$ constant is a solution to the equations. So we differentiate equation [6.36] with respect to η but consider only the terms proportional to Φ, dropping all terms proportional to derivatives of Φ. Using the fact that $(d/d\eta)(aH)^{-1} = 1/2$, we see that the remaining terms are

$$\begin{aligned}\left[\frac{i\dot{v}_{\rm DM}}{k} + \frac{\Phi}{3}\right]\left(\frac{9a^2H^2}{2} + k^2\right) &- \left[\frac{iaHv_{\rm DM}}{k} + \frac{2\Phi}{3}\right]\frac{d}{d\eta}\frac{9a^2H^2}{2} \\ &= -\left[\frac{iaHv_{\rm DM}}{k} + \frac{2\Phi}{3}\right]\left(9a^2H^2 + k^2\right)\end{aligned} \tag{6.37}$$

where I have eliminated $\dot{v}_{\rm DM}$ by using the velocity equation again. But equation [6.36] tells us that the term in square brackets is proportional to $\dot{\Phi}$. So there are no terms in the second order equation proportional to Φ. Constant potentials are therefore a solution in the matter dominated era. Since the initial conditions pick out this mode, constant potential is *the* solution in the matter dominated era.

Potentials remain constant as long as the universe is matter dominated. At much later times ($a > 1/10$), it is conceivable that the universe becomes dominated by some other form of energy – vacuum energy for example – or, less likely by curvature. If so, then the potentials will decay. This decay is described by the growth function though (§6.5) and does not affect the transfer function. The main result of this section is that the transfer function as defined in equation [6.3] is very close to unity on all scales that enter the horizon after the universe becomes matter dominated. You will show in Problem (4) that the relevant scale is

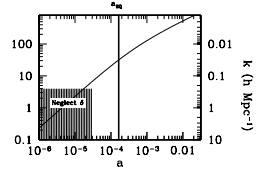
$$k_{\rm EQ} = 0.073{\rm Mpc}^{-1}\Omega_m h^2. \tag{6.38}$$

In the limit in which we are working, where baryons and anisotropic stresses are neglected, the transfer function depends only on $k/k_{\rm EQ}$. To get a feel for when the large scale approximations of this section are valid, look back at figure 6.6, plotted for the standard CDM model with $\Omega_m = 1$ and $h = 0.5$. The transfer function for the curve labelled .01 is seven percent lower (84/.9) than unity. For that mode, $k/k_{\rm EQ} = .01/(.073h) = 0.27$. So if we are interested in ten percent accuracy in the transfer function, then we can use the large scale approximation for $k < k_{\rm EQ}/3$.

6.3 Small Scales

We were able to solve for the evolution of large scale perturbations in the previous section because the modes crossed the horizon well *after* the epoch of equality. Therefore, the problem neatly divided into (i) super-horizon modes passing through the epoch of equality and then (ii) modes in the matter dominated era which cross the horizon. The converse is true for the small scale modes considered in this section. They cross the horizon when the universe is deep in the radiation era. So the problem divides neatly into (i) modes in the radiation era crossing the horizon and then (ii) sub-horizon modes passing through the epoch of equality. Step (i) we treat in §6.3.1, step (ii) in §6.3.2. Notice that we are unable to treat analytically modes which enter the horizon around the epoch of equality.

6.3.1 Horizon crossing



When the universe is radiation dominated, the potential is determined by perturbations to the radiation. The dark matter perturbations – the ones we are interested in in this chapter – are influenced by the potential, but do not themselves influence the potential. So the situation is as depicted in figure 6.7. To solve for matter perturbations in this epoch, therefore, is a two-step problem. First, we must solve the coupled equations for $\Theta_{R,0}$, $\Theta_{R,1}$, and Φ. Then we solve the equation for matter evolution using the potential as an external driving force.

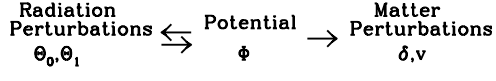


Figure 6.7: Coupling of perturbations in the radiation era. Radiation perturbations and the gravitational potential affect each other. Matter perturbations do not affect the potential but are driven by it.

To solve for the potential in the radiation dominated era, we choose equation [6.15]. Dropping the matter source terms, we have

$$\Phi = \frac{6a^2H^2}{k^2}\left[\Theta_{R,0} + \frac{3aH}{k}\Theta_{R,1}\right] \tag{6.39}$$

since $H^2 = 8\pi G\rho_R$ in the radiation era. Also in the radiation era, $aH = 1/\eta$. Armed with this fact, we can use Einstein's equation [6.39] to eliminate $\Theta_{R,0}$ from the two radiation equations, [6.10] and [6.11]. These become

$$-\frac{3}{k\eta}\dot{\Theta}_{R,1} + k\Theta_{R,1}\left[1 + \frac{3}{k^2\eta^2}\right] = -\dot{\Phi}\left[1 + \frac{k^2\eta^2}{6}\right] - \Phi\frac{k^2\eta}{3} \tag{6.40}$$

$$\dot{\Theta}_{R,1} + \frac{1}{\eta}\Theta_{R,1} = \frac{-k}{3}\Phi\left[1 - \frac{k^2\eta^2}{6}\right] \tag{6.41}$$

We can turn these two first order equations for Φ and $\Theta_{R,1}$ into one second order equation for the potential. Use equation [6.41] to eliminate $\dot{\Theta}_{R,1}$ from the first equation, which then becomes

$$\dot{\Phi} + \frac{1}{\eta}\Phi = \frac{-6}{k\eta^2}\Theta_{R,1}. \tag{6.42}$$

We now have an expression for $\Theta_{R,1}$ solely in terms of the potential and its first derivative. To arrive at a second order equation for Φ, we differentiate. When we do, we wll encounter terms proportional to $\Theta_{R,1}$ and its derivative. Each of these can be eliminated with equation [6.41] and equation [6.42]. The resulting second order equation is

$$\ddot{\Phi} + \frac{4}{\eta}\dot{\Phi} + \frac{k^2}{3}\Phi = 0. \tag{6.43}$$

To determine the behavior of the potential in the radiation dominated era, we must solve equation [6.43] subject to the initial conditions that Φ is constant. It can be solved analytically by defining $R \equiv \Phi\eta$. Then equation [6.43] becomes

$$\ddot{R} + \frac{2}{\eta}\dot{R} + \left(\frac{k^2}{3} - \frac{2}{\eta^2}\right)R = 0. \tag{6.44}$$

This is the spherical Bessel equation of order one with solutions $j_1(k\eta/\sqrt{3})$ – the spherical Bessel function – and $n_1(k\eta/\sqrt{3})$ – the spherical Neumann function. The latter blows up as η gets very small, so we discard it on the basis of the initial conditions. The spherical Bessel function of order one is easily expressed in terms of trigonometric functions, so

$$\Phi = 3\Phi_{\rm p}\left(\frac{\sin(k\eta/\sqrt{3}) - (k\eta/\sqrt{3})\cos(k\eta/\sqrt{3})}{(k\eta/\sqrt{3})^3}\right). \tag{6.45}$$

The factor of three in front here arises because the $\eta \to 0$ limit of the expression in parentheses is 1/3.

Equation [6.45] tells us that, as soon as a mode enters the horizon during the radiation dominated era, its potential starts to decay. After decaying, the potential oscillates, as depicted in figure 6.8. Qualitatively, we could have anticipated as much. By definition, radiation travels at the speed of light. Perturbations on scales smaller than the horizon, therefore, can dissipate, leading to the decay of the gravitational potentials. Afterwards, sound waves are set up in the radiation leading to oscillations in the radiation density,

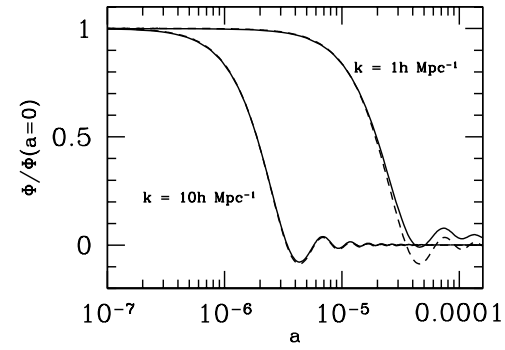


Figure 6.8: Evolution of the potential in the radiation dominated era. For two small scale modes which enter the horizon well before equality, the exact (solid curve) solution is shown along with the approximate analytic solution (dashed curve) of equation [6.45].

and therefore also in the potential coupled to it. Note from figure 6.8 though that this approximate description – in which the effect of matter on the potential is neglected – is valid only deep in the radiation era. The analytic solution for the $k = 1$ h Mpc^{-1} mode already begins to depart from the exact solution at $a \simeq 3 \times 10^{-5}$, well before equality (here, in the sCDM model I have taken for illustrative purposes, at $a \simeq 2 \times 10^{-4}$).

Armed with knowledge of the potential in the radiation dominated era, we can now determine the evolution of the matter perturbations, the second half of figure 6.7. To do this, we turn the two matter evolution equations – [6.12] and [6.13] – into one second order equation with the potentials serving as an external source. Differentiate equation [6.12] and use equation [6.13] to eliminate $\dot{v}_{\rm DM}$

$$\ddot{\delta}_{\rm DM} + ik\left(-\frac{\dot{a}}{a}v_{\rm DM} + ik\Phi\right) = -3\ddot{\Phi}. \tag{6.46}$$

Now we can use equation [6.12] to eliminate $v_{\rm DM}$ leading to

$$\ddot{\delta}_{\rm DM} + \frac{1}{\eta}\dot{\delta}_{\rm DM} = S(k,\eta) \tag{6.47}$$

where the source term is

$$S(k,\eta) = -3\ddot{\Phi} + k^2\Phi - \frac{3}{\eta}\dot{\Phi}. \tag{6.48}$$

The two solutions to the homogeneous equation ($S = 0$) associated with equation [6.47] are $\delta_{\rm DM} =$ constant and $\delta_{\rm DM} = \ln(a)$ (or, equivalently in the radiation dominated era $\ln[\eta]$). In general, the solution to a second order equation is a linear combination of the two homogeneous solutions and a particular solution. In the absence of a revelation about the

particular solution, one can construct it from scratch from the two homogeneous solutions (call them s_1 and s_2) and the source terms. It is the integral of the source term weighted by the Green's function $[s_1(\eta)s_2(\eta') - s_1(\eta')s_2(\eta)]/[\dot{s}_1(\eta')s_2(\eta') - s_1(\eta')\dot{s}_2(\eta')]$. So here, we have

$$\delta_{\rm DM}(k,\eta) = C_1 + C_2\ln(\eta) - \int_0^\eta d\eta' S(k,\eta')\eta'\left(\ln(k\eta') - \ln(k\eta)\right). \tag{6.49}$$

At very early times the integral is small, so our initial conditions ($\delta_{\rm DM}$ constant) dictate that the coefficient of $\ln(\eta)$, C_2, vanishes and $C_1 = \delta_{\rm DM}(k,\eta=0) = 3\Phi_{\rm p}/2$. Now let us consider the integral in equation [6.49]. The source function decays to zero along with the potential as the mode enters the horizon. Thus, the dominant contribution to the integral comes from the epochs during which $k\eta$ is of order one. The integral over $S(\eta')\ln(k\eta')$ therefore will just asymptote to some constant, while the integral over $S(\eta')\ln(k\eta)$ will lead to a term proportional to $\ln(k\eta)$ with the constant of proportionality being just that, a constant. Thus, we expect that after the mode has entered into the horizon,

$$\delta_{\rm DM}(k,\eta) = A\Phi_{\rm p}\ln(Bk\eta), \tag{6.50}$$

i.e. a constant ($A\Phi_{\rm p}\ln[B]$) plus a logarithmic growing mode ($A\Phi_{\rm p}\ln[k\eta]$).

We can determine the constants A and B in equation [6.50] by referring to the relevant parts of equation [6.49]. The constant term, $A\Phi_{\rm p}\ln(B)$, is equal to C_1 plus the integral over $\ln(\eta')$, or

$$A\Phi_{\rm p}\ln(B) = \frac{3}{2}\Phi_{\rm p} + \int_0^\infty d\eta' S(k,\eta')\eta'\ln(k\eta'), \tag{6.51}$$

while the coefficient of the $\ln(k\eta)$ term is set by the remaining integral

$$A\Phi_{\rm p} = \int_0^\infty d\eta' S(k,\eta')\eta'. \tag{6.52}$$

Note that in both integrals here, I have set the upper limit to infinity in accord with our expectation that the integrals asymptote to some constant value at large η. Using the expression for the source term, equation [6.48], and our analytic approximations to the potential, equation [6.45], we can evaluate the integrals here and determine A and B. I find $A = 9.0$ and $B = 0.62$. Hu and Sugiyama (1996), who introduced this method for following the dark matter evolution at early times, found that integrating the exact potentials (instead of the approximate ones of equation [6.45]) leads to slightly different values, $A = 9.6$ and $B = 0.44$.

Figure 6.9 shows the exact solution for $\delta_{\rm DM}$ in the radiation era along with the approximation of equation [6.50]. Setting aside the details for a moment, we see that matter perturbations do indeed grow even during the radiation era. The growth is not as prominent as during the matter era (when the constant potentials derived in §6.2 imply $\delta_{\rm DM} \propto a$) due to the pressure of the radiation but it still exists. For both modes shown in figure 6.9 the perturbations do indeed settle into the logarithmic growing mode once they enter the horizon. As the universe gets closer to matter domination though, the pressure of the radiation eases up, and the perturbations begin to grow faster. Indeed, you might be worried that our approximation for the $k = 1{\rm h\ Mpc}^{-1}$ mode is not very useful. Fortunately, we will be using these solutions only to set the initial conditions for growth in the sub-horizon epoch (next

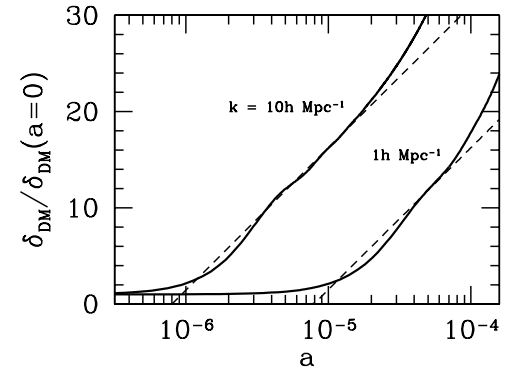
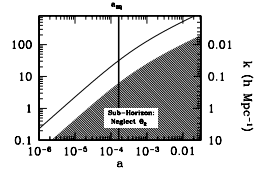


Figure 6.9: Matter perturbations in the radiation dominated era. The two modes shown here both enter the horizon in the radiation era and lock onto the logarithmically growing mode after some oscillations. Heavy solid curves are the exact solutions, light dashed curves the logarithmic mode of equation [6.50]. The perturbations have been normalized by their values at early times: inflation predicts a larger initial amplitude (by a factor of $10^{3/2}$) for the larger scale mode.

sub-section), so the approximation need be valid only for a very limited range of times. As long as we choose the matching epoch appropriately, the logarithmic approximation will be extremely good.

6.3.2 Sub-horizon evolution



We saw in the last sub-section that radiation pressure causes the gravitational potentials to decay as modes enter the horizon during the radiation era. Although I did not focus on the radiation perturbations themselves (we will do this in the next chapter), you might expect that the pressure suppresses any growth in $\Theta_{R,0}$. This is correct, and it is in sharp contrast to the matter perturbations which, we just saw, grow logarithmically. Although initially the potential is determined by the radiation (since the universe is radiation dominated), eventually the growth in the matter perturbations more than offsets the fact that there is more radiation than matter. That is, eventually $\rho_{\rm DM}\delta_{\rm DM}$ becomes larger than $\rho_R\Theta_{R,0}$ even if $\rho_{\rm DM}$ is smaller than ρ_R. Once this happens, the gravitational potential and the dark matter perturbations evolve together and do not care what happens to the radiation. In this sub-section, we want to solve the set of equations governing the matter perturbations and the potential and then match on to the logarithmic solution ([6.50]) set up during the epoch in which the potential

decays.

Once again our starting point is the set of equations governing dark matter evolution, [6.12] and [6.13], and the algebraic equation for the gravitational potential, [6.15]. And, once again, we want to reduce this set of three equations (two of which are first order differential equations) to one second order equation. We will want to follow the sub-horizon dark matter perturbations through the epoch of equality, so it proves convenient to again use y (equation [6.20]) – the ratio of the scale factor to its value at equality – as the evolution variable. In terms of y, the three equations become

$$\delta'_{\rm DM} + \frac{ikv_{\rm DM}}{aHy} = -3\Phi' \tag{6.53}$$

$$v'_{\rm DM} + \frac{v_{\rm DM}}{y} = \frac{ik\Phi}{aHy} \tag{6.54}$$

$$k^2\Phi = \frac{3y}{2(y+1)}a^2H^2\delta_{\rm DM}. \tag{6.55}$$

Several comments are in order about this version of our fundamental equations. First, notice that the time derivatives in the first two equations has been replaced with derivatives with respect to y (indicating by primes), and this transformation leads to the factors of $\dot{y} = aHy$ in the denominators of the unprimed terms. Second, the gravitational potential is now expressed solely in terms of $\delta_{\rm DM}$: there is no dependence on radiation perturbations because of our arguments above that these are sub-dominant, and there is no $aHv_{\rm DM}/k$ dependence because the perturbations are well within the horizon and $aH/k << 1$. Finally, the coefficient of the $\delta_{\rm DM}$ source term is $4\pi G\rho_{\rm DM}a^2 \to (3/2)a^2H^2y/(y+1)$ since we are interested in times early enough that any curvature or vacuum energy are negligible. You might be tempted to throw out all the gravitational source terms in the equations for the dark matter perturbations since Φ is proportional to $(aH/k)^2$. This would be too hasty though as we will see presently.

We now go through the familiar routine of turning equations [6.53] and [6.54] into a second order equation for $\delta_{\rm DM}$: differentiate the first of these to get

$$\delta''_{\rm DM} - \frac{ik(2+3y)v_{\rm DM}}{2aHy^2(1+y)} = -3\Phi'' + \frac{k^2\Phi}{a^2H^2y^2} \tag{6.56}$$

where $v'_{\rm DM}$ has been eliminated using the velocity equation. Also I have used the fact that $d/dy(1/aHy) = -(1+y)^{-1}(2aHy)^{-1}$. Now my rationale for holding on to the Φ terms becomes clearer: in the process of getting the second order equation for $\delta_{\rm DM}$, we obtained one term proportional to Φ which is enhanced by the very large $k^2/(aH)^2$ factor needed to make it relevant. Using equation [6.55], we immmediately recognize this combination as $3y\delta_{\rm DM}/(2y+2)$. The other terms proportional to Φ have no such enhancement so can be neglected. Thus, the combination $ikv_{\rm DM}/(aHy)$ can be simply replaced by $-\delta'_{\rm DM}$ (using equation [6.53]) leaving

$$\delta''_{\rm DM} + \frac{2+3y}{2y(y+1)}\delta'_{\rm DM} - \frac{3}{2y(y+1)}\delta_{\rm DM} = 0.. \tag{6.57}$$

This is the *Meszaros equation* governing the evolution of sub-horizon cold, dark matter perturbations once radiation perturbations have become negligible.

To understand the growth of dark matter perturbations, we need to obtain the two independent solutions to the Meszaros equations and then match on to the logarithmic mode established in the previous sub-section. To solve this differential equation, we can use our knowledge of the solution deep in the matter era. We know that sub-horizon perturbations in the matter era grow with the scale factor, so one of the solutions to equation [6.57] is a polynomial in y of order one. Therefore, for one mode at least, $\delta''_{\rm DM}$ vanishes. Therefore, the equation governing this first mode, the growing mode, is $D'_1/D_1 = 1/(y+2/3)$, the solution to which is

$$D_1(y) = y + 2/3. \tag{6.58}$$

To find the second solution, notice that the Meszaros equation tells us that $u \equiv \delta_{\rm DM}/D_1$ satisfies

$$(1+3y/2)u'' + \frac{u'}{y(y+1)}[(21/4)y^2 + 3y + 1] = 0. \tag{6.59}$$

Since there is no term proportional to u, equation [6.59] is actually a first order equation for u'[8]. We can therefore integrate to obtain a solution for u' and then integrate again to get the second Meszaros solution. The first integral gives

$$u' \propto (y+2/3)^{-2}y^{-1}(y+1)^{-1/2}. \tag{6.60}$$

Integrating again leads to the second Meszaros solution

$$D_2(y) = D_1(y)\ln\left[\frac{\sqrt{1+y}+1}{\sqrt{1+y}-1}\right] - 2\sqrt{1+y}. \tag{6.61}$$

At lates times ($y >> 1$), the growing solution D_1 scales as y while the decaying mode D_2 falls off as $y^{-3/2}$.

The general solution to the Meszaros equation is therefore

$$\delta_{\rm DM}(k,y) = C_1D_1(y) + C_2D_2(y) \qquad y >> y_H \tag{6.62}$$

where y_H is the scale factor when the mode enters the horizon divided by the scale factor at equality (Problem (5)). To determine the constants C_1 and C_2 we can match on to the logarithmic solution of equation [6.50]. That solution is valid within the horizon but before equality: $y_H << y << 1$. So we can hope to arrive at a reasonable approximation for the evolution of dark matter perturbations only for those modes that enter the horizon before equality. For those modes, we match the two solutions and their first derivatives

$$\begin{aligned} A\Phi_{\rm p}\ln(By_m/y_H) &= C_1D_1(y_m) + C_2D_2(y_m) \\ \frac{A\Phi_{\rm p}}{y_m} &= C_1D'_1(y_m) + C_2D'_2(y_m) \end{aligned} \tag{6.63}$$

where the matching epoch y_m must satisfy $y_H << y_m << 1$. Note that I have replaced the argument of the log in equation [6.50] – $k\eta$ – with y/y_H, valid as long as the matching epoch

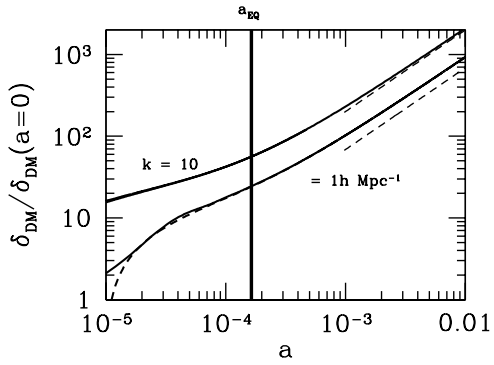


Figure 6.10: Evolution of small scale, sub-horizon, dark matter perturbations. Solid curves are exact solutions; dashed curves (almost imperceptible due to the goodness of fit in the 10h Mpc^{-1} case) the Meszaros solution with coefficients given by the matching condition, equation [6.63]. The dashed straight lines at $a > 10^{-3}$ are the asymptotic solution of equation [6.66].

is deep in the radiation era. Figure 6.10 shows the evolution of two modes along with the analytic solutions to the Meszaros equation with coefficients set by the matching conditions laid out in equation [6.63]. Not suprisingly, for larger scale modes than the ones shown, the approximation breaks down.

6.4 Numerical Results and Fits

In §6.2 and §6.3, we derived analytic solutions following the dark matter perturbations deep into the matter era. Here, we assimilate these results and spline them together to form the transfer function. Also, I will present a well-known fitting function for the transfer function.

First, we need to transform our expression ([6.62] along with equations [6.63]) for the small scale matter density into an expression for the transfer function. The transfer function is determined by the behavior of $\delta_{\rm DM}$ well after equality when the decaying mode has long since vanished. We can extract an even simpler form for $\delta_{\rm DM}$ in this $a >> a_{\rm EQ}$ limit. The key constant in that case is C_1, the coefficient of the growing mode. Multiplying the first matching condition by D_2' and the second by D_2 and then subtracting leads to

$$C_1 = \frac{D_2'(y_m) A \ln(B y_m/y_H) - D_2(y_m)(A/y_m)}{D_1(y_m) D_2'(y_m) - D_1'(y_m) D_2(y_m)} \tag{6.64}$$

The denominator $D_1 D_2' - D_1' D_2 = -(4/9) y_m^{-1} (y_m + 1)^{-1/2}$, which is approximately equal to

[§]Indeed this is a general trick for obtaining the second solution to a differential equation once the first is known. We will use it again later on to obtain the growth factor.

$-4/9 y_m$ since $y_m << 1$. Similarly for small y_m, $D_2 \to (2/3)\ln(4/y) - 2$ and $D_2' \to -2/3y$. Therefore,

$$C_1 \to \frac{-9A\Phi_{\rm p}}{4}\left[\frac{-2}{3}\ln(B y_m/y_H) - (2/3)\ln(4/y_m) + 2\right] \tag{6.65}$$

which fortuitously does not depend on y_m. Therefore, at late times we have an exact solution for the small scale dark matter perturbations

$$\delta_{\rm DM}(k,a) = \frac{3A\Phi_{\rm p}}{2}\ln\left[\frac{4Be^{-3}a_{\rm EQ}}{a_H}\right]\frac{a}{a_{\rm EQ}} \qquad a >> a_{\rm EQ}. \tag{6.66}$$

On very small scales, the argument of the log simplifies because $a_{\rm EQ}/a_H = \sqrt{2}k/k_{\rm EQ}$ (Problem (5)). To turn equation [6.66] into a transfer function, we need to remember both the definition of the transfer function, equation [6.5], and the way that $\delta_{\rm DM}$ is related to Φ at late times, equation [6.7]. Then, the transfer function on small scales is

$$T(k) = \frac{5A\Omega_m H_0^2}{2k^2 a_{\rm EQ}}\ln\left[\frac{4Be^{-3}\sqrt{2}k}{k_{\rm EQ}}\right]. \tag{6.67}$$

Recall that $k_{\rm EQ}$ is defined as $a_{\rm EQ}H(a_{\rm EQ}) = \sqrt{2}H_0 a_{\rm EQ}^{-1/2}$, so the prefactor is also a function of $k/k_{\rm EQ}$ only. Then, plugging in numbers leads to

$$T(k) = \frac{12k_{\rm EQ}^2}{k^2}\ln\left[\frac{k}{8k_{\rm EQ}}\right]. \tag{6.68}$$

Figure 6.11 shows the power spectrum for a standard CDM model ($n = 1$; $h = 0.5$; but no baryons) matching the large scale transfer function ($T = 1$) with the small scale transfer function of equation [6.68]. Also shown is the exact solution (again in the no-baryon limit), or equivalently, the fitting form of Bardeen, Bond, Kaiser, and Szalay (BBKS),

$$T(x \equiv k/k_{\rm EQ}) = \frac{\ln[1. + .171x]}{(.171x)}\left[1. + .284x + (1.18x)^2 + (0.399x)^3 + (.490x)^4\right]^{-0.25}. \tag{6.69}$$

Note that the BBKS form agrees very well with the analytic solution on small scales (i.e. both aymptote to $\ln(k)/k^2$ with the same coefficients.

Several final comments are in order. First, our analytic work has enabled us to understand the origin of the asymptotic, small scale behavior of the power spectrum. Had there been no logarithmic growth in the radiation era, the modes which entered very early on would have experienced no growth from horizon entry until the epoch of equality. Their amplitude relative to large scale modes would then have been suppressed by a factor of order $(k_{\rm EQ}/k)^2$. The logarithmic growing mode in the radiation era somewhat ameliorates this suppression. Second, although our analytic expression and its BBKS counterpart are good approximations, it is important to be aware of some small effects which affect the transfer function in the real world. We have assumed no anisotropic stresses ($\Phi = -\Psi$). Dropping this assumption changes the factor of 9/10 by which the potential drops for large scale modes to 0.86, resulting in a corresponding rise in the small scale transfer function. Including a realistic amount of baryons leads to even more severe small scale changes. We will address

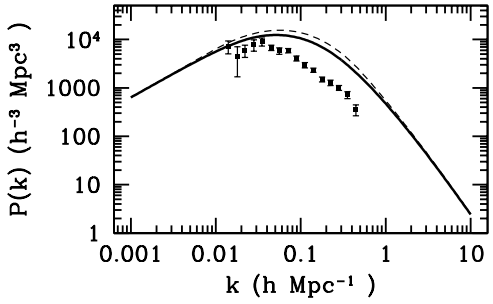


Figure 6.11: The power spectrum in a standard CDM model with a Harrison-Zel'dovich spectrum. The thick solid curve uses the BBKS transfer function; the dashed curves interpolates between the analytic transfer function on large scales (equal to one) and small scales (equation [6.67]). The data points are a compilation (and interpretation) by Peacock and Dodds (1994).

these in §6.6. Third, all of our work in this section has been on the transfer function, i.e. on the evolution of perturbations early on when the only components of the universe were matter and radiation. At very late times, the growth function depends on other hypothetical components, the most likely of which is dark energy. Finally, the data in Figure 6.11 tell us that (i) the large scale normalization is roughly correct and that (ii) the shape of the standard CDM power spectrum is wrong. The sCDM power spectrum turns over on very small scales, in distinct disagreement with the data. The universe as we observe it appears to have a smaller $k_{\rm EQ}$ than sCDM. This observation motivates consideration of variations on the standard model; we will consider these in §6.6. sc

6.5 Growth Function

At late times ($z < 10$) all modes of interest have entered the horizon. You might think, then, that the $y >> 1$ limit of the Meszaros equation, which describes sub-horizon modes in the matter era, would apply. This is true if $\Omega_m = 1$. If the energy budget of the universe has another item at late times – either dark energy or curvature – then we must retrace the steps which led to the Meszaros equation. Before doing this, I want to point out that, no matter what constitutes the energy budget today, all modes will experience the same growth factor. We saw this in the previous section, where the Meszaros equation is independent of k. And we now see it again, when we generalize the Meszaros equation to account for other forms of energy. This uniform growth is a direct result of the fact that cold, dark matter has zero pressure. Therefore, once a mode enters the horizon, there is no way for pressure to smooth out the inhomogeneities and all modes evolve identically.

We want to derive an evolution equation analagous to the Meszaros equation, but allowing for the possibility of energy other than matter or radiation. We can take the $y >> 1$ limit of

equations [6.53]-[6.55] , but we must rethink the coefficient of the source term in the Poisson equation. Since radiation can be ignored, the coefficient multiplying $\delta_{\rm DM}$ in equation [6.55] is now $4\pi G\rho_{\rm DM} = (3/2)H_0^2\Omega_m a^{-3}$. Also when differentiating equation [6.53] previously, we set $(1/aHy)' = -(1+y)^{-1}(2aHy)^{-1}$; here we need to account for other contributions to H' so equation [6.56] becomes

$$\delta''_{\rm DM} + ikv_{\rm DM}\left(\frac{d(aHy)^{-1}}{dy} - \frac{1}{aHy^2}\right) = \frac{3\Omega_m H_0^2}{2y^3a^2H^2a_{\rm EQ}}\delta_{\rm DM} \tag{6.70}$$

Replacing the velocity term using the continuity equation as before leads to

$$\frac{d^2\delta_{\rm DM}}{da^2} + \left(\frac{d\ln(H)}{da} + \frac{3}{a}\right)\frac{d\delta_{\rm DM}}{da} - \frac{3\Omega_m H_0^2}{2a^5H^2}\delta_{\rm DM} = 0. \tag{6.71}$$

Here I have divided by $a_{\rm EQ}^2$ and we will now use a as the variable instead of y. In this large y limit, all factors of $a_{\rm EQ}$ disappear.

There are two solutions to equation [6.71]. One solution is $\delta_{\rm DM} \propto H$. It is easy to check this if all the energy is non-relativistic matter, so that the solution is proportional to $a^{-3/2}$. Then all three terms scale as $a^{-7/2}$; the coefficient of the first is $15/4$, the second $-9/4$, and the last $3/2$. The sum of these does indeed vanish. In Problem (6), you will be asked to show that $\delta_{\rm DM} \propto H$ is a solution if there are other components of energy in the universe. This solution is pretty, but it is not the one we want. To see why, note from equation [1.3] that H is decreasing today no matter what constitutes the dominant energy density. The modes we are interested in – those that remain long after horizon-crossing – are the growing modes. So we are interested in the other solution of equation [6.71].

To obtain the growing mode, we try a solution of the form $u = \delta_{\rm DM}/H$. The evolution equation for u then becomes

$$\frac{d^2u}{da^2} + 3\left[\frac{d\ln(H)}{da} + \frac{1}{a}\right]\frac{du}{da} = 0. \tag{6.72}$$

This first order equation for u' can be integrated to obtain

$$\frac{du}{da} \propto (aH)^{-3}. \tag{6.73}$$

Integrating again and remembering that this second solution is uH leads to an expression for the growth factor

$$D_1(a) \propto H(a)\int^a \frac{da'}{(a'H(a'))^3}. \tag{6.74}$$

I have glossed over the proportionality constant. This is fixed by the definition of equation [6.4], which says that, early on when matter still dominates (say at $z \simeq 10$), D_1 should be equal to a. At those times, $H = H_0\Omega_m^{1/2}a^{-3/2}$ so the growth factor is

$$D_1(a) = \frac{5\Omega_m}{2}\frac{H(a)}{H_0}\int_0^a \frac{da'}{(a'H(a')/H_0)^3}. \tag{6.75}$$

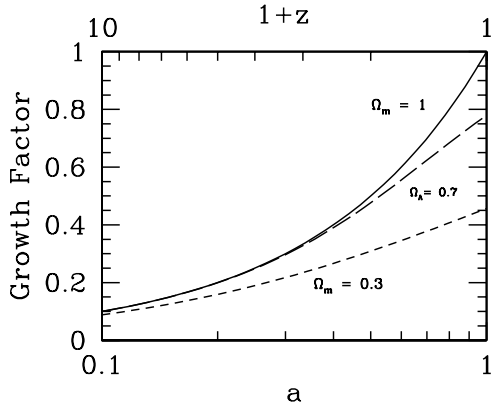


Figure 6.12: The growth factor in three cosmologies. Top two curves are for flat universes without and with a cosmological constant. Bottom curve is for an open universe.

The growth factor in an open universe with no cosmological constant can be derived analytically (see Problem (7)).

Figure 6.12 shows the growth factor for three different cosmologies. As mentioned above, if the universe is flat and matter dominated, the growth factor is simply equal to the scale factor. In both open and lambda cosmologies, though, growth is suppressed at late times. This leads to an important qualitative conclusion: structure in an open or lambda cosmology likely developed much earlier than in a flat, matter dominated universe. There has been relatively little evolution at recent times if the universe is open or lambda- dominated. Therefore, whatever structure is observed today was likely in place at much earlier times.

6.6 Beyond Cold Dark Matter

There is more to the universe than just cold dark matter. Although CDM is the main component in most cosmological modes, so that the transfer function we derived above is a good approximation to reality, there are trace amounts of other stuff. To be completely accurate we need to account for this other stuff. Here I focus on three additional components, only one of which undeniably exists. First, we consider the effect of the baryons, which constitute roughly ten percent of the total matter in most models, on the transfer function. Then, we entertain the possibility that neutrinos have mass and examine the resultant effect on the transfer function. Finally, dark energy – one model for which is the cosmological constant – is considered.

Figure 6.13 shows the transfer functions accounting for these components. A realistic

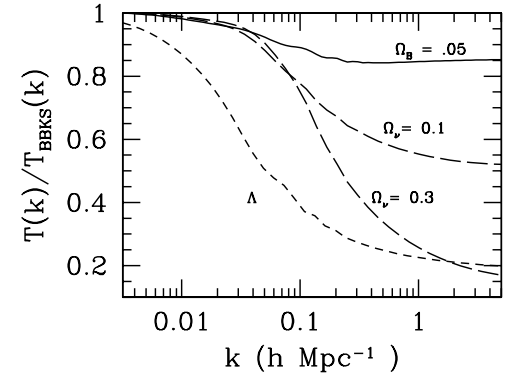


Figure 6.13: The ratio of the transfer function to the BBKS transfer function (equation [6.69]) which describes dark matter only (no baryons) perturbations. Top curve (and all other curves as well) has five percent baryons. Two middle curves show different values for a massive neutrino. Bottom curve has a cosmological constant $\Omega_\Lambda = 0.7$.

baryon fraction suppresses the transfer function on small scales. A massive neutrino does the same, with the nature and amplitude of the suppression depending on the neutrino mass. Dark energy, or a cosmological constant moves the epoch of equality to later times, thereby shifting $k_{\rm EQ}$. The break in the transfer function therefore comes on much larger scales than in the standard CDM model.

6.6.1 Baryons

Baryons account for between five and ten percent of the total energy density in the universe. As such, their effect on the matter power spectrum is small. A careful examination of figure 6.13 reveals two signatures of a non-zero baryon density. The first is that the power spectrum is suppressed on small scales. This is not surprising: at early times, before decoupling, baryons are tightly coupled to photons. Therefore, just as radiation perturbations decay when entering the horizon, so too do baryon over-densities. After decoupling, baryons are released from the relatively smooth radiation field and fall into the gravitational potentials set up by the dark matter. The depth of these wells is smaller than we estimated in §6.3, though, because only a fraction $\Omega_{\rm cdm}/\Omega_m$ of the total matter was involved in the collapse.

The second effect of baryons is less noticable in figure 6.13 and indeed may never get measured in real life either. Nonetheless, it is extremely important if only because it hints at a fundamental feature of the radiation field. In all the curves in figure 6.13, except the $\Omega_\nu = 0.3$ case, you can see small oscillations in the transfer function centered around $k \simeq 0.1$h Mpc^{-1}. These are not numerical artifacts. Rather, they are manifestations of the oscillations

that the combined baryon/photon fluid experience before decoupling. We got a glimpse of these in §6.3.2 (e.g. figure 6.8) when we considered the potential in the radiation dominated era. Just as the potential oscillates in this era, the baryon/photon fluid also oscillates. It is the traces of these oscillations that are imprinted on the matter transfer function. They are barely (if at all) detectable because baryons are such a small fraction of the total matter. However, you might think that these oscillations would be prominent in the spectrum of the radiation perturbations. We will see in the next chapter that this is precisely what happens.

6.6.2 Massive Neutrinos

Neutrinos are known to exist, and the standard Big Bang model gives a definite prediction for how many there are in the universe (equation [2.43]). One thing we do not yet know for sure about neutrinos, though, is whether or not they have mass. If they do, then the expression for the neutrino energy in equation [2.43], which assumes massless neutrinos, is incorrect. Rather, the energy density in one species of neutrinos with mass m_ν (Problem 5.9) is equal to

$$\Omega_\nu = \frac{m_\nu}{93.8h^2\mathrm{eV}}. \tag{6.76}$$

Thus a neutrino with a mass of 15 eV has $\Omega_\nu \sim 0.3$.

The reason why even a small admixture of massive neutrinos impact the power spectrum is that, especially if they are light, neutrinos can move fast (they are not *cold* dark matter) and stream out of high density regions. Perturbations on scales smaller than the free-streaming scale are therefore suppressed. Indeed, a long time ago, cosmologists considered the possibility that all the dark matter in the universe was in the form of neutrinos. If this were so, then there would be no power on small scales and structure would have to form from the "top down."

We can estimate the scale on which perturbations are damped by computing the comoving distance a massive neutrino can travel in one Hubble time at equality. This calculation is trivial however if the neutrino mass is in the eV range. For then, the average velocity, T_ν/m_ν, is of order unity at equality. So neutrinos can free-stream out of horizon-scale perturbations at equality. This leads to a suppression in power on all scales smaller than k_{EQ}.

Figure 6.13 shows this suppression. Note, though, that the effect is a little subtle. A lighter neutrino can free-stream out of larger scales, so the suppression begins at lower k for the $\Omega_\nu = 0.1$ mass than for the $\Omega_\nu = 0.3$ case. On the other hand, the more massive neutrino constitutes more of the total density so it suppresses small scale power more than does the lighter neutrino.

6.6.3 Dark Energy

Cosmologists have recently accumulated tantalizing evidence for dark energy in the universe above and beyond the dark matter that we have spent so much time on in this chapter. If dark energy exists, how does it affect the matter perturbations?

The first effect of dark energy is indirect. Since theoretical prejudice and evidence both indicate that the universe is flat, $\Omega_{\mathrm{dark}} \simeq 0.6 - 0.7$ implies that the matter density, Ω_m, is

less than one. This has a huge impact on the power spectrum, because we have seen that the power spectrum turns over at k_{EQ}, which is proportional to Ω_m. So dark energy leads to a turnover in the power spectrum on a scale much larger than predicted in standard CDM. In fact, as we saw in figure 6.11, this is one of the pieces of evidence for dark energy. The turnover in the power spectrum does not seem to appear on the scale predicted by standard CDM.

The second effect of the dark energy on the density inhomogeneities is more direct and more model dependent. At late times, amplification of perturbations is controlled by the growth factor of equation [6.75]. The evolution of the Hubble rate depends on the model of dark energy, so different models of dark energy predict different growth factors.

One way to parametrize the different models is to introduce the equation of state for the dark energy

$$w \equiv \frac{P}{\rho}. \tag{6.77}$$

If w is constant, then the dark energy density scales as $a^{-3[1+w]}$. (Check this scaling for matter ($w = 0$), radiation ($w = 1/3$), and the original dark energy, a cosmological constant ($w = -1$).) With this scaling, the Hubble rate in a flat universe evolves as

$$\frac{H(z)}{H_0} = \left[\frac{\Omega_m}{a^3} + \frac{\Omega_{\mathrm{dark}}}{a^{3[1+w]}}\right]^{1/2} \tag{6.78}$$

at late times.

To sum up, dark energy affects the power spectrum by changing k_{EQ} (this depends only on Ω_{dark}) and by changing the growth factor at late times (depends on both Ω_{dark} and w). Careful observations of the matter spectrum therefore may enable us to learn about dark energy.

Suggested Reading

Once again *The Large Scale Structure of the Universe* (Peebles) is a useful reference. Since it was written before the implications of cold dark matter and inflation were explored, though, it does not contain a transfer function or power spectrum like the ones we have derived (although Peebles himself was instrumental in computing these things several years after the book was published). A more up-to-date book, which is particularly strong on large scale structure is *Structure Formation in the Universe* (Padmanabhan).

The first papers to work out the CDM transfer function are particularly instructive to read, not least because they also focus on some of the physical implications of the hierarchical theories. See Blumenthal et al. (1984) and Peebles (1982). The most important recent paper is Seljak and Zaldarriaga (1996), not so much because it contains a concise description of the set of coupled equations to be solved (although it does that), but because it makes available CMBFAST, a code which computes transfer functions and CMB anisotropy spectra. It is available at `http://www.ias.edu/ matiasz/CMBFAST`. The treatment in this chapter follows most closely the small scale analytic solution of Hu and Sugiyama (1996), a paper which is extremely rich, and well-worth reading. A more recent paper by Eisenstein and Hu (1998) employs the analytic small scale solution to derive accurate fitting formulae.

Problems

6.1 Derive equations [6.10] and [6.11].

(a) First neglect the scattering term in equation [3.96], the one proportional to $\dot{\tau}$. Then the photon evolution equation is identical to the neutrino evolution equation, [3.103]. Show that this collisionless equation reduces to the two equations for the monopole and dipole. To get the monopole equation, multiply equation [3.103] by $d\mu/2P_0(\mu) = d\mu/2$ and integrate from $\mu = -1$ to 1. To get the dipole, multiply by $d\mu/2P_1(\mu)$ and integrate.

(b) Show that, in the limit of small baryon density, the scattering term in equation [3.96] can indeed be neglected. Neglect Π, since the quadrupole and polarization are very small. Then show that the scattering term is proportional to R, 3/4 times the baryon to photon ratio. You will want to use equation [3.102]. It cannot be emphasized enough that this series of approximations is valid only for the purposes of this chapter, wherein we are interested in the matter distribution.

6.2 Solve the set of five equations ([6.10]-[6.13] and [6.14]) numerically to obtain the transfer function for dark matter. Use the initial conditions derived in Chapter 5 and read the footnote on page ???. Plot the transfer function for sCDM (with Hubble constant $h = 0.5$) and LCDM (with $\Omega_\Lambda = 0.7$ and $h = 0.7$). Compare with the BBKS transfer function of equation [6.69].

6.3 Fill in some of the algebraic detail left out of §6.2.1.

(a) Show that equation [6.23] leads to equation [6.24] by carrying out the differentiation.

(b) Show that equation [6.24] is equivalent to equation [6.26] when the definition of u from equation [6.25] is used.

(c) Show that the integral in equation [6.30] can be done analytically with the result given in equation [6.31]. One way to do the integral is to define a dummy variable $x \equiv \sqrt{1+y}$.

6.4 Find the wavenumber of the mode which equals the inverse comoving Hubble radius at equality. That is, define $k_{\rm EQ}$ to be equal to $a_{\rm EQ}H(a_{\rm EQ})$. Show that this definition implies

$$k_{\rm EQ} = \sqrt{\frac{2\Omega_m H_0^2}{a_{\rm EQ}}}. \tag{6.79}$$

Then use equation [2.46] to show that $k_{\rm EQ}$ is given by equation [6.38]. Show that if you define $k_{\rm EQ}$ by setting it to $1/\eta_{\rm EQ}$, you get a number 17% lower.

6.5 Define a_H, the scale factor at which wavelength k equals the comoving Hubble radius, via $a_H H(a_H) \equiv k$. Express $a_H/a_{\rm EQ}$ in terms of k and $k_{\rm EQ}$. Show that in the limit $k >> k_{\rm EQ}$, this expression reduces to

$$\lim_{k>>k_{\rm EQ}} \frac{a_H}{a_{\rm EQ}} = \frac{k_{\rm EQ}}{\sqrt{2}k}. \tag{6.80}$$

6.6 Show that $\delta_{\rm DM} \propto H$ is a solution to the evolution equation [6.71] if the universe is flat with a cosmological constant. You will need to use equation [1.3]. Show also that the solution is valid if the universe has zero cosmological constant, but is open with $\Omega_m < 1$.

6.7 Derive the growth factor for an open universe with $\Omega_m < 1$:

$$D_1(a,\Omega_m) = \frac{5\Omega_m}{2(1-\Omega_m)}\left[3\frac{\sqrt{1+x}}{x^{3/2}}\ln\left(\sqrt{1+x}-\sqrt{x}\right)+1+\frac{3}{x}\right] \tag{6.81}$$

where $x \equiv (1-\Omega_m)a/\Omega_m$. There may be easier ways to do this (e.g. you might want to check *The Large Scale Structure of the Universe*, §11), but I found it easiest to define a dummy variable $y \equiv \Omega_m/a$; write the integral as

$$\int_{\Omega_m/a}^{\infty} \frac{dy}{y^2(y+1-\Omega_m)^{3/2}} = 2\left[\frac{d}{d\epsilon}\frac{d}{d\lambda}\int_{\Omega_m/a}^{\infty}\frac{dy}{(y+\epsilon)\sqrt{y+\lambda}}\right]_{\epsilon=0,\lambda=1-\Omega_m} \tag{6.82}$$

and then use 2.246 from Gradshteyn and Ryzhik.

6.8 Using the result of Problem (2.9), show that the energy density of one species of massive neutrino is given by equation [6.76].

6-9 One popular way to characterize power on a particular scale is to compute the expected RMS overdensity in a sphere of radius R,

$$\sigma_R^2 \equiv < \delta_R^2(x) > . \tag{6.83}$$

Here

$$\delta_R(\vec{x}) \equiv \int d^3x' \delta(\vec{x}') W_R(\vec{x} - \vec{x}') \tag{6.84}$$

where $W_R(x)$ is the *tophat* window function, equal to one for $x < R$ and zero otherwise; the angular brackets denote the average over all space.

(a) By Fourier transforming, express σ_R in terms of an integral over the power spectrum.

(b) Use the BBKS transfer function to compute σ_8 ($R = 8\mathrm{h}^{-1}\mathrm{Mpc}$) for a standard CDM model ($h = 0.5, n = 1, \Omega_m = 1$). We will see in Chapter 7 that COBE normalization for this model is

$$\delta_H = 1.9 \times 10^{-5}. \tag{6.85}$$

The value of σ_8 you find is yet another sign of the sickness of the model. For galaxies, σ_8 is known to be unity (or less, depending on galaxy type). A model with $\sigma_8 > 1$ then requires galaxies to be less clustered than the dark matter. Present models of galaxy formation suggest that this is unlikely.

(c) In the same model, plot σ_R as a function of R. Since σ_R monotically increases, small scales tend to go non-linear before large scales, the signature of a hierarchical model.

6-10 Rewrite σ from Problem (9) as

$$\sigma_R^2 = \int_0^\infty \frac{dk}{k} \Delta^2(k) \bar{W}_R^2(k), \tag{6.86}$$

where $\bar{W}_R$ is the Fourier tranform of the tophat window function and $\Delta^2 = d\sigma^2/d\ln(k)$ is the contribution to the variance per $ln(k)$. A useful transition point is the value of k at which Δ exceeds one. Scales larger than this are linear, while smaller scales have gone non-linear. Find k_{NL} defined in this way for the sCDM model described in Problem (9).

6-11 Compute the growth factors in a universe with $\Omega_{\mathrm{dark}} = 0.7, \Omega_m = 0.3$ and $w = -0.5$. Compare with the cosmological constant model ($w = -1$) with the same energy.

Chapter 7

Anisotropies

The primordial perturbations set up during inflation manifest themselves in the photons as well as in the matter distribution. By understanding the evolution of the photon perturbations, we can make predictions about the expected anisotropy spectrum today. This evolution is again completely determined by the Einstein-Boltzmann equation we derived in Chapters 3 and 4, and one way to go would be to just simply stick all the relevant equations in those chapters on a computer and solve them numerically. Historically, this is a pretty good caricature of what happened. Long before we developed deep insight into the physics of anisotropies, various groups had codes which determined the expected spectra from different models. Only much later did we come to understand both qualitatively and quantitatively why the spectra look like they do *. In this chapter, I will mangle the history and simply explain what we have learned about the physics of anisotropies.

Perturbations to the photons evolve completely differently before and after the epoch of decoupling at $z \simeq 1100$. Before decoupling the photons are tightly coupled to the electrons and protons; all together they can be described as a single fluid (dubbed the "baryon-photon" fluid). After decoupling, photons free-stream from the "surface of last scattering" to us today. After an overview which qualitatively explains the anisotropy spectrum, sections 7.2-7.4 work through the physics of the baryon-photon fluid before decoupling. Then sections 7.5-7.7 treat the post-decoupling era, culminating in the predicted spectrum of anisotropies today. Finally section 7.8 discusses how these spectra vary when the cosmological parameters change.

7.1 Overview

Let's begin as we did in the last chapter, by cheating and looking at the answers first. Figure 7.2 shows the evolution of four modes corresponding to different wavelengths. Qualitatively, the most important feature of figure 7.2 is that perturbations to the photons do not grow with time. This stands in stark contrast to the matter perturbations which do grow. And this contrast is something we should have expected: the pressure of the photons is so large that

*Understanding the anisotropies actually helped make the codes much more efficient. The prime example of this is the popular code CMBFAST which is based on the analytic solution presented in this chapter.

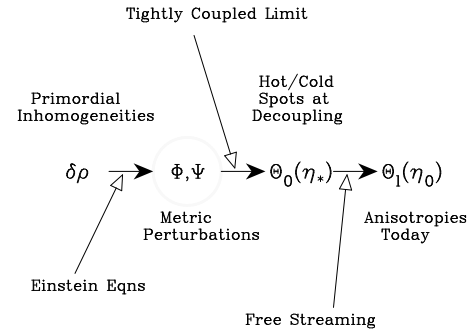


Figure 7.1: The physics of anisotropies. Inhomogeneities in the matter determine the gravitational potentials, as we saw in the last chapter. The potentials affect the photon perturbations at decoupling ($\Theta_0(\eta_*)$); After decoupling, photons freestream, producing anisotropies today, $\Theta_l(\eta_0)$.

it can withstand any tendency towards collapse. This means that the small perturbations set up during inflation stay small; they remain linear all the way up to the present. Before going further and examining the evolution of the different modes in more detail, a technical note: I have plotted not simply the perturbation to the photons but rather the combination $k^{3/2}(\Theta_0 + \Psi)$. The $k^{3/2}$ factor balances the fact that the rms perturbations (in a simple inflationary model) scale as $k^{-3/2}$. I have added the gravitational potential Ψ because the photons we see today had to travel out of the potentials they were in at the time of decoupling. As they emerged from these potential wells, their wavelengths were stretched (if the region was overdense and $\Psi < 0$), thereby decreasing their energy. Thus, the temperature we see today is actually Θ_0 at decoupling plus Ψ.

The large scale mode in figure 7.2 evolves hardly at all. This is not surprising: no causal physics can affect perturbations with wavelengths larger than the horizon, so a super-horizon mode should exhibit little evolution. This means that when we observe large scale anisotropies – which are sensitive to modes with wavelengths larger than the horizon at decoupling – we are observing perturbations in their most pristine form, as they were set down at very early times, presumably during inflation.

Figure 7.2 shows that the smaller scale modes evolve in a more complicated way than the super-horizon modes. Consider the curve labeled "First Peak." As the mode enters the horizon, the perturbation begins to grow until it reaches an apparent maximum at the time

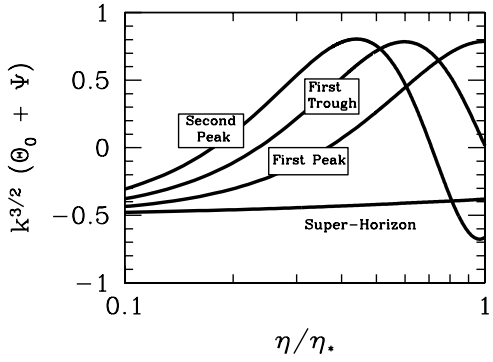


Figure 7.2: Evolution of photon perturbations of four different modes before decoupling. Normalization is arbitrary, but the relaive normalization of the 4 curves is appropriate for perturbations with a Harrison-Zel'dovich ($n = 1$) spectrum. Model is standard CDM with $h = 0.5$, $\Omega_m = 1$, and $\Omega_B = 0.06$. Starting from the bottom left and moving upwards, the wavenumbers for the modes are $k = (7 \times 10^{-4}, 0.022, 0.034, 0.045)$ Mpc^{-1} or $(8, 260, 400, 540)/\eta_0$.

of decoupling. If we observe anisotropies on scales corresponding to this mode, we would expect to see large fluctuations. Hence the label: the anisotropy spectrum will have a peak at the angular scales corresponding to the mode which has just reached its peak at decoupling.

The mode in figure 7.2 which enters the horizon slightly earlier peaks earlier. This mode makes it quite clear that oscillations are being set up by the pressure of the relativistic photons. The phase of this mode is such that, at decoupling, its amplitude is zero. When we observe anistropies today corresponding to these scales, we expect very small fluctuations. There will be a trough in the anisotropy spectrum on these angular scales.

And on it goes. The curve labeled "Second Peak" entered the horizon even earlier and has gone through one full oscillation by decoupling. As such, this mode will have large fluctuations and lead to a second peak in the anisotropy spectrum. You might expect that there will be a never-ending series of peaks and troughs in the anisotropy spectrum corresponding to modes that entered the horizon earlier and earlier. And you would be right: this is exactly what happens.

We can see this more clearly by looking at the spectrum of perturbations at one time, the time of decoupling. Figure 7.3 shows this spectrum for two different models, one with a very low baryon content. We do indeed see this pattern of peaks and troughs. There are two more quantitative features of these oscillations that are important. First, note that – at least in the higher baryon model – the heights of the peaks seem to alternate: the odd peaks seem higher than the even peaks. Second, and this is clearest in the low baryon model, perturbations on small scales $k\eta_0 > 500$ are damped.

To understand the first of these features, we can write down a cartoon version of the

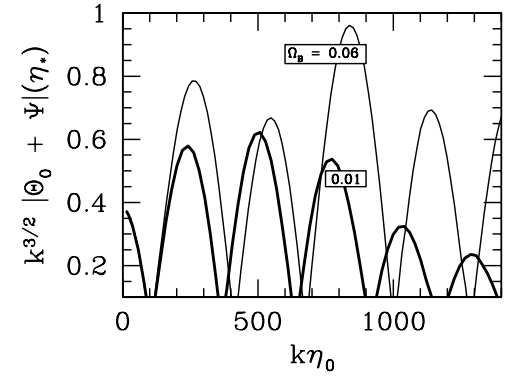


Figure 7.3: Perturbations to the photon distribution at decoupling in two models. The larger damping length of the low Ω_B model is clearly evident in the suppression of perturbations for modes with $k > 500/\eta_0$.

equation governing perturbations. Very roughly, this equation is

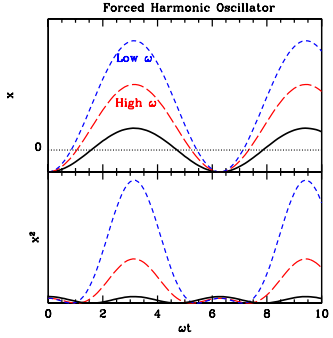
$$\ddot{\Theta}_0 + k^2 c_s^2 \Theta_0 = F \tag{7.1}$$

where F is a driving force due to gravity and c_s is the sound speed of the combined baryon-photon fluid. This is the equation of the forced harmonic oscillator (see box). Qualitatively, it predicts the oscillations we have seen above. But is also explains something about the heights of the peaks. As we add more baryons to the universe, the sound speed goes down (baryons are heavy so they reduce the speed). Thus the frequency of the oscillations goes down. The peaks at $n\pi/\omega$ should be shifted to larger k (you really should read that box!). And, as the frequency goes down, the disparity between the heights of the odd and even peaks gets larger. We clearly see both of these features in figure 7.3. Another way of understanding the alternating peak heights is to note that the perturbations for the first peak mode have been growing since they entered the horizon. By decreasing the pressure (or equivalently increasing the importance of gravity) these modes will grow even more. The second peak mode on the other hand, corresponds to an *underdensity* of photons in the potential wells. Decreasing the pressure makes it harder for photons to escape the well and therefore reduces the magnitude of the perturbation (makes it less underdense).

Consider a simple harmonic oscillator with mass m and force constant k. In addition to the restoring force, the oscillator is acted on by an external force F_0. Thus the full force is $F_0 - kx$ where x is the oscillator's position. The equation of motion is

$$\ddot{x} + \frac{k}{m}x = \frac{F_0}{m}. \tag{7.2}$$

We will shortly use the analogue of this equation when considering small scale CMB anisotropies, so it is useful to reiterate what the different terms stand for. The term on the right hand side – representing the external force – is driving the oscillator to large values of x. The restoring force on the other hand tries to keep the oscillator as close to the origin as possible. The solution therefore will be that oscillations will be set up around a new zero point, at positive x.



The solution to equation [7.2] is the sum of the general solution to the homogeneous equation (with the right hand side set to zero) and a particular solution. The general solution has two modes, best expressed as a sine and cosine with arguments ωt, with the frequency ω defined as $\omega \equiv \sqrt{\frac{k}{m}}$. A particular solution to equation [7.2] is constant $x = F_0/m\omega^2$, so the full solution is the sum of the sine and cosine modes plus this constant. Let us assume that the oscillator is initially at rest. Then, since $\dot{x}(0)$ is proportional to the coefficient of the sine mode, this coefficient must vanish, leaving

$$x = A\cos(\omega t) + \frac{F_0}{m\omega^2}. \tag{7.3}$$

This solution is shown in the figure at right. The solid line is the unforced solution: oscillations about the origin. The dashed curves are the forced solutions for two different choices of frequencies. In both cases, the oscillations are not around $x = 0$ as they would be if the system was unforced. Once an external force is introduced,the zero point of the oscillations shifts in the direction of the force. Two curves are drawn to show that this shift is more dramatic for lower frequencies. The bottom panel shows the square of the oscillator position as a function of time. All three oscillators experience a series of peaks at $t = n\pi/\omega$ corresponding to the minima/maxima of the cosine mode. (Note that if only the sine mode was present these peaks would be at $t = (2n+1)\pi/\omega$.) The heights of these peaks are identical in the case of the unforced oscillator and equal to the height at $t = 0$. In the forced case, though, the height of the odd peaks – those at $t = (2n+1)\pi/\omega$ – is greater than that of the even peaks. The effect is most dramatic for low frequencies. If the frequency is low, the force has a greater effect, producing the greater zero point offset, and hence the greater odd/even disparity. The other feature of this example is that the even peaks correspond to negative positions of the oscillator: points at which it is farthest from where the force wants it to go.

To understand the damping evident in 7.3, we need to remember that the approximation of the photons and electrons and baryons moving together as a single fluid is just that, an approximation. It is valid only if the scattering rate of photons off of electrons is infinite. Of course this is not true: photons travel a finite distance in between scatters. Consider figure 7.4, which depicts the path of a single photon as it scatters off a sea of electrons. It travels

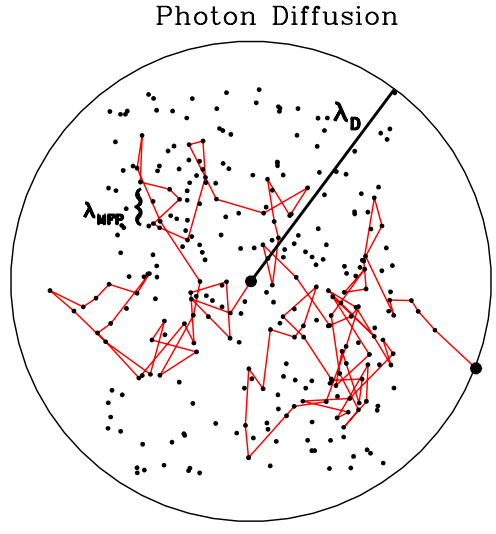


Figure 7.4: Photon diffusion through the electron gas. Electrons are denoted as points. Shown is a typical photon path as it scatters off electrons. The mean free path is λ_{MFP}. After a Hubble time, the photon has scattered many times, so that it has moved a distance of order λ_D.

a mean distance λ_{MFP} in between each scatter. In our case this distance is $1/n_e\sigma_T$. If the density of electrons is very large, then the mean free path is correspondingly small. In the course of a Hubble time, H^{-1}, a photon scatters of order $n_e\sigma_T H^{-1}$ times (simply the product of the rate and the time). As depicted in 7.9, each scatter contributes to the random walk of the photon. We know that the total distance travelled in the course of a random walk is the mean free path times the square root of the total number of steps. Therefore, a cosmological photon moves a mean distance

$$\begin{aligned} \lambda_D &\sim \lambda_{MFP}\sqrt{n_e\sigma_T H^{-1}} \\ &= \frac{1}{\sqrt{n_e\sigma_T H}} \end{aligned} \tag{7.4}$$

Any perturbation on scales smaller than λ_D can be expected to be washed out. In Fourier space this will correspond to damping of all high $k-$ modes. Note that this crude estimate gets the Ω_B dependence right. Models with small baryon density have a larger λ_D (since n_e is proportional to Ω_B when the universe is ionized). Therefore, the damping sets in at larger scales, or smaller k. This is precisely what we saw in figure 7.3.

The final step is to relate the perturbations at decoupling, as depicted in figure 7.3, to the anisotropies we observe today. The math of this is a little complicated, but the physics is perfectly straightforward. Consider one Fourier mode, a plane wave perturbation. Figure 7.5 shows the temperature variations for one mode at deocupling. Photons from hot and cold spots separated by a typical (comoving) distance k^{-1} travel to us coming from an angular separation $\theta \simeq k^{-1}/(\eta_0 - \eta_*)$ where $\eta_0 - \eta_*$ is the (comoving) distance between us and the surface of last scattering[†]. If we decompose the temperature field into multipole moments, then an angular scale θ roughly corresponds to $1/l$. So, using the fact that $\eta_* << \eta_0$, we project inhomogeneities on scales k onto anisotropies on angular scales $l \simeq k\eta_0$.

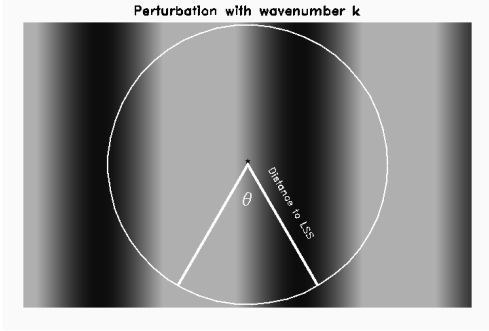


Figure 7.5: Freestreaming. Perturbations in the temperature (Θ_0) at decoupling from one plave wave with wavenumber k. Hot and cold spots are shaded light and dark. After decoupling, photons from the hot and cold spots travel freely to us, here denoted by the star at the center. This mode contributes anisotropy on a scale $\theta \sim k^{-1}$/(Distance to last scattering surface).

There is one final caveat to this picture of freestreaming. We have been implicitly assuming that nothing happens to the photons on their journey from the last scattering surface to Earth. In fact, if the universe was flat and matter dominated through this whole time, then gravitational potentials remain constant, and this assumption is correct. However, decoupling takes places not too much later than the epoch of equality, so the remnant radiation density means potentials are not exactly constant right after deocupling. Also, at late times, dark energy does not behave like matter and leads to potential decay. You can imagine

[†]This is true only in a flat universe. In an open universe, the distance to the last scattering surface is larger, so the same physical scale is projected onto a smaller angular scale. See Problem (1)

other disruptions to matter domination. All of these so-called *Integrated Sachs-Wolfe Effects* produce new perturbations to the photons, leading to changes typically of order ten percent.

And that's it; we now understand how primordial perturbations are processed to form the present day anisotropy spectrum. Let's work through it again quantitatively.

7.2 Large Scale Anisotropies

To find the large scale solution for the photon perturbation, we make use of the superhorizon equation, [6.16]. This immediately tells us that $\Theta_0 = -\Phi$ plus a constant. The initial conditions are such that $\Theta_0 = \Phi/2$, so the constant is $3\Phi(a = 0)/2$. We have an exact expression for the large scale evolution of Φ, equation [6.31]. If decoupling takes place long after the epoch of equality, then we can take the $y >> 1$ limit of this expression, $\Phi \to 9\Phi(0)/10$. Therefore, at decoupling, large scale photon perturbations satisfy

$$\Theta_0(k, \eta_*) = \frac{3\Phi(k, 0)}{5} \tag{7.5}$$

$$= \frac{2\Phi(k, \eta_*)}{3}. \tag{7.6}$$

The observed anisotropy is $\Theta_0 + \Psi$, which to a good approximation is $\Theta_0 - \Phi$. Therefore,

$$(\Theta_0 + \Psi)(k, \eta_*) = \frac{1}{3}\Psi(k, \eta_*). \tag{7.7}$$

Another useful way of expressing the large scale perturbations at decoupling is in terms of the density field. The initial conditions derived in Chapter 5 were that $\delta_{\rm DM} = 3\Phi/2$. Integrating the large scale evolution equation, $\dot{\delta}_{\rm DM} = -3\dot{\Phi}$, leads to

$$\delta_{\rm DM}(\eta_*) = \frac{3}{2}\Phi(0) - 3[\Phi(\eta_*) - \Phi(0)]$$

$$= 2\Phi(\eta_*). \tag{7.8}$$

So the observed anisotropy expressed in terms of the dark matter overdensity is

$$(\Theta_0 + \Psi)(k, \eta_*) = -\frac{1}{6}\delta_{\rm DM}(\eta_*). \tag{7.9}$$

Equations [7.7] and [7.9] will be useful to us when we compute the large scale anisotropy spectrum. However, even now, they contain a fascinating piece of information. From the Fourier transform of equation [7.9], we see that the observed anisotropy of an overdense region will be *negative*. This is such a surprising result that it is worth repeating. For large scale perturbations, overdense regions do indeed contain hotter photons at decoupling than do underdense regions: i.e. $\Theta_0 > 0$ when $\Psi < 0$. However, to get to us today, these photons must travel out of their potential wells. In so doing they lose energy, and this energy loss more than compensates for the fact that the photons were initially hotter than average: i.e.

$\Theta_0 + \Psi$ is negative when $\Psi < 0$. To sum up, when we observe large scale hot spots on the sky today, we are actually observing regions that were underdense at the time of decoupling.

The other important feature of equation [7.9] is the coefficient 1/6. It enables us to relate "$\delta T/T$" to "$\delta\rho/\rho$." Very roughly speaking, an anisotropy of order 10^{-5} corresponds to an overdensity of 6×10^{-5}. One of the important questions which must be addressed by the picture of gravitational instability is whether the observed anisotropy is consistent with the overdensities needed to form structure by today. This factor of six is a huge help. Almost all models of structure formation other than inflation, this factor of six is replaced by a number much closer to unity (see e.g. Problem (2)). Therefore, they struggle with the fact that the observed level of anisotropy is too small to account for the clustering of matter in the universe (in our language the matter power spectrum is normalized too low).

7.3 Acoustic Oscillations

7.3.1 Tightly Coupled Limit of the Boltzmann Equations

When all electrons were ionized, before η_*, the mean free path for a photon was much smaller than the horizon of the universe. Compton scattering caused the electron-proton fluid to be tightly coupled with the photons. We now proceed to explore this regime quantitatively using the Boltzmann equations.

The tightly coupled limit corresponds to the scattering rate being much larger than the expansion rate: $\tau >> 1$, where τ is the optical depth defined in equation [3.57]. I want to argue that in the $\tau >> 1$ limit, the only non-negligible moments, Θ_l, are the monopole ($l = 0$) and the dipole ($l = 1$). All others are suppressed. In this sense, photons behave just like a fluid, which can be described with only two variables: the density ρ and the velocity $\vec{v}$. In order to show this, let's go back to the Boltzmann equation ([3.96]) for photons. We want to turn this differential equation for $\Theta(\eta, \mu)$ into an infinite set of coupled equations for $\Theta_l(\eta)$. The advantage is that – as we will see – the higher moments are small and so can be neglected. The strategy is to multiply by $P_l(\mu)$ and then integrate over μ. Using equation [3.104], the Boltzmann equation for $l > 2$ becomes

$$\dot{\Theta}_l + \frac{k}{(-i)^{l+1}} \int_{-1}^{1} \frac{d\mu}{2} \mu P_l(\mu) \Theta(\mu) = \dot{\tau}\Theta_l. \tag{7.10}$$

Note that all other terms (e.g. $-\dot{\Phi}$) have simple μ dependence (scale as μ^0 or μ^1) so all $l > 2$ moments vanish for them. To do the integral, we make use of the recurrence relation for Legendre polynomials,

$$\mu P_l(\mu) = \frac{l}{2l+1} P_{l-1} + \frac{l+1}{2l+1} P_{l+1}. \tag{7.11}$$

Thus,

$$\dot{\Theta}_l - \frac{kl}{2l+1}\Theta_{l-1} + \frac{k(l+1)}{2l+1}\Theta_{l+1} = \dot{\tau}\Theta_l. \tag{7.12}$$

Let us consider the order of magnitude of the terms in equation [7.12]. The first term on the left is of order Θ_l/η which is much smaller than the term on the right which is of order

$\tau\Theta_l/\eta$. Neglecting the Θ_{l+1} term for the moment, this tells us that in the tightly coupled regime

$$\Theta_l \sim \frac{k\eta}{2\tau}\Theta_{l-1}. \tag{7.13}$$

For modes of order the horizon $k\eta \sim 1$, this means that $\Theta_l << \Theta_{l-1}$. (By the way, this is justification for throwing out the Θ_{l+1} term in making our estimate.) This estimate is valid for all modes higher than the dipole, so all such modes are very small compared to the monopole and dipole.

Before making use of this fact and deriving the tightly coupled equations in the limit in which only the monopole and dipole are non-zero (the fluid approximation), I want to explain *why* higher moments are damped in a tightly coupled environment. Indeed this observation is extremely important not only in cosmology but in all settings in which the fluid approximation is used. To understand the fluid approximation, consider one plane wave perturbation as depicted in figure 7.6. An observer sitting at the center of the perturbation sees photons arriving from a distance of order η/τ. A wavelength of order the horizon is much larger than this distance, so the photons arriving at the observer all have the same temperature. There is very little anisotropy. You might think that a perturbation with a very small wavelength (with $k\eta \sim \tau$) would lead to anisotropy. In fact, though, such a mode has a wavelength much smaller than the damping scale. So all perturbations on such small scales are smoothed out, again leading to no anisotropy. The bottom line is there is essentially no anisotropy beyond the monopole and the dipole in the tightly coupled regime.

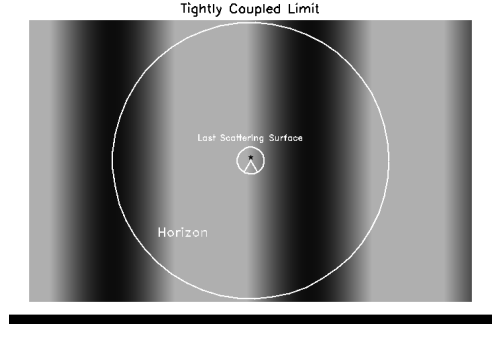


Figure 7.6: Anisotropies in the tightly coupled era. Perturbations on the scale of the horizon cannot be observed by an observer [denoted by the star here]. For, the photons observed come from the last scattering surface a distance η/τ away. This last scattering surface is so close that photons arriving from all angles have virtually identical temperatures.

Armed with this knowledge, we can now return to the equations for the first two moments, which – after disposing of Θ_2 – read:

$$\dot{\Theta}_0 + k\Theta_1 = -\dot{\Phi} \tag{7.14}$$

$$\dot{\Theta}_1 - \frac{k\Theta_0}{3} = \frac{k\Psi}{3} + \dot{\tau}(\Theta_1 - \frac{iv_{\rm B}}{3}). \tag{7.15}$$

These are supplemented by the equations for the electron-baryon fluid, equations (3.101) and (3.102). Let us first rewrite the velocity equation, equation [3.102], as

$$v_{\rm B} = -3i\Theta_1 + \frac{R}{\dot{\tau}}\left[\dot{v}_{\rm B} + \frac{\dot{a}}{a}v_{\rm B} + ik\Psi\right]. \tag{7.16}$$

The second term on the right here is much smaller than the first since it is suppressed by a relative factor of order τ^{-1}. Thus, to lowest order, $v_{\rm B} = -3i\Theta_1$. A systematic way to expand, then, is to use this lowest order expression everywhere in the second term, leading to

$$v_{\rm B} \simeq -3i\Theta_1 + \frac{R}{\dot{\tau}}\left[-3i\dot{\Theta}_1 - 3i\frac{\dot{a}}{a}\Theta_1 + ik\Psi\right]. \tag{7.17}$$

Now let us insert this expression into equation [7.15], eliminating $v_{\rm B}$. After rearranging terms, we find

$$\dot{\Theta}_1 + \frac{\dot{a}}{a}\frac{R}{1+R}\Theta_1 - \frac{k}{3[1+R]}\Theta_0 = \frac{k\Psi}{3}. \tag{7.18}$$

We now have two first order coupled equations for the first two photon moments, equations (7.14) and (7.18). We can turn these into one second order equation by differentiating equation [7.14] and using equation [7.18] to eliminate $\dot{\Theta}_1$:

$$\ddot{\Theta}_0 + k\left[\frac{k\Psi}{3} - \frac{\dot{a}}{a}\frac{R}{1+R}\Theta_1 + \frac{k}{3[1+R]}\Theta_0\right] = -\ddot{\Phi}. \tag{7.19}$$

Finally, we use equation [7.14] to eliminate Θ_1 here. This leaves:

$$\ddot{\Theta}_0 + \frac{\dot{a}}{a}\frac{R}{1+R}\dot{\Theta}_0 + k^2c_s^2\Theta_0 = -\frac{k^2}{3}\Psi - \frac{\dot{a}}{a}\frac{R}{1+R}\dot{\Phi} - \ddot{\Phi} \equiv F(k,\eta) \tag{7.20}$$

where I have defined the forcing function on the right as F and the sound speed of the fluid as

$$c_s \equiv \sqrt{\frac{1}{3(1+R)}}. \tag{7.21}$$

The sound speed depends on the baryon density in the universe. In the absence of baryons, it has the standard value for a relativistic fluid, $c_s = 1/\sqrt{3}$. The presence of baryons though makes the fluid heavier, thereby lowering the sound speed. We will see shortly that the fluid oscillates in both space and time, with a period which is determined by the sound speed, and hence by the baryon density. Note that equation [7.20] is the "grown-up" version of equation [7.1]; it differs only through the $\dot{\Theta}_0$ damping[‡] term (see Problem (3)). The presence of this term does not change any of the qualitiative conlcusions we reached in section 7.1. Finally, note that Φ enters on the right in a very similar way as Θ_0 does on the left. An alternate version of equation [7.20] takes adavantage of this:

$$\left\{\frac{d^2}{d\eta^2} + \frac{\dot{R}}{1+R}\frac{d}{d\eta} + k^2c_s^2\right\}[\Theta_0 + \Phi] = \frac{k^2}{3}\left[\frac{1}{1+R}\Phi - \Psi\right]. \tag{7.22}$$

[‡]This "damping" term is not to be confused with the damping of perturbations on small scales treated in the next section. Completely different effects.

7.3.2 Tightly Coupled Solutions

The equation we have derived governing acoustic oscillations of the photon-baryon fluid, [7.22], is a second order ordinary differential equation. To solve it, we will again use the Green's method to find the full solution. First we find the two solutions to the homogeneous equation. Then we use these to construct the particular solution.

In prnciple, to obtain the homogeneous solution, we must solve the damped, harmonic oscillator equation, [7.22] with the right hand side equal to zero. In practice, the damping term is of order $R(\Theta_0+\Phi)/\eta^2$ while the pressure term is much larger, of order $k^2c_s^2(\Theta_0+\Phi)$ (at least it's larger when modes are within the horizon or when R is small). Physically we expect pressure to induce oscillations in the photon temperature; the time scale for these oscillations is much shorter than the damping introduced by the expansion of the universe. To a first approximation, then, let us neglect the damping term and simply obtain the oscillating solutions. You can rectify this by applying the WKB approximation in Problem (6). In this limit, the two homogeneous solutions are

$$S_1(k,\eta) = \sin[kr_s(\eta)] \quad ; \quad S_2(k,\eta) = \cos[kr_s(\eta)] \tag{7.23}$$

where I have defined the *sound horizon* as

$$r_s(\eta) \equiv \int_0^\eta d\eta' c_s(\eta'). \tag{7.24}$$

The sound horizon is the comoving distance travelled by a sound wave by time η. The tightly coupled solution for the photon temperature can be constructed from the homogeneous solutions of equation [7.23]:

$$\begin{aligned}\Theta_0(\eta) &+ \Phi(\eta) = C_1S_1(\eta) + C_2S_2(\eta)\\ &+ \frac{k^2}{3}\int_0^\eta d\eta'\,[\Phi(\eta') - \Psi(\eta')]\frac{S_1(\eta')S_2(\eta) - S_1(\eta)S_2(\eta')}{S_1(\eta')\dot{S}_2(\eta') - \dot{S}_1(\eta')S_2(\eta')}\end{aligned} \tag{7.25}$$

Here again, I have dropped all powers of $1+R$ except in the argument of the rapidly varying sines and cosines. We can fix the constants C_1 and C_2 in equation [7.25] by appealing to the initial conditions, when both Θ_0 and Φ are constants. The coefficient of the sin term therefore, C_1, must vanish, and $C_2 = \Theta_0(0) + \Phi(0)$. The denominator in the integrand reduces to $-kc_s(\eta') \to -k/\sqrt{3}$ in the limit in which we are working. Finally, the difference of the products in the numerator of the integrand is simply $-\sin[k(r_s - r_s')]$, so

$$\begin{aligned}\Theta_0(\eta) + \Phi(\eta) &= [\Theta_0(0) + \Phi(0)]\cos(kr_s)\\ &+ \frac{k}{\sqrt{3}}\int_0^\eta d\eta'\,[\Phi(\eta') - \Psi(\eta')]\sin[k(r_s(\eta) - r_s(\eta'))].\end{aligned} \tag{7.26}$$

Equation [7.26] is an expression for the anisotropy in the tightly coupled limit, first derived by Hu & Sugiyama in 1995. If you are not impressed with this solution since it still involves an integral over the gravitational potentials, I urge you to reconsider. First, look at figure 7.7 which compares the solution of equation [7.26] with exact results obtained by

integrating the full set of coupled Einstein-Boltzmann equations. The approximate solution gets the peak locations dead on, and it does fairly well with the heights as well. The later peaks – those at $k\eta_0 > 500$ are clearly overestimated by our solution, but we will shortly rectify this when we include damping due to diffusion in the next section. A second reason to respect the approximate solution is that it divides the problem neatly into (i) first a calculation of the external gravitional potentials generated by the dark matter and then (ii) the effect of these potentials on the anisotropies. Third, the solution clearly illustrates that the cosine mode is the one excited by inflationary models. This is important, because it is very hard to imagine this mode excited by any other mechanism. If causality is respected, then there should be no perturbations with $k\eta << 1$ early on. We know that inflation evades this constraint by changing the true horizon; it is tempting to say that if this mode is observed, we are seeing evidence for inflation. And the final reason equation [7.26] is impressive is that the full set of Einstein-Boltzmann equations involve literally thousands of coupled variables (e.g. the Θ_l's). Reducing those thousands of differnetial equations to just one is a huge leap in knowledge.

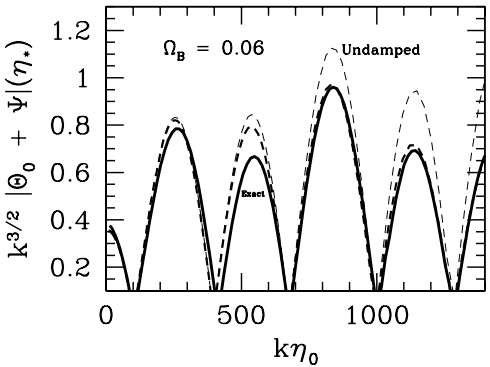


Figure 7.7: The monopole at decoupling in a standard CDM model. The exact solution is the heavily weighted solid line. The light dashed line is the undamped solution of this section, equation [7.26]; the heavier curve in the middle accounts for damping using the treatment of the next section.

In addition to the monopole, the photon distribution has a non-negligible dipole at decoupling. Using equation [7.14], we can obtain an analytic solution for the dipole by differentiating equation [7.26]:

$$\begin{aligned}\Theta_1(\eta) &= \frac{1}{\sqrt{3}}\left[\Theta_0(0)+\Phi(0)\right]\sin(kr_s) \\ &- \frac{k}{3}\int_0^\eta d\eta'\left[\Phi(\eta')-\Psi(\eta')\right]\cos\left[k(r_s(\eta)-r_s(\eta'))\right]. \end{aligned} \tag{7.27}$$

The first term is completely out of phase with the monopole ($sin(kr_s)$ vs. $\cos(kr_s)$). Figure 7.8 shows that this feature remains even after accounting for the integral term. This mismatch of phase will have important implications for the final anisotropy spectrum.

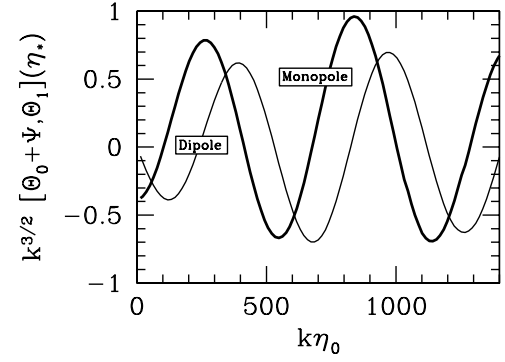


Figure 7.8: The monopole and dipole at decoupling in a standard CDM model. The dipole vanishes for the longest wavelength modes that have not enetered the horizon by decoupling. It is completely out of phase with the monopole.

7.4 Diffusion Damping

Figure 7.7 makes it clear that we must account for diffusion to get accurate CMB spectra. To analyze diffusion quantitatively, we must return to the equations for the moments of the photon distribution, equations (7.14), (7.15) and (7.12). Until now, we have neglected Θ_2 and all higher moments. Diffusion is characterized by a small, but non-negligible quadrupole.

We must therefore supplement the set of equations we wrote down in the last section with an equation for the quadrupole, Θ_2. Our task is somewhat simplified by the fact that we will be interested in phenomena occurring only on small scales. On these scales, the potentials are irrelevant, so we can drop Φ and Ψ everywhere. Also, we will see that diffusion manifests itself in the moments by making each successive moment proportional to a higher power of $1/\dot{\tau}$. Thus we will need to keep only the $l = 2$ mode; all higher ones can be neglected. With these approximations, we have

$$\dot{\Theta}_0 + k\Theta_1 = 0 \tag{7.28}$$

$$\dot{\Theta}_1 + k\left(\frac{2}{5}\Theta_2 - \frac{1}{3}\Theta_0\right) = \dot{\tau}\left(\Theta_1 - \frac{iv_B}{3}\right) \tag{7.29}$$

$$\dot{\Theta}_2 - \frac{2k}{5}\Theta_1 = \frac{9}{10}\dot{\tau}\Theta_2. \tag{7.30}$$

These three equations need to be supplemented by an equation for $v_{\rm B}$. This is best expressed as a slight rewriting of equation [3.102]:

$$3i\Theta_1 + v_{\rm B} = \frac{R}{\dot\tau}\left[\dot v_{\rm B} + \frac{\dot a}{a} v_{\rm B}\right], \tag{7.31}$$

where again I have dropped the gravitational potential.

To solve this set of equations, we appeal to the high frequency nature of damping. Let us write the time dependence of the velocity as

$$v_{\rm B} \propto e^{i\int \omega d\eta} \tag{7.32}$$

and similarly for all other variables. We already know that $\omega \simeq kc_s$ in the tightly coupled limit. Now we are searching for damping, an imaginary part to ω. Since damping occurs on small scales, or high frequencies,

$$\dot v_{\rm B} = i\omega v_{\rm B} >> \frac{\dot a}{a} v_{\rm B}; \tag{7.33}$$

the latter is of order $v_{\rm B}/\eta$ while the former is of order $kv_{\rm B}$. The velocity equation then becomes

$$\begin{aligned} v_{\rm B} &= -3i\Theta_1\left[1 - \frac{i\omega R}{\dot\tau}\right]^{-1} \\ &\simeq -3i\Theta_1\left[1 + \frac{i\omega R}{\dot\tau} - \left(\frac{\omega R}{\dot\tau}\right)^2\right] \end{aligned} \tag{7.34}$$

where I have expanded out to $\dot\tau^{-2}$ because $v_{\rm B} + 3i\Theta_1$ is multiplied by $\dot\tau$ in equation [7.29].

The equation for the second moment of the photon field, equation [7.30], can be reduced similarly. First we can drop the $\dot\Theta_2$ term since it is much smaller than $\dot\tau\Theta_2$. This leaves simply

$$\Theta_2 = -\frac{4k}{9\dot\tau}\Theta_1 \tag{7.35}$$

which shows that our approximation scheme is controlled: higher moments are suppressed by additional powers of $k/\dot\tau$. The equation for the zeroth moment becomes:

$$i\omega\Theta_0 = -k\Theta_1. \tag{7.36}$$

Inserting all of these into equation [7.29] gives the dispersion relation

$$i\omega - \frac{8k^2}{27\dot\tau} + (k^2/3i\omega) = \dot\tau\left(1 - \left[1 + \frac{i\omega R}{\dot\tau} - \left(\frac{\omega R}{\dot\tau}\right)^2\right]\right). \tag{7.37}$$

Collecting terms we get

$$\omega^2(1+R) - \frac{k^2}{3} + \frac{i\omega}{\dot\tau}\left[\omega^2R^2 + \frac{8k^2}{27}\right] = 0. \tag{7.38}$$

The first two terms on the left , the leading ones in the expansion of $1/\dot\tau$, recover the result of the previous section, that the frequency is the wavenumber times the speed of sound. We can write the frequency as this zero order piece plus a first order correction, $\delta\omega$. Then, inserting the zero order part into the terms inversely proportional to $\dot\tau$ leads to

$$\delta\omega = -\frac{ik^2}{2(1+R)\dot\tau}\left[c_s^2R^2 + \frac{8}{27}\right]. \tag{7.39}$$

Therefore, the time dependence of the perturbations is

$$\Theta_0, \Theta_1 \sim \exp\left\{ik\int d\eta c_s - \frac{k^2}{k_D^2}\right\} \tag{7.40}$$

where the damping wavenumber is defined via

$$\frac{1}{k_D^2(\eta)} \equiv \int_0^\eta \frac{d\eta'}{6(1+R)n_e\sigma_T a(\eta')}\left[\frac{R^2}{(1+R)} + \frac{8}{9}\right]. \tag{7.41}$$

Putting aside factors of order unity, this equation says that $1/k_D \sim [\eta/n_e\sigma_T a]^{1/2}$ which agrees with our heuristic estimate at the beginning of this chapter.

As a first estimate of the damping scale, we can work in the pre-recombination limit, in which all electrons (except those in helium) are free. In Chapter 2 we estimated the optical depth in this limit, but ignored helium. The mass fraction of helium is usually denoted as Y_p and is approximately 0.24. Since each helium atom contains four nucleons, the ratio of helium atoms to the total number of atoms is approximately $Y_p/4$. Each of these absorbs two electrons (one for each proton), so when counting the number of free electrons before hydrogen recombination, we must multiply our estimate of equation [2.51] by $1-Y_p/2$. Using the fact that $H_0 = 3.33\times10^{-4}$ h Mpc^{-1}, we have, in the pre-recombination limit,

$$n_e\sigma_T a = 2.3\times10^{-5}{\rm Mpc}^{-1}\Omega_B h^2 a^{-2}\left(1 - \frac{Y_p}{2}\right). \tag{7.42}$$

Using this, you can show (Problem (9)) that an approximation for the damping scale is

$$k_D^{-2} = 3.1\times10^6{\rm Mpc}^2 a^{5/2} f_D(a/a_{\rm EQ})\left(\Omega_B h^2\right)^{-1}\left(1 - \frac{Y_p}{2}\right)^{-1}\left(\Omega_m h^2\right)^{-1/2} \tag{7.43}$$

where f_D, defined in equation [7.88], goes to one as $a/a_{\rm EQ}$ gets large.

Figure 7.9 shows the evolution of the damping scale before recombination. Neglecting recombination is a good approximation at early times but, as expected, leads to quantitative errors right near η_*, when using equation [7.42] for the free electron density does not accurately account for the electrons swooped up into neutral hydrogen. In the absence of recombination, k_D scales as $\Omega_B^{1/2}$. Note from the late time behavior in figure 7.9 that the messy details of recombination change this simple scaling: k_D for the $\Omega_B = 0.06$ case is less than $2^{0.5}$ as big as the $\Omega_B = 0.03$ case.

Figure 7.9 suggests one final comment. The damping of anisotropies due to photon diffusion is sometimes referred to as being caused by the "finite thickness of the last scattering

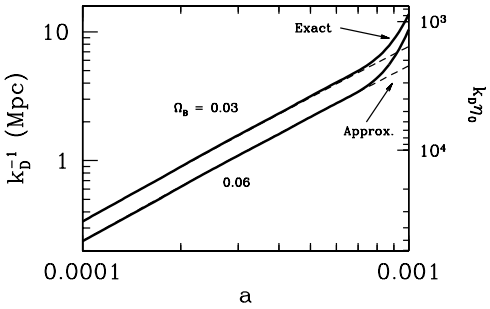


Figure 7.9: Damping scale as a function of the scale factor for two different values of Ω_B (with $h = 0.5$). Heavy curves (exact) numerically integrate over the standard recombination history, while light curves are the approximation of equation [7.43] which assume electrons remain ionized. Right axis shows the equivalent $k_D\eta_0$; damping occurs on angular scales $l > k_D\eta_0$.

surface." That is, it is argued that if recombination took place instantaneously at η_*, then there would be no damping. Figure 7.9 shows that this is patently false. Even if recombination occurred in this way, the universe before recombination is *not* inifinitely optically thick. Photons can still stream a reasonable distance and hence damp anisotropies. In the examples shown, the damping scale would have been smaller (larger l) by less than a factor of two if recombination had occurred instantaneously.

7.5 Inhomogeneities to Anisotropies

We now have a good handle on the perturbations to the photons at decoupling. It is time to transform this understanding into predictions for the anisotropy spectrum today. First, we will solve for the moments Θ_l today in the next subsection. Then we will spend a bit of time relating these moments to the observables. Thus the main purpose of the following subsections is to derive equation [7.57], which relates the moments today to the monopole and dipole at deocupling, and equation [7.68], which expresses the CMB power spectrum in terms of the Fourier moments today.

7.5.1 Free Streaming

We want to derive a formal solution for the photon moments today $\Theta_l(\eta_0)$ in terms of the monopole and dipole at decoupling. A formal solution can be obtained by returning to equation [3.96]. Subtracting $\dot{\tau}\Theta$ from both sides leads to

$$\dot{\Theta} + (ik\mu - \dot{\tau})\Theta = e^{-ik\mu\eta+\tau}\frac{d}{d\eta}\left[\Theta e^{ik\mu\eta-\tau}\right] = \bar{S} \tag{7.44}$$

where the source function is defined as

$$\bar{S} \equiv -\dot{\Phi} - ik\mu\Psi - \dot{\tau}\left[\Theta_0 + \mu v_{\rm B} - \frac{1}{2}P_2(\mu)\Pi\right]. \tag{7.45}$$

Hold your curiosity about the $\bar{}$ in the definition. Multplying both sides of equation [7.44] by the exponential and then integrating over η leads directly to

$$\Theta(\eta_0) = \Theta(\eta_{\rm init})e^{ik\mu(\eta_{\rm init}-\eta_0)}e^{-\tau(\eta_{\rm init})} + \int_{\eta_{\rm init}}^{\eta_0} d\eta \bar{S}(\eta)e^{ik\mu(\eta-\eta_0)-\tau(\eta)} \tag{7.46}$$

where I have used the fact that $\tau(\eta = \eta_0) = 0$ since τ is defined as the scattering rate integrated from η up to η_0. We also know that, if the initial time $\eta_{\rm init}$ is early enough, then the optical depth $\tau(\eta_{\rm init})$ will be extremely large. Therefore, the first term on the right side of equation [7.46] vanishes. This corresponds to the fact that any initial anisotropy is completely erased by Compton scattering. By the same reasoning, we can set the lower limit on the integral to zero: any contribution to the integrand from $\eta < \eta_{\rm init}$ is completely negligible. Thus, the solution for the perturbations is

$$\Theta(k, \mu, \eta_0) = \int_0^{\eta_0} d\eta \bar{S}(k, \mu, \eta)e^{ik\mu(\eta-\eta_0)-\tau(\eta)} \tag{7.47}$$

Equation 7.47 looks simple, but of course all of the complication is hidden in the source function $\bar{S}$. Notice that $\bar{S}$ depends somewhat on the angle μ. If it did *not* depend on μ, we could immediately turn equation [7.47] into an equation for each of the Θ_l's. For, we could multiply each side by the Legendre polynomial $P_l(\mu)$ and then integrate over all μ. By equation [3.104], the left side would give $(-i)^l\Theta_l$ and the right would contain the integral

$$\int_{-1}^{1}\frac{d\mu}{2}P_l(\mu)e^{ik\mu(\eta-\eta_0)} = \frac{1}{(-i)^l}j_l\left[k(\eta-\eta_0)\right] \tag{7.48}$$

where j_l is a spherical Bessel function. This approach looks so promising that we should pursue it to its end, again forgetting for the moment that $\bar{S}$ really does have some μ dependence. The expression for Θ_l would be

$$\Theta_l(k, \eta_0) = (-1)^l\int_0^{\eta_0} d\eta \bar{S}(k, \eta)e^{-\tau(\eta)}j_l\left[k(\eta-\eta_0)\right]. \tag{7.49}$$

What about the μ dependence in $\bar{S}$? We can account for this by noting that $\bar{S}$ multiplies the exponential $e^{ik\mu(\eta-\eta_0)}$ in equation [7.47]. Thus, everywhere we encounter a factor of μ in $\bar{S}$ we can replace it with a time derivative:

$$\mu \to \frac{1}{ik}\frac{d}{d\eta}. \tag{7.50}$$

Let me demonstrate this explicitly with the $-ik\mu\Psi$ term in $\bar{S}$. The integral is

$$\begin{aligned} -ik\int_0^{\eta_0} d\eta\ \mu\Psi e^{ik\mu(\eta-\eta_0)-\tau(\eta)} &= -\int_0^{\eta_0} d\eta\Psi e^{-\tau(\eta)}\frac{d}{d\eta}e^{ik\mu(\eta-\eta_0)} \\ &= \int_0^{\eta_0} d\eta e^{ik\mu(\eta-\eta_0)}\frac{d}{d\eta}\left[\Psi e^{-\tau(\eta)}\right] \end{aligned} \tag{7.51}$$

where the last line follows by integration by parts. Note that the surface terms can be dropped: at $\eta = 0$ they are damped by the $e^{-\tau(0)}$ factor. The terms at $\eta = \eta_0$ are not small, but they are irrelevant since they have no angular dependence. They alter the monopole, an alteration which we cannot detect. Thus, accounting for the integration by parts changes the substitution rule of equation [7.50] by a minus sign, with the understanding that the derivative does *not* act on the oscillating part of the exponential, $e^{ik\mu(\eta-\eta_0)}$. The solution in equation [7.49] therefore becomes:

$$\Theta_l(k,\eta_0) = \int_0^{\eta_0} d\eta S(k,\eta) j_l\left[k(\eta_0-\eta)\right] \tag{7.52}$$

with the source function now defined as

$$\begin{aligned} S(k,\eta) &\equiv e^{-\tau}\left[-\dot{\Phi} - \dot{\tau}(\Theta_0 + \frac{1}{4}\Pi)\right] \\ &+ \frac{d}{d\eta}\left[e^{-\tau}\left(\Psi - \frac{iv_{\rm B}\dot{\tau}}{k}\right)\right] - \frac{3}{4k^2}\frac{d^2}{d\eta^2}\left[e^{-\tau}\dot{\tau}\Pi\right] \end{aligned} \tag{7.53}$$

In equation [7.52], I have also used the property of spherical Bessel functions: $j_l(x) = (-1)^l j_l(-x)$.

At this stage, it is useful to introduce the *visibility function*

$$g(\eta) \equiv -\dot{\tau}e^{-\tau}. \tag{7.54}$$

The visibility function has some interesting properties. The integral $\int_0^{\eta_0} d\eta g(\eta) = 1$, so we can think of it as a probability density. It is the probability that a photon last scattered at η. In the standard recombination, since τ is so large early on, this probability is essentially zero for η earlier than the time of recombination. It also declines rapidly after recombination, because the prefactor $-\dot{\tau}$, which is the scattering rate, is quite small. Figure 7.10 shows the visibility function for two values of the baryon density.

The source function in equation [7.53] can now be expressed in terms of the visibility function. If we drop the polarization tensor Π in the source since it is very small, then the source function becomes

$$\begin{aligned} S(k,\eta) &\simeq g(\eta)\left[\Theta_0(k,\eta) + \Psi(k,\eta)\right] \\ &+ \frac{d}{d\eta}\frac{iv_{\rm B}(k,\eta)g(\eta)}{k} \\ &+ e^{-\tau}\left[\dot{\Psi}(k,\eta) - \dot{\Phi}(k,\eta)\right]. \end{aligned} \tag{7.55}$$

We can take our analytic solution one step further by performing the time integral in equation [7.52]. The source term proportional to $v_{\rm B}$ is best treated by integrating by parts. Then,

$$\begin{aligned} \Theta_l(k,\eta_0) &= \int_0^{\eta_0} d\eta\, g(\eta)\ \left[\Theta_0(k,\eta) + \Psi(k,\eta)\right]\ j_l\left[k(\eta_0-\eta)\right] \\ &- \int_0^{\eta_0} d\eta\, g(\eta)\ \frac{iv_{\rm B}(k,\eta)}{k}\ \frac{d}{d\eta} j_l\left[k(\eta_0-\eta)\right] \\ &+ \int_0^{\eta_0} d\eta\, e^{-\tau}\left[\dot{\Psi}(k,\eta) - \dot{\Phi}(k,\eta)\right]\ j_l\left[k(\eta_0-\eta)\right] \end{aligned} \tag{7.56}$$

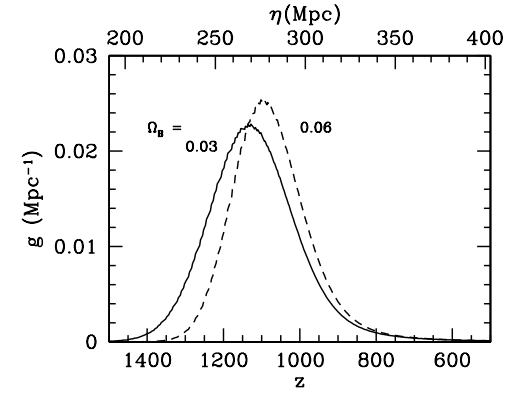


Figure 7.10: The visibility function. Most electrons last scatter at around $z \simeq 1100$ with little dependence on the baryon density. Note that the integral of g over conformal time is one. Here $h = 0.5$.

There are two types of terms in equation [7.56]. First, there are those wherein the integral is weighted by $e^{-\tau}$. These contribute as long as $\tau < 1$, that is, at all times after recombination. Note that these are also proportional to derivatives of the potentials. If the potentials are constant after recombination, these terms vanish. In many theories, as we saw in Chapter 6, this is precisely what happens: the universe is purely matter-dominated after recombination and in such an environment, the potentials generally remain constant. The corrections due to changing potentials are therefore important to get things right quantitatively, but do not affect the qualitative structure of the anisotropy spectrum. Rather, the dominant terms in equation [7.56] are the ones with integrals weighted by the visibility function.

Since the visibility function is so sharply peaked, the integrals in the first two terms become very simple. To see why, consider figure 7.11 which shows the three parts of the integrand of the the term (the monopole) in equation [7.56]. Since the visibility function changes rapidly compared with the other two functions, we can evaluate those other functions at the peak of the visibility function, i.e. at $\eta = \eta_*$, and remove them from the integral. But then, the integral is simply $\int d\eta g(\eta) = 1$. Thus, we are left with

$$\Theta_l(k,\eta_0) \simeq \left[\Theta_0(k,\eta_*) + \Psi(k,\eta_*)\right]\ j_l\left[k(\eta_0-\eta_*)\right]$$

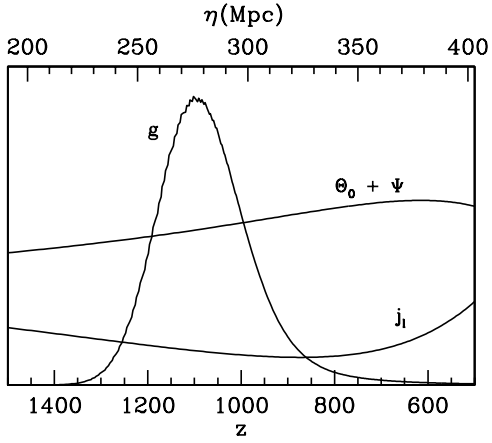


Figure 7.11: The three components of the integrand in the monopole term of equation [7.56]. The visibility function is sharply peaked, so changes rapidly compared with the monopole $\Theta_0 + \Psi$ and the Bessel function $j_l(k[\eta - \eta_0])$. Figure is for $l = 100, k = 0.013\text{h Mpc}^{-1}$, but the qualitative feature that the visibility function changes rapidly compared to the other terms in the integrand holds for all modes.

$$+ \quad 3\Theta_1(k,\eta_*)\Big(j_{l-1}\left[k(\eta_0-\eta_*)\right] - \frac{(l+1)j_l\left[k(\eta_0-\eta_*)\right]}{k(\eta_0-\eta_*)}\Big)$$

$$+ \quad \int_0^{\eta_0} d\eta\; e^{-\tau}\left[\dot{\Psi}(k,\eta) - \dot{\Phi}(k,\eta)\right]\; j_l\left[k(\eta_0-\eta)\right]. \tag{7.57}$$

Here I have used the spherical Bessel function identity:

$$\frac{d}{dx}j_l(x) = j_{l-1}(x) - \frac{l+1}{x}j_l(x) \tag{7.58}$$

and the fact that $v_B \simeq -3i\Theta_1$ at η_*. Equation 7.57 is the basis for semi-analytic calculations of C_l spectra which agree with the exact (numerical) solutions to within ten percent. From equation [7.57], we see that, to solve for the anisotropies today, we must know the monopole (Θ_0), dipole ($Theta_1$), and potential (Ψ) at the time of recombination. Further, there will be small but noticable corrections if the potentials are time dependent. These corrections, encoded in the first line of equation [7.57], are often called Integrated Sachs-Wolfe (ISW) terms.

The monopole term – the first in equation [7.57] – is precisely what we expected from the rough arguments of §7.1. In particular, the spherical Bessel function, $j_l[k(\eta_0 - \eta_*)]$, determines how much anisotropy on an angular scale l^{-1} is contributed by a plane wave with

wavenumber k. On very small angular scales,

$$\lim_{l\to\infty} j_l(x) = \frac{1}{l}\left(\frac{x}{l}\right)^{l-1/2}. \tag{7.59}$$

That is, $j_l(x)$ is extremely small for large l when $x < l$. In our case, this means that $\Theta_l(k,\eta_0)$ is very close to zero for $l > k\eta_0$. This makes sense physically. Returning to Figure 7.5, we see that very small angular scales will see little anisotropy from a perturbation with a large wavelength. The converse is also true: angular scales larger than $1/(k\eta_0)$ get little contribution from such a perturbation. To sum up, a perturbation with wavenumber k contributes predominantly on angular scales of order $l \sim k\eta_0$. One last comment about the monopole term: the final anisotropy today depends on not just Θ_0, but rather $\Theta_0 + \Psi$, again something we anticipated since photons must climb out of their potential wells to reach us today.

7.5.2 The C_l's

How is the observed anisotropy pattern today related to the rather abstract $\Theta_l(k,\eta_0)$? To answer this question, we must first describe the way in which the temperature field is characterized today and then relate this characterization to Θ_l.

Recall that in equation [3.32], we wrote the temperature field in the universe as

$$T(\vec{x},\hat{\gamma},\eta) = T^{(0)}(\eta)\left[1 + \Theta(\vec{x},\hat{\gamma},\eta)\right]. \tag{7.60}$$

While this field is defined at every point in space and time, we can observe it only here (at x_0) and now (at η_0)[§]. Our only handle on the anisotropies is their dependence on the direction of the incoming photons, $\hat{\gamma}$. So all the richness we observe comes from the changes in the temperature as the direction vector γ changes. Observers typically makes maps, wherein the temperature is reported at a number of incoming directions, or "spots on the sky." These spots are labeled not by the $\hat{\gamma}_x, \hat{\gamma}_y, \hat{\gamma}_z$ components of $\hat{\gamma}$, but rather by polar coordinates θ, ϕ. However, it is a simple matter to move back and forth between the 3D unit vector $\hat{\gamma}$ and polar coordinates[¶]. I'll stick with $\hat{\gamma}$ in the ensuing derivation.

We now expand the field in terms of spherical harmonics. That is, we write

$$\Theta(\vec{x},\hat{\gamma},\eta) = \sum_{l=1}^{\infty}\sum_{m=-l}^{l} a_{lm}(\vec{x},\eta)Y_{lm}(\hat{\gamma}). \tag{7.61}$$

The subscripts l, m are conjugate to the real space unit vector $\hat{\gamma}$, just as the variable $\vec{k}$ is conjugate to the Fourier transform variable $\vec{x}$. We are all familiar with Fourier transforms, so it is useful to think of the expansion in terms of spherical harmonics as a kind of generalized Fourier transform. While the complete set of eigenfunctions for the Fourier transform are

[§] We do make small excursions from this point in space-time. For example, satellites are not located on Earth and anisotropy measurements have been made over the last thirty years. These are completely insignificant on scales over which the temperature is varying, which are of order the Hubble time (or distance).

[¶] $\hat{\gamma}_z = \cos\theta$, $\hat{\gamma}_x = \sin\theta\cos\phi$, and $\hat{\gamma}_y = \sin\theta\sin\phi$.

$e^{i\vec{k}\cdot\vec{x}}$, here the complete set of eigenfunctions for expansion on the surface of a sphere are $Y_{lm}(\hat{\gamma})$. All of the information contained in the temperature field T is also contained in the space-time dependent amplitudes a_{lm}. As an example of this, consider an experiment which maps the full sky with an angular resolution of seven degrees. The full sky has $4\pi\text{radians}^2 = 41,000\text{degrees}^2$, so there are 840 pixels with area of $(7°)^2$. Thus, such an experiment would have 840 independent pieces of information. Were we to characterize this information with a_{lm}'s instead of temperatures in pixels, there would be some $l_{\max}$ above which there is no information. One way to determine this $l_{\max}$ is to set the total number of recoverable a_{lm}'s as $\sum_{l=0}^{l_{\max}}(2l+1) = (l_{\max}+1)^2 = 840$. So the information could be equally well characterized by specifying all the a_{lm}'s up to $l_{\max} = 28$. Incidentally, this is a fairly good caricature of the COBE experiment. They presented temperature data over many more pixels, but many of these pixels were overlapping. So, the independent information was contained in multipoles up to $l \sim 30$. Experiments currently under way or well along in the planning stage are capable of measuring the moments all the way up to $l \sim 3000$.

We want to relate the a_{lm}'s which are observable to the Θ_l we have been dealing with. To do this, we can use the orthogonality property of the spherical harmonics. The Y_{lm}'s are normalized so that

$$\int d\Omega Y_{lm}(\hat{\gamma})Y^*_{l'm'}(\hat{\gamma}) = \delta_{ll'}\delta_{mm'} \tag{7.62}$$

where Ω is the solid angle spanned by $\hat{\gamma}$. Therefore the expansion of Θ in terms of spherical harmonics, equation [7.61], can be inverted by multiplying both sides by $Y^*_{lm}(\hat{\gamma})$ and integrating:

$$a_{lm}(\vec{x},\eta) = \int \frac{d^3k}{(2\pi)^3} e^{i\vec{k}\cdot\vec{x}} \int d\Omega Y^*_{lm}(\hat{\gamma})\Theta(\vec{k},\hat{\gamma},\eta). \tag{7.63}$$

Here I have written the right hand side in terms of the Fourier transform ($\Theta(\vec{k})$ instead of $\Theta(\vec{x})$), since that is the quantity for which we obtained solutions.

As with the density perturbations, we cannot make predictions about any particular a_{lm}, just about the distribution from which they are drawn. Figure 7.12 illustrates this distribution. The mean value of all the a_{lm}'s is zero, but they will have some non-zero variance. The variance of the a_{lm}'s is called C_l. Thus,

$$< a_{lm} >= 0 \qquad ; \qquad < a_{lm}a^*_{l'm'} >= \delta_{ll'}\delta_{mm'}C_l. \tag{7.64}$$

It is very important to note that, for a given l, each a_{lm} has the same variance. For $l = 100$, say, all 201 $a_{100,m}$'s are drawn from the same distribution. When we measure these 201 coefficients, we are sampling the distribution. This much information will give us a good handle on the underlying variance of the distribution. On the other hand, if we measure the five components of the quadrupole ($l = 2$), we do not get very much information about the underlying variance, C_2. Thus, *there is a fundamental uncertainty in the knowledge we may get about the C_l's.* This uncertainty, which is most pronounced at low l, is called *cosmic variance*. Quantitatively, the uncertainty scales simply as the inverse of the square root of the number of possible samples, or

$$\left(\frac{\Delta C_l}{C_l}\right)_{\text{cosmic variance}} = \sqrt{\frac{2}{2l+1}}. \tag{7.65}$$

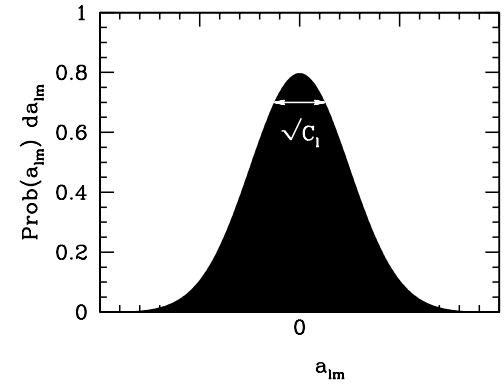


Figure 7.12: The distribution from which the a_{lm}'s are drawn. The distribution has expectation equal to zero and a width of $C_l^{1/2}$.

We can now obtain an expression for C_l in terms of $\Theta_l(k)$. First we square a_{lm} in equation [7.63] and take the expectation value of the distribution. For this we use

$$< \Theta(\vec{k},\hat{\gamma})\Theta(\vec{k}',\hat{\gamma}') >= (2\pi)^3\delta^3(\vec{k}-\vec{k}')\Theta(k,\hat{k}\cdot\hat{\gamma})\Theta(k,\hat{k}\cdot\hat{\gamma}'). \tag{7.66}$$

Then we expand $\Theta(k,\hat{k}\cdot\hat{\gamma}')$ in terms of spherical harmonics using the inverse of equation [3.104], $\Theta(k,\hat{k}\cdot\hat{\gamma}) = \sum_l(-i)^l(2l+1)P_l(\hat{k}\cdot\hat{\gamma})\Theta_l(k)$. This leaves

$$\begin{aligned} C_l &= \int \frac{d^3k}{(2\pi)^3}\sum_{l'l''}(-i)^{l'}(i)^{l''}(2l'+1)(2l''+1)\Theta_{l'}(k)\Theta_{l''}(k) \\ &\quad \times \int d\Omega P_{l'}(\hat{k}\cdot\hat{\gamma})Y_{lm}(\hat{\gamma})\int d\Omega' P_{l''}(\hat{k}\cdot\hat{\gamma}')Y^*_{lm}(\hat{\gamma}'). \end{aligned} \tag{7.67}$$

The two angular integrals here (Problem (10)) are identical. They are non-zero only if $l' = l$ and $l'' = l$, in which case they are equal to $4\pi Y_{lm}(\hat{k})/(2l+1)$ (or the complex conjugate). The angular part of the d^3k integral then becomes an integral over $|Y_{lm}|^2$, which is just equal to one, leaving

$$C_l = \frac{2}{\pi}\int_0^\infty dk\, k^2|\Theta_l(k)|^2. \tag{7.68}$$

So the mean square variance of the a_{lm}'s is the sum of the variances of Θ_l for all $k-$ modes. This is the function for which we have labored so hard to find solutions.

7.6 The Anisotropy Spectrum Today

7.6.1 Sachs-Wolfe Effect

Large angle anisotropies are not affected by any microphysics: at the time of recombination, the perturbations responsible for these anisotropies were on scales far larger than could be connected via causal processes. On these largest of scales, only the monopole contributes to the anisotropy. Equation [7.57] tells us that the relevant perturbations are $\Theta_0+\Psi$ evaluated at recombination. The large scale solution we found in equation [7.7] was that this combination is equal to $\Psi(\eta_*)/3$. In most cosmological models, recombination occurs far enough after matter/radiation equality that we can approximate the potential back then to be equal to the potential today modulo the growth factor, so

$$\Theta_0(\eta_*) + \Psi(\eta_*) \simeq \frac{1}{3D_1(a=1)}\Psi(\eta_0) = -\frac{1}{3D_1(a=1)}\Phi(\eta_0). \tag{7.69}$$

The last equality holds here because at very late times, there are no appreciable anisotropic stresses, and $\Phi = -\Psi$.

We may use equation [6.7] to express the potential Φ today in terms of the dark matter distribution, so that

$$\Theta_0(\eta_*) + \Psi(\eta_*) \simeq -\frac{\Omega_m H_0^2}{2k^2 D_1(a=1)}\delta_{\rm DM}(\eta_0). \tag{7.70}$$

This gives us what we need: an expression for the sum of $\Theta_0 + \Psi$ at recombination that we can plug into the monopole term in equation [7.57]. To get the anisotropy spectrum today, we then integrate as in equation [7.68], leaving

$$C_l^{\rm SW} \simeq \frac{\Omega_m^2 H_0^4}{2\pi D_1^2(a=1)} \int_0^\infty \frac{dk}{k^2} j_l^2\left[k(\eta_0-\eta_*)\right] P(k) \tag{7.71}$$

where the superscript denotes *Sachs-Wolfe*, in honor of the first people to compute the large angle anisotropy. The power spectrum is given by equation [6.9] with the transfer function set to one (since we're considering very large scales). Therefore,

$$C_l^{\rm SW} \simeq \pi H_0^{1-n}\delta_H^2 \int_0^\infty \frac{dk}{k^{2-n}} j_l^2\left[k(\eta_0-\eta_*)\right]. \tag{7.72}$$

The large scale anisotropies in the form of equation [7.72] can be computed analytically. First, we will use the fact that $\eta_* << \eta_0$ and define the dummy variable $x \equiv k\eta_0$. Then the spectrum can be rewritten as

$$C_l^{\rm SW} \simeq \pi(\eta_0 H_0)^{1-n}\delta_H^2 \int_0^\infty \frac{dx}{x^{2-n}} j_l^2(x). \tag{7.73}$$

The integral over the spherical Bessel functions can be analytically expressed (Gradshteyn & Ryzhik, 6.574.2) in terms of Gamma functions, leaving

$$C_l^{\rm SW} \simeq 2^{n-4}\pi^2(\eta_0 H_0)^{1-n}\delta_H^2 \frac{\Gamma\left(l+\frac{n}{2}-\frac{1}{2}\right)\Gamma(3-n)}{\Gamma\left(l+\frac{5}{2}-\frac{n}{2}\right)\Gamma^2\left(2-\frac{n}{2}\right)}. \tag{7.74}$$

If the spectrum is Harrison-Zel'dovich, $n = 1$, then

$$l(l+1)C_l^{\rm SW} = \frac{\pi}{2}\delta_H^2 \tag{7.75}$$

a constant. Indeed, this is the reason why workers in the field typically plot $l(l+1)C_l$: at low l, where the Sachs-Wolfe approximation is a good one, we expect a plateau.

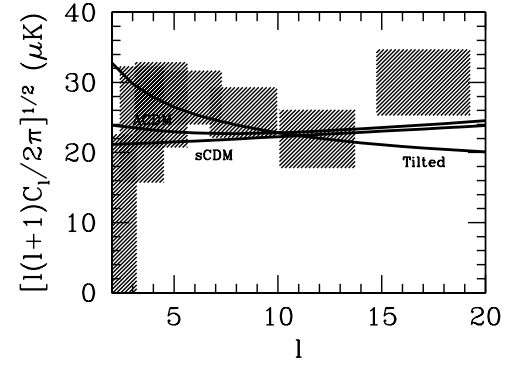


Figure 7.13: Large scale anisotropies. Hatched boxes show measurements by COBE satellite. Curves show the spectra for standard CDM, ΛCDM, (both with $n = 1$). The *tilted* model is identical to standard CDM, except $n = 0.5$. The late time Integrated Sachs-Wolfe effect enhances anisotropy on the largest scales in ΛCDM. Note that here, and in subsequent C_l figures, the RMS anisotropy plotted is proportional to $C_l^{1/2}$ and implicitly multiplied by the present background temperature, $T^{(0)}$.

Figure 7.13 shows the COBE measurements of the large angular scale anisotropies along with the Boltzmann solutions of three CDM models. For an $n = 1$ spectrum, the best fit values of δ_H from COBE are:

$$\begin{aligned} \delta_H &= 1.92\times10^{-5} \qquad \Omega_m = 1 \\ \delta_H &= 1.39\times10^{-5} \qquad \Omega_m = 0.3; \Omega_\Lambda = 0.7. \end{aligned} \tag{7.76}$$

Note that, even for $n = 1$, the true spectrum is *not* completely flat as suggested by equation [7.75]. The dipole at decoupling (neglected in equation [7.74]) contributes slightly. The Integrated Sachs-Wolfe effect also is not completely negligible, especially in the Λ model, wherein the potential starts to decay once the universe becomes vacuum dominated at late times.

Also shown in Figure 7.13 is a *tilted* model, one in which the primordial spectral index n is not equal to one. In such models, the anisotropy should scale as l^{n-1} compared with

the Harrison-Zel'dovich $n = 1$ spectrum. You can see this scaling from equation [7.74] or more directly from the integral in equation [7.73]. The integrand peaks at $k \sim l/\eta_0$, so very roughly every appearence of k there can be replaced by l/eta_0. The change in the integrand from k^{-1} to k^{n-2} therefore leads to a change in the spectrum that scales as l^{n-1}. As indicated in Figure 7.13, the COBE data has the greatest weight at $l \sim 10$, but covers a range of l spanning an order of magnitude. Extreme values of tilt are therefore ruled out by COBE. To get much better constraints on the tilt, though, measurements spanning a larger range of l are necessary.

7.6.2 Small Scales

The small scale anisotropy spectrum depends not only on the monopole, but also on the dipole and the ISW effect. Figure 7.14 shows all these contributions to the spectrum. Let's consider each in turn.

The monopole at decoupling $\Theta_0 + \Psi(k, \eta_*)$ freestreams to us today creating anisotropies on angular scales $l \sim k\eta_0$. This is what we expected back in Figure 7.5, showed to be true in equation [7.57], and can now see directly in the top panel of 7.14. There are two interesting features of the quantitative aspect of the freestreaming process. First, note that the "zeroes" in the monopole spectrum, here at 400, 650, and 970, are smoothed out because many modes contribute to anisotropy on a given angular scale. If only the $k = 400/\eta_0$ modes contributed to the anisotropy at $l = 400$, then C_{400} would really be zero. But many non-zero modes, with wavenumber greater than $400/\eta_0$, contribute. These change the zero to a trough in the C_l spectrum. The second feature of freestreaming worth noticing is that our initial estimate that the anisotropy shows up at $l = k\eta_0$ is not exactly right. There is a noticable shift in the top panel, suggesting that a given $k-$ mode contributes to slightly smaller l than we anticipated. This shift arises from the spherical Bessel function in equation [7.57]. As indicated in Figure 7.15, the peak in the Bessel function comes not when $l = k\eta_0$, but rather at slightly smaller values of l.

The dipole at decoupling is smaller than the monopole and out of phase with it. The middle panel in Figure 7.14 shows that adding in the dipole raises the overall anisotropy level, but particularly fills in the troughs. Without the dipole (in this model) the ratio of the height of the first peak (at $l \sim 200$) to the height of the first trough (at $L \sim 400$) is about $2.5 : 1$; the dipole lowers this ratio to $1.5 : 1$. This is a direct manifestation of the dipole and monopole being out of phase with one another.

The Integrated Sachs-Wolfe effect is also important if the potential changes after decoupling. To see which scales are affected by the ISW effect, consider the integral in equation [7.57]. Suppose the potential changes at time η_c, with all sub-horizon scales ($k\eta_c > 1$) being affected. The Bessel function peaks at $l \sim k(\eta_0 - \eta_c)$; so all angular scales $l > (\eta_0 - \eta_c)/\eta_c$ are affected. The largest effect is typically at the horizon.

The best, and most prevalent, example of the ISW effect is that due to residual radiation at decoupling. If the universe were purely matter dominated, there would be no such effect. But, the transition to pure matter domination is not abrupt, and even for $a_{\rm EQ} \sim 10^{-4}$, an ISW effect occurs right after decoupling. This early ISW effect is particularly important because it adds in phase with the monopole. To see this, integrate the last term in equa-

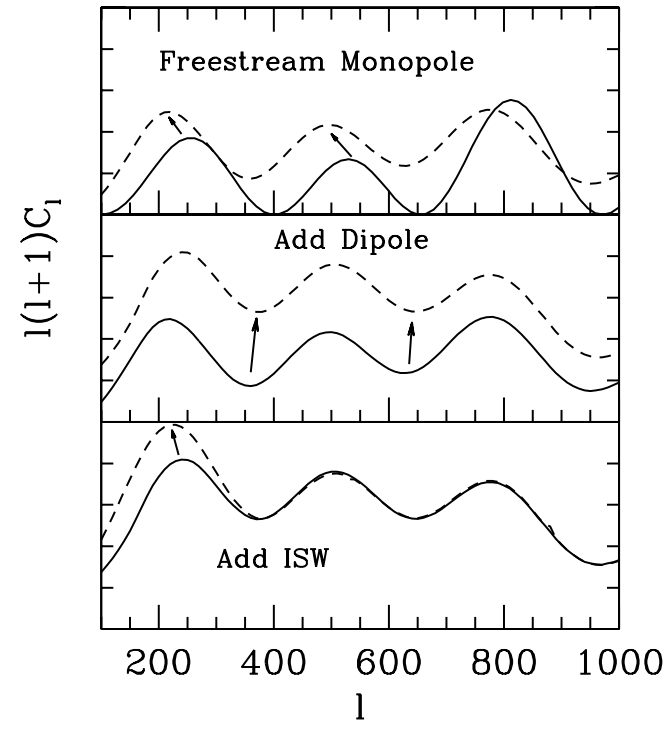


Figure 7.14: Small scale anisotropy. *Top panel:* The monopole at decoupling $\Theta_+\Psi(k = l/\eta_0, \eta_*)$ contains most of the structure of the final anisotropy spectrum. When free-streamed via the integral in equation [7.57], the spectrum shifts slightly to lower l. *Middle panel:* Accounting for the dipole raises the anisotropy spectrum. Since the dipole is out of phase with the monopole, the troughs become less pronounced. *Bottom panel:* The Integrated Sachs-Wolfe effect enhances the anisotropy on scales comparable to the horizon. In this case, the potentials changes near decoupling since the universe is not purely matter dominated then. Thus the first peak gets most of the excess power. Throughout, $h = 0.5, \Omega_B = 0.06$.

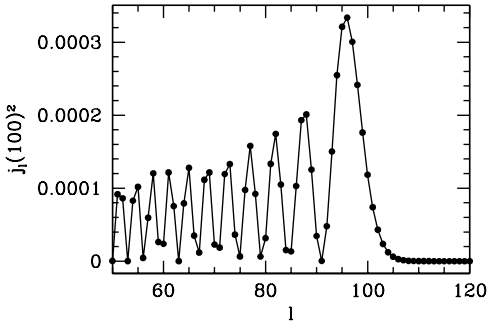


Figure 7.15: The spherical Bessel function, $j_l(100)$. Note that the peak occurs at $l \simeq 90$, slightly smaller than the argument.

tion [7.57] by parts. Then, the dominant contribution comes from $\eta \simeq \eta_*$, so the Bessel function can be evaluated there, leaving the trivial integral which gives

$$\Theta_l(k,\eta_0)^{\text{early ISW}} = \left[\Psi(k,\eta_0) - \Psi(k,\eta_*) - \Phi(k,\eta_0) + \Phi(k,\eta_*)\right] j_l\left[k(\eta_0 - \eta_*)\right]. \tag{7.77}$$

This adds exactly in phase with the monopole (which is proportional to the same Bessel function) so even though the magnitude of the effect on Θ_l is much smaller than is the dipole, the effect on the anisotropy spectrum is disproportionate. A thirty percent dipole leads to a ten percent shift in the C_l's, while a five percent ISW effect leads to the same ten percent shift in the C_l's. The bottom panel shows that the large scales, those with $l \sim \eta_0/\eta_*$, get a big boost from this early ISW effect.

7.7 Cosmological Parameters

The anisotropy spectrum depends on cosmological parameters. This fundamental realization initially caused great consternation ("We will never be able to measure any one parameter because there is too much degeneracy"). As more quantitative studies were carried out, the pendulum swung to the other side ("We will be able to disentangle the degeneracies and measure cosmological parameters to percent accuracy"). More recently, the community has settled into a state of cautious optimism. Indeed, just eight years after the initial discovery of large scale anisotropies by COBE, there were a host of experiments which together seemed to pin down one parameter (the total energy density) by measuring the location of the first peak. Several of these had measured a low second peak, which, as we will see, argues for high Ω_B.

We now have developed the theoretical tools needed to participate in the parameter determination discussion. In this section, we apply these tools to understand how the anisotropy spectrum varies as cosmological parameters vary.

One very important decision that must be made is which parameters will be allowed to vary. I will consider eight parameters:

- Total density, $\Omega = \Omega_m + \Omega_B + \Omega_\Lambda$
- Normalization, C_{10}
- Primordial Tilt, n
- Tensor modes, r
- Reionization, parametrized by τ back to decoupling
- Baryon density, $\Omega_B h^2$
- Matter density, $\Omega_m h^2$
- Cosmological constant, Λ

There are two aspects of this list worth stressing. The first is that obviously it does not include all possible cosmological parameters. Some favorites missing are a neutrino mass, the equation of state for dark energy w, and tensor tilt n_T. The second important point is that I have deliberately chosen very specific combinations of these parameters, e.g. $\Omega_B h^2$, not Ω_B and h separately. While there is good reason for this (e.g. the alternating peaks effect depends on $\Omega_B h^2$), it also is a source of confusion. A common complaint is that, within the context of a flat universe (the first parameter, the total density equal to critical), why should both the cosmological constant and the matter density be allowed to vary? Mustn't their sum equal one? It is true that $\Omega_m + \Omega_\Lambda$ must equal one in a flat universe. But that does not preclude us from varying both $\Omega_m h^2$ and Ω_Λ, since h can change while the sum of the two densities is one.

To harp on this point, consider two analysts. Analyst A works in the context of a flat universe and uses $\Omega_m h^2$ and Ω_Λ as her two free parameters. Analyst B also assumes the universe is flat, but takes h and Ω_Λ as his two parameters. When A raises Ω_Λ, the matter density ($\Omega_m h^2$) is kept fixed, so the epoch of equality is kept fixed. However, when analyst B raises his Ω_Λ, to keep the universe flat, he must lower Ω_m. He is therefore also lowering the matter density (since h is kept fixed), thereby moving $a_{\rm EQ}$ closer to today. That change in $a_{\rm EQ}$ will lead to an enhanced ISW effect, and therefore a larger first peak. Analyst A, who had the foresight to separate out this effect by choosing $\Omega_m h^2$ as one of her parameters, sees no such enhancement. And, indeed the enhancement is caused only indirectly by Ω_Λ: rather it is the direct result of a smaller $\Omega_m h^2$.

Let's now consider the effect of each parameter in turn.

7.7.1 Curvature

If the universe is not flat, then the simple picture of 7.5 is no longer accurate since photons do not travel in straight lines. In an open universe, photons starting out parallel to each other slowly diverge. Consider the implication of this divergence for anisotropies. Suppose

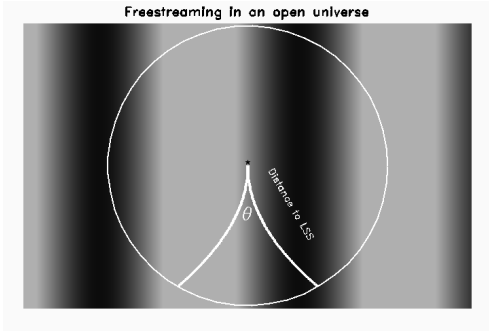


Figure 7.16: Photon trajectories in an open universe diverge. Perturbations at last scattering turn up on smaller scales in an open universe than they do in a flat universe.

the identical pattern of inhomogeneities was in place at decoupling in both a flat and open universe. As shown in Figure 7.16, the physical scale with maximal anisotropy (the first peak) gets projected onto a much smaller angular scale in an open universe. The peaks should therefore be shifted to higher l. As shown in figure 7.17, this is precisely what happens. In Problem (1), you will compute the magnitude of this effect, but it is already clear from figure 7.17 that the shift is dramatic. In fact, as we walk through shifts in all the other parameters of interest, we will see that the only way to get such a dramatic shift is by changing the curvature. Therefore, the positions of the peaks and troughs offer a robust test of flatness.

Figure 7.17 also shows data circa 2000. There is a clear rise up to a first peak at $l \sim 200$ and an almost equally clear fall past this first peak. When the data first started coming in (around 1998), a skeptic could plausibly claim that no one data set spanned the whole peak, and it is difficult to combine data sets. Within a year or two, though, this objection vanished as larger data sets such as TOCO, Boomerang, and Maxima all contained enough information by themselves to rule out an open universe. Of course, a truly flat universe is only one point in parameter space, the point in which the sum of the energy densities exactly equals the critical density, and no data will ever rule out all values except for this one point. Rather, the data now suggest that the total density is equal to the critical density with an error of about ten percent. The classic open universe favored by astronomers has thirty percent of the critical density, so is ruled out with very high confidence.

7.7.2 Parameters I

Figure 7.18 shows the results of varying four parameters. Before considering each in turn, it is important to state the obvious. All of these parameters change the spectrum in very similar ways. The shape of the spectrum varies hardly at all; rather these parameters simply

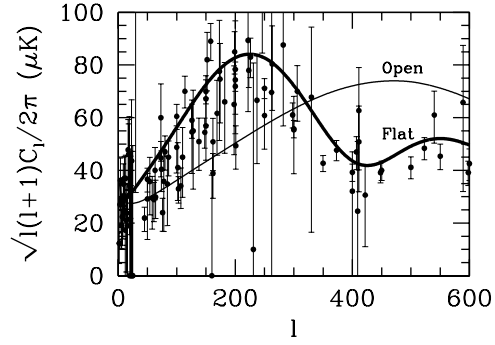


Figure 7.17: The anisotropy spectrum in flat vs. open universe. The pattern of peaks and troughs persists in the open universe but is shifted to smaller scales. The data clearly favor the flat case. Open curve has $\Omega_m = 0.3$ with no cosmological constant. Flat has $h = 0.7, \Omega_\Lambda = 0.7$.

move the spectrum up and down.

Normalization. The parameter C_{10} trivially moves the spectrum up or down. Note that, of the four parameters varied in Figure 7.18, it is the only one which can raise the amplitude of the spectrum.

Tilt. We have already considered the large angle effects of a tilted ($n \neq 1$) spectrum. If $n < 1$, then the small scale anisotropies are smaller than in the $n = 1$ model. Figure 7.18 shows that, as smaller and smaller scales are probed, the effect becomes more pronounced. So of the four parameters considered here, tilt has the most distinctive shape – it is not a simple up-down shift – and perhaps will be most easily extracted. Quantitatively, the spectrum scales as

$$\frac{C_l(n)}{C_l(n=1)} \simeq \left(\frac{l}{l_{\rm pivot}}\right)^{n-1} \tag{7.78}$$

where here $l_{\rm pivot} = 10$. Accounting for the fact that $\sqrt{C_l}$ is plotted in figure 7.18, we see from the point at $l = 1000$ that this scaling works extremely well.

Reionization.

The universe was almost certainly reionized at late times. We see this in the absorbtion spectra of high redshift quasars, where no evidence is seen of a uniform background of neutral hydrogen. Reionization brings the CMB back in contact with electrons. If enough scattering takes place, that is if the optical depth is high enough, equilibrium is restored; equivalently, primordial anisotropies are washed out.

There are several ways to see the effect of reionization quantitatively. One is to imagine a photon travelling in our direction with temperature $T + \Theta$, where Θ is the perturbation for which we have solved. If these photons hit a region with optical depth τ, only a fraction $e^{-\tau}$ will escape and continue on their way to us. In addition to these, we will also get a fraction

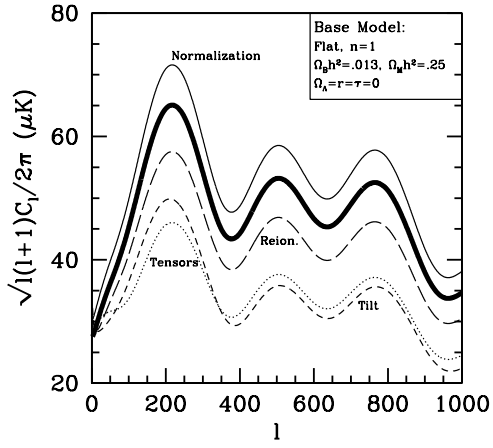


Figure 7.18: Changes in the anisotropy spectrum as C_{10}, τ, r, and n vary. Here $r(=1)$, $\tau(=.1)$, and $n(=.8)$ were chosen so the three separate curves could be distinguished. With different choices, these three variations (as well as the normalization) are close to degenerate.

$1-e^{-\tau}$ from the ionized region. All of these have the equilibrated temperature, T. So the temperature we see today is

$$(T+\Theta)e^{-\tau} + T\left(1-e^{-\tau}\right) = T + \Theta e^{-\tau}. \tag{7.79}$$

Subtracting from this the mean temperature T tells us that the anisotropy will be Θ, the primordial one set up at $z \simeq 1100$, multiplied by $e^{-\tau}$. Of course this argument can affect only those scales within the horizon at the time of reionization, so multipoles l larger than $\eta_0/\eta_{\rm reion}$ will be suppressed by $e^{-\tau}$; small l will be unaffected. This is clearly seen in figure 7.18, where the reionization curve agrees with the base model on large scales but is uniformly suppressed by ten percent (corresponding to $\tau = 0.1$) on small scales.

Tensors. We saw in Chapter 4 that once they enter the horizon, the amplitude of gravitational waves dies away. Therefore, gravity waves affect the anisotropy spectrum only on scales larger than the horizon at decoupling. Typically, this translates into angular scales $l < 100$. Indeed the *tensors* curve in figure 7.18 shows that tensors die out after $l > 100$. If tensor perturbations were produced during inflation, and if the total (scalar plus tensor) anisotropy spectrum is fit to the large scale (COBE) data, then the scalar amplitude is smaller than it would otherwise be. Therefore, on scales $l > 100$ where only scalars remain, the anisotropy spectrum is identical to the base model in figure 7.18, but with a lower amplitude.

7.7.3 Parameters II

The final variations we will consider are changes in the baryon density $\Omega_B h^2$, the matter density $\Omega_m h^2$, and the cosmological constant. As can be seen from figure 7.19, these changes lead to richer variations in the anisotropy spectrum; as such they are somewhat harder to understand (but easier to extract from the data!) than the parameters in the previous subsection.

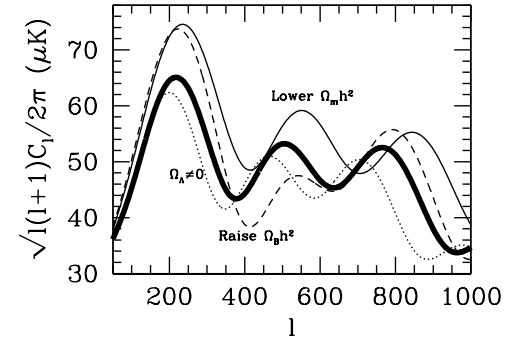


Figure 7.19: Changes in the anisotropy spectrum as baryon density, matter density, and cosmological constant vary. Same base model as figure 7.18.

Baryon Density. We have already touched on all the ways in which the anisotropy spectrum depends on the baryon density. The foremost, clearly visible in figure 7.19 is that odd peaks (first and third in the figure) go up when the baryon density is raised, while even peaks go down. This is a direct ramification of the lower frequency of oscillations due to the massive baryons. The second change is that an increased baryon density reduces the diffusion length. Therefore, a larger baryon density means damping moves to smaller angular scales, so the anisotropy spectrum at $l \sim 800-1000$ say is larger in a high $\Omega_B h^2$ model. One last feature of raising the baryon density is also evident from figure 7.19. The spectrum shifts to smaller angular scales. We should have anticipated this: the sound horizon defined in equation [7.24] gets smaller as the baryon density goes up. This means that the peaks, which occur at $kr_s \simeq n\pi$ move to *larger* k, and ultimately to larger l.

Cosmological Constant. If the matter density ($\Omega_m h^2$) is kept fixed, as we are doing, then introducing a cosmological constant impacts the anisotropy spectrum in two ways. Before discussing them, though, it should be clear that a cosmological constant is a late time phenomenon. It was not around at decoupling, and therefore could not have affected perturbations then. Therefore, the only possible effects of a cosmological constant are on freestreaming and on the largest angular scales just entering the horizon at recent times. The change due to freestreaming is evident in figure 7.19. The spectrum is shifted to larger angular scales. You will show in Problem (16) that this can be readily explained by comparing the

conformal times in a Λ universe and a matter dominated universe. Figure 7.19 also shows that the anisotropy spectrum is slightly lower on small scales in a Λ universe. This is a direct result of the large angle normalization. In a Λ universe, there is a late time ISW effect, which enhances the anisotropies on large angles. If we normalize on these scales, then the small scale anisotropy gets correspondingly smaller.

Matter Density. The conformal time today is inversely proportional to $[\Omega_m h^2]^{1/2}$. Therefore, lowering the matter density leads to a larger conformal time today, i.e. a larger distance to the last scattering surface. This shifts anisotropies to smaller angular scales, an effect which is readily apparent in figure 7.19. The corollary to this is that the damping scale $l_D = k_D \eta_0$ gets correspondingly larger (moves to smaller scales). Therefore, anisotropies on small scales are *less* damped if the matter density is low. Finally, if the matter density is low, the epoch of equality occurred closer to decoupling. This leads to both a larger ISW effect and a stronger driving force (so that Θ_0 is large) than in a purely matter dominated universe. All of these features serve to raise the anisotropies on small scales if the matter density is low.

Suggested Reading

The large scale Sachs-Wolfe effect was first predicted by Sachs & Wolfe (1967), just several years after the discovery of the CMB. Several groups initiated the study of anisotropies in the tightly coupled limit, Doroshkevich, Zel'dovich, and Sunyaev (1978), Atrio-Barandela & Doroshkevich (1994) and Jorgenson et al. (1994). The approach was perfected by Hu & Sugiyama (1995), which is the basis for the semi-analytic treatment of this chapter. It is well worth reading. Again, CMBFAST described in Seljak and Zaldarriaga (1996) is a crucial tool for fast, accurate numerical work. Diffusion damping is sometimes called *Silk* damping because of the Silk (1968) paper recognizing its importance. Two recent papers of interest are Zaldarriaga & Harari (1997) which discusses the effect of polarization on the damping scale (see Problem (8)) and Hu & White (1997) which, among other things, gives fits to the damping scale valid for a wide range of parameters.

The question of how the anisotropy spectrum depends on cosmological parameters has been explored in literally hundreds of papers over the past decade. Dick Bond, one of the pioneers in the field, gave a talk in 1992 (at a conference about the early COBE data) waving his hands through an invisible multi-dimensional parameter space, explaining that our goal now was to navigate through this space. Among the most important realizations were the dependence on curvature (Kamionkowski, Spergel, & Sugiyama 1994), the degeneracy of the height of the first peak (Bond et al. 1994), and breaking of this degeneracy by smaller scale information (Jungman et al. 1995). More recently, Hu et al. (2000) is a good reference.

I have given short shrift (or no shrift) to some important parameters. The effect of dark energy on the CMB has now been well-studied: first by Coble, Dodelson, & Frieman (1997) and then more generally by Caldwell, Dave, and Steinhardt (1998). Massive neutrinos affect the anisotropy spectrum at the 5-10% level (Ma & Bertschinger 1995 and Dodelson, Gates & Stebbins 1996). The anisotropies due to tensors became a hot topic after the COBE discovery. For a semi-analytic treatment and references to the dozens of papers relating the tensor anisotropy to parameters in the potential, see Turner, White & Lidsey (1993). While

the effect of reionization on the primary anisotropies generated before decoupling is well-understood, a hot topic now is *secondary* anisotropies, those generated after reionization. These will likely be probed by the post-Planck generation of experiments.

3K: The Cosmic Microwave Background (Partridge) is a good introduction to some of the experimental issues I have neglected in this book. The COBE discovery paper is Smoot et al. (1992) with the four year observations presented in Bennett et al. (1996). There were many good analyses papers written on the COBE data; I've relied on Bunn & White (1997), which is especially good for using COBE to normalize the matter power spectrum, and Tegmark (1997), from which the points in Figure 7.13 are taken.

Problems

7.1 What is the distance to the last scattering surface in an open universe? Express your answer as a function of Ω_m, and assume zero cosmological constant. What does this imply about the location of the first peak in the anisotropy spectrum?

7.2 Most of this book is devoted to understanding adiabatic perturbations with the initial conditions derived in 5. Another class of perturbations are *isocurvature* perturbations with initial conditions $\Theta_0 = \Psi = \Phi = 0$. Show that these initial conditions imply that

$$\Theta_0(\eta_*) + \Psi(\eta_*) = 2\Psi(\eta_*). \tag{7.80}$$

7.3 The equation for a damped harmonic oscillator is

$$m\ddot{x} + b\dot{x} + kx = 0. \tag{7.81}$$

Find the solutions to this equation if $k/m > (b/2m)^2$. What is the frequency of oscillations? How does this differ from the undamped ($b = 0$) solution? What is the other effect of non-zero b besides the change in frequency?

7.4 Determine $R(\eta_*)$ when $\Omega_B = 0.01, 0.06$. Plot the sound speed as a function of the scale factor for these two values of Ω_B.

7.5 Show that the sound horizon can be expressed in terms of the conformal time as

$$r_s(\eta) = \frac{2}{3k_{\rm EQ}}\sqrt{\frac{6}{R(\eta_{\rm EQ})}}\ln\left\{\frac{\sqrt{1+R}+\sqrt{R+R(\eta_{\rm EQ})}}{1+\sqrt{R(\eta_{\rm EQ})}}\right\}, \tag{7.82}$$

where $k_{\rm EQ}$ is given in equation [6.38].

7.6 Obtain the WKB solution to equation [7.20]. Write

$$\Theta_0 = Ae^{iB} \tag{7.83}$$

with A and B real. Show that the homogeneous part of equation [7.20] breaks up into two equations, coming from the real and imaginary part:

$$\text{Real} \quad : \quad -(\dot{B})^2 + \frac{\ddot{A}}{A} + \frac{\dot{R}}{1+R}\frac{\dot{A}}{A} + k^2 c_s^2 = 0 \tag{7.84}$$

$$\text{Imaginary} \quad : \quad 2\dot{B}\frac{\dot{A}}{A} + \ddot{B} + \frac{\dot{R}}{1+R}\dot{B} = 0. \tag{7.85}$$

Find B using the real part and the fact that B changes much more rapidly than A. Then, use the imaginary equation to determine A. Show that the homogeneous solutions obtained in this way differ from the simple oscillatory solutions of equation [7.23] by a factor of $(1+R)^{1/4}$.

7.7 Obtain a semi-analytic solution for $\Theta_0 + \Psi$ and Θ_1 at decoupling by carrying out the integrals in equation [7.26] and equation [7.27]. To do this you will need expressions for the gravitational potentials. Hu & Sugiyama (1995) provided the following convenient fits:

$$\begin{aligned}
\Phi(k,y) &= \bar{\Phi}(k,y)\left\{[1-T(k)]\exp[-0.11(ky/k_{\rm EQ})^{1.6}] + T(k)\right\} \\
\Psi(k,y) &= \bar{\Psi}(k,y)\left\{[1-T(k)]\exp[-0.097(ky/k_{\rm EQ})^{1.6}] + T(k)\right\}
\end{aligned}$$

where $y \equiv a/a_{\rm EQ}$, T is the BBKS transfer function and the large scale potentials are

$$\begin{aligned}
\bar{\Phi}(k,y) &= \frac{3}{4}\left(\frac{k_{\rm EQ}}{k}\right)^2 \frac{y+1}{y^2}\bar{\Delta}_T(y) \\
\bar{\Psi}(k,y) &= -\frac{3}{4}\left(\frac{k_{\rm EQ}}{k}\right)^2 \frac{y+1}{y^2}\left(\Delta_T(y) + 0.65 N_2/(1+y)\right).
\end{aligned} \tag{7.86}$$

Finally the two functions N_2 and Δ_T are

$$\begin{aligned}
N_2(y) &= -0.1\frac{20y+19}{3y+4}\frac{y^2}{y+1}\Phi_{ls} - \frac{8}{3}\frac{y}{3y+4} + \frac{8}{9}\ln[3y/4+1] \\
\Delta_T &= \left[1.16 - \frac{0.48y}{y+1}\right]\Phi_{ls}\frac{y^2}{y+1}.
\end{aligned} \tag{7.87}$$

Here Φ_{ls} is the large scale solution of equation [6.31].

7.8 Our treatment of diffusion damping neglected the effect of polarization. Go through the same expansion in $\dot{\tau}^{-1}$ that we carried out in §7.4 this time accounting for polarization. Show that this changes the factor of 8/9 in equation [7.41] to 16/15. This beautiful result was first obtained by Zaldarriaga and Harari (1996) when the first author was an undergraduate in Argentina!

7.9 Assume that all electrons associated with hydrogen stay ionized and set $R = 0$. Evaluate the damping scale, k_D, defined in equation [7.41]. Show that in this limit, the damping scale is given by equation [7.43] where

$$f_D(x) = 5\sqrt{1+1/x} - \frac{20}{3}(1+1/x)^{3/2} + \frac{8}{3}\left[\left(1+1/x\right)^{5/2} - 1/x^{5/2}\right]. \tag{7.88}$$

7.10 Show that

$$\int d\Omega Y_{lm}(\hat{\gamma}) P_l(\hat{\gamma}\cdot\hat{k}) = \frac{4\pi}{2l+1}. \tag{7.89}$$

7.11 A simple way to estimate the COBE normalization of δ_H is to fix C_{10}. From Figure 7.13, estimate C_{10}. Use this and the Sachs-Wolfe formula, equation [7.75], to estimate δ_H for a flat, matter dominated universe and for a Lambda model. Compare with the numbers given in equation [7.76].

7.12 Determine the effects of reionization using the Boltzmann equation. Neglect the gravitational potentials, the velocity, and Θ_0 in the Boltzmann equation for photons. Start with a spectrum $\Theta_l(\eta)$ and evolve till today. Show that the moments are indeed suppressed by $e^{-\tau}$.

7.13 Assume that decoupling took place instantaneously. Show that the solution for the lth moment due to tensor perturbations (equation [3.111]) is

$$\Theta_l^T = -\frac{1}{2}\int_{\eta_*}^{\eta_0} d\eta\ \dot{h}\ j_l[k(\eta_0 - \eta)]. \tag{7.90}$$

7.14 Using the decomposition for tensor modes given in equation [3.110], find the contribution to the C_l's from $\Theta_l^T(k)$. That is, show that the analogue of equation [7.68] for tensors is

$$\begin{aligned}
C_{l,i}^T &= \frac{(l-1)l(l+1)(l+2)}{\pi}\int_0^\infty dk\ k^2 \\
&\quad \times \left|\frac{\Theta_{l-2,i}^T}{(2l-1)(2l+1)} + 2\frac{\Theta_{l,i}^T}{(2l-1)(2l+3)} + \frac{\Theta_{l+2,i}^T}{(2l+1)(2l+3)}\right|^2,
\end{aligned} \tag{7.91}$$

where i denotes the two different components $+$ and $\times$.

7.15 Determine the spectrum of anisotropies due to gravity waves produced during inflation.

(a) Combine the results of the previous two problems, your solution to Problem (4.11), and the primordial amplitude of gravity waves in equation [5.103] to find the large angle C_l's due to inflation-produced gravity waves.

(b) Tensor anisotropies are often parametrized by

$$r \equiv \frac{C_2^T}{C_2^S} \tag{7.92}$$

where C_2^T is the variance of the quadrupole due to tensors and C^S is the same due to scalars. We already derived an expression for the scalar C_2 in equation [7.75]. Find C_2^T and compute r to first order in the slow roll parameter ϵ.

(c) The results of part (b) and equation [5.107] imply a *consistency relation* – a robust prediction of inflation – between the two observables n_T and r. What is the consistency relation?

7.16 Compute the conformal time today in a flat universe with $\Omega_\Lambda = 0.7$. (This cannot be done analytically.) Compare with the conformal time in a flat universe with no cosmological constant. By how much do you expect the anisotropy spectrum to shift in $l-$ space between the two models. Compare with figure 7.19.

7.17 Compute the conformal time today in a model with dark energy $\Omega_{\rm dark} = 0.7$ today with $w = -0.5$. Compare the expected shift in the anisotropy spectrum with the cosmological constant problem of the previous problem.

Solutions to Selected Problems

Chapter 1

Problem (6) An inverse wavelength is ν/c, so replacing ν everywhere in equation [1.12] by c/λ leads to

$$I_\nu = \frac{4\pi\hbar c}{\lambda^3}\frac{1}{\exp\{2\pi\hbar c/\lambda k_B T\}-1}. \tag{7.93}$$

This is energy per Hz; we want energy per cm^{-1}, we need to multiply by c, leaving

$$I_{1/\lambda} = \frac{4\pi\hbar c^2}{\lambda^3}\frac{1}{\exp\{2\pi\hbar c/\lambda k_B T\}-1}. \tag{7.94}$$

Plugging in numbers leads to

$$I_{1/\lambda} = 1.2\times 10^{-5}\mathrm{erg\ sec^{-1}\ cm^{-1}sr^{-1}}\left(\frac{\mathrm{cm}}{\lambda}\right)^3\frac{1}{\exp\{0.53\mathrm{cm}/\lambda\}-1}. \tag{7.95}$$

A quick check verifies that this agrees with figure 1.10.

To find the peak, differente I with respect to $1/\lambda$ and set equal to zero. This leaves:

$$\lambda = \frac{1}{3}\frac{(2\pi\hbar c/k_B T)}{1-\exp\{-2\pi\hbar c/\lambda k_B T\}}. \tag{7.96}$$

So $1/\lambda_{\mathrm{peak}}$ is $3/.53\mathrm{cm}^{-1}$. The exact coefficient, accounting for the exponential is 2.82, so $1/\lambda_{\mathrm{peak}} = 5.3\mathrm{cm}^{-1}$, exactly where it occurs in figure 1.10.

Chapter 2

Problem (5) Accumulating the various Γ's leads to

$$\frac{d^2x^i}{d\lambda^2} = -2\frac{\dot a}{a}\frac{dt}{d\lambda}\frac{dx^i}{d\lambda}. \tag{7.97}$$

Change to differentiation with respect to η using the facts that $dt/d\lambda = E$ and $d\eta/d\lambda = E/a$. Then the geodesic equation becomes

$$\frac{E}{a}\frac{d}{d\eta}\left(\frac{E}{a}\frac{dx^i}{d\eta}\right) = -2\frac{\dot a}{a}\frac{E^2}{a}\frac{dx^i}{d\eta}. \tag{7.98}$$

Since $E/a \propto a^{-2}$, when the derivative on the left acts on E/a, the resulting term (proportional to $dx^i/d\eta$ exactly cancels the term on the right, leaving the result of equation [2.64].

Chapter 3

Problem (1) First integrate equation [3.6] over all momentum. This gives

$$\frac{\partial n}{\partial t} + \frac{\partial(nv)}{\partial x} = 0 \tag{7.99}$$

the $\partial f/\partial p$ term vanishing after integrating by parts and noticing that $f = 0$ at $p = \pm\infty$ (there are no particles with infinite momentum). This is the continuity equation. To get the Euler equation, first multiply by p/m and then integrate over all momentum. This gives

$$\frac{\partial(nv)}{\partial t} + \frac{\partial}{\partial x}\int_{-\infty}^{\infty}\frac{dp}{2\pi}\frac{p^2}{m^2} + \frac{kx}{m}n = 0 \tag{7.100}$$

where the last term follows from an integration by parts. The integral over p^2 yields two terms, one a *bulk velocity* term, v^2, and the second a pressure term, P. Using the continuity equation reduces this to

$$\dot v + v\frac{\partial v}{\partial x} + \frac{1}{n}\frac{\partial P}{\partial x} + \frac{kx}{m} = 0. \tag{7.101}$$

Problem (3) From equation [2.54], the electron distribution function peaks at zero momentum, with a maximum value of

$$n_e\left(\frac{2\pi}{m_e T}\right)^{3/2}. \tag{7.102}$$

Divide equation [2.49] by the Thomson cross-section to get $n_e = 1.12\times 10^{-5}\Omega_B h^2\mathrm{cm}^{-3}$ today including both ionized and captured electrons. Taking the electron temperature to be equal to the photon temperature today gives $2\pi/m_e T = 2.04\times 10^{-11}\mathrm{cm}^2$. Putting back in the factors of a leads to

$$f_e^{\mathrm{MAX}} = 10^{-21}\Omega_B h^2 a^{-3/2}. \tag{7.103}$$

This expression holds only up to $T \le m_e$, corresponding to $a \simeq 4.6\times 10^{-10}$. So, as long as the temperature is well below the electron mass, f_e is very small.

Problem (5) The difference between the amplitude we used in the derivation in §3.3 and the more accurate one given in the problem is $2\pi\sigma_T m_e^2[3\cos(\hat p\cdot\hat p') - 1]$. The combination is square brackets is twice the second Legendre polynomial. Rewrite using the addition formula of spherical harmonics; then the difference becomes:

$$2\pi\sigma_T m_e^2\frac{8\pi}{5}\sum_{m=-2}^{2} Y_{2m}(\hat p)Y_{2m}^*(\hat p'). \tag{7.104}$$

This is the quantity we need to insert into the multiple integral in equation [3.47] in place of $\mathcal{M}^2$. When we do this, only the $m = 0$ term will contribute since all other $Y_{2m}(\hat p')$ have

an azimuthal dependence which integrates to zero. Therefore, the new collision term due to anisotropic Compton scattering is

$$\begin{aligned}\delta C[f(\vec{p})] &= \frac{\pi^2 n_e \sigma_T}{p} P_2(\mu) \int \frac{d^3p'}{(2\pi)^3 p'} P_2(\hat{\gamma}' \cdot \hat{k}) \\ &\times \left\{ \delta(p-p') + (\vec{p} - \vec{p}\,') \cdot \vec{v}\, \frac{\partial \delta(p-p')}{\partial p'} \right\} \{ f(\vec{p}\,') - f(\vec{p}) \}. \end{aligned} \tag{7.105}$$

where I have used the fact that $Y_{20} = -\sqrt{5}P_2/\sqrt{4\pi}$. The only term which survives the angular integral is the one proportional to $\delta(p-p')f(\vec{p}')$, leaving

$$\begin{aligned}\delta C[f(\vec{p})] &= -\frac{n_e \sigma_T}{2p} P_2(\mu) \int_0^\infty dp' p' \delta(p-p') p' \frac{\partial f^{(0)}}{\partial p'} \\ &\times \int_{-1}^{1} \frac{d\mu}{2} P_2(\mu) \Theta(\mu). \end{aligned} \tag{7.106}$$

The angular integral gives $-\Theta_2$. Then integrating over the $\delta-$ function yields

$$\delta C[f(\vec{p})] = +p \frac{\partial f^{(0)}}{\partial p} \frac{n_e \sigma_T}{2p} P_2(\mu) \Theta_2. \tag{7.107}$$

This adds a factor of $-P_2\Theta_2/2$ inside the square brackets of equation [3.52] and explains the corresponding factor in equation [3.96].

Chapter 4

Problem (6) (a) By definition,

$$\Gamma^i_{jk} = \frac{g^{ii'}}{2} \left[g_{i'j,k} + g_{i'k,j} - g_{jk,i'} \right]. \tag{7.108}$$

All derivatives here are spatial, and the only spatially varying part of the metric is the first order piece $\mathcal{H}$. Therefore, we can again use the zero order $g^{ii'} = \delta_{ii'}/a^2$, leaving equation [4.42].

(b) The product $\Gamma^\alpha_{\beta j}\Gamma^\beta_{i\alpha}$ vanishes when both indices α and β are zero (because $\Gamma^0_{0i} = 0$) and when both indices are spatial (because then each Christoffel symbol is first order). Therefore, this product is

$$\begin{aligned}\Gamma^\alpha_{\beta j}\Gamma^\beta_{i\alpha} &= \Gamma^0_{kj}\Gamma^k_{i0} + \Gamma^k_{0j}\Gamma^0_{ik} \\ &= \Gamma^0_{kj}\Gamma^k_{i0} + (i \leftrightarrow j). \end{aligned} \tag{7.109}$$

But

$$\begin{aligned}\Gamma^0_{kj}\Gamma^k_{i0} &= \frac{1}{2}\left(2\frac{da/dt}{a} g_{jk} + a^2 \mathcal{H}_{jk,0}\right)\left(\frac{da/dt}{a}\delta_{ik} + \frac{1}{2}\mathcal{H}_{ik,0}\right) \\ &= \left(\frac{da/dt}{a}\right)^2 g_{ij} + a\frac{da}{dt}\mathcal{H}_{ij,0}. \end{aligned} \tag{7.110}$$

We must remember to add back in the same set of terms with i and j interchanged. This just introduces a factor of two, so

$$\Gamma^\alpha_{\beta j}\Gamma^\beta_{i\alpha} = 2\left(\frac{da/dt}{a}\right)^2 g_{ij} + 2a\frac{da}{dt}\mathcal{H}_{ij,0}. \tag{7.111}$$

Chapter 5

Problem (5) There are 420 photons per cm^{-3} today; the Hubble volume is $(4\pi/3)[3000\mathrm{h}^{-1}\mathrm{Mpc}]^3$ $3.3 \times 10^{84}\mathrm{h}^{-3}\mathrm{cm}^3$. So the total number of photons is $1.4 \times 10^{87}\mathrm{h}^{-3}$. This number remains roughly constant throughout the matter and radiation eras since the number density scales as T^3, the physical volume as a^3, and the temperature as a^{-1}. So another problem of the classical cosmology is: Why is the entropy of the universe so large?

Inflation solves this problem. At first the solution seems obvious: inflation makes the scale factor grow exponentially fast, thereby increasing the product aT and hence the entropy. In fact, the solution is not quite that simple because during inflation, the exponential expansion is adiabatic: the temperature still falls as a^{-1}. So near the end of inflation the temperature has dropped rapidly enough so that if the entropy was initially of order unity, it remained of order unity.

The production of entropy actually takes place at the end of inflation during the reheating process. Even though the temperature at the end of inflation is extremely small, the energy density (which is almost completely in the scalar field) is not. When the energy in the scalar field transforms into radiation, the temperature of the radiation shoots up from its very low value of T to $\rho^{1/4} >> T$. Thus, the reheating process is responsible for the large entropy we see today. Another way to say this is to point out that inflation is a very ordered state: the universe supercools while the field is trapped in a false vacuum. The transition to the true vacuum is a transition to the very disordered state of equilibrium.

Chapter 6

Problem (3)

(c) To do the integral, introduce a new dummy variable $x \equiv \sqrt{1+y}$. Then equation [6.30] becomes

$$\Phi = \frac{3\Phi(0)}{2}\frac{\sqrt{1+y}}{y^3}\int_1^{\sqrt{1+y}} dx \frac{(x^2-1)^2(3x^2+1)}{x^2}. \tag{7.112}$$

Now integrate by parts using the fact that the integral of $1/x^2$ is equal to $-1/x$. The surface term is proportional to the numerator and so vanishes at the lower limit, when $x = 1$. Therefore,

$$\begin{aligned}\Phi &= \frac{3\Phi(0)}{2}\frac{\sqrt{1+y}}{y^3}\left[-\frac{y^2(4+3y)}{\sqrt{1+y}} + \int_1^{\sqrt{1+y}} dx(18x^4 - 20x^2 + 2)\right] \\ &= \frac{3\Phi(0)}{2}\frac{\sqrt{1+y}}{y^3}\left[-\frac{y^2(4+3y)}{\sqrt{1+y}} + (\frac{18}{5}x^5 - \frac{20}{3}x^3 + 2x)|_1^{\sqrt{1+y}}\right]. \end{aligned} \tag{7.113}$$

Evaluating the terms in parenthesis at the upper and lower limits leads to equation [6.31].

Problem (9)

$$\begin{aligned}\sigma_R^2 &= < \left[\int d^3x\delta(x)W_R(x)\right]^2 > \\ &= < \left[\frac{d^3k}{(2\pi)^3}\bar{\delta}(\vec{k})\bar{W}_R^*(\vec{k})\right]^2 > \end{aligned} \tag{7.114}$$

where $^-$ denotes Fourier transform, and I have used the fact that since $W_R(x)$ is real, $\bar{W}_R(\vec{k}) = \bar{W}_R^*(-\vec{k})$. Also I have evaluated δ_R at the origin; the angular brackets denote the average, now over all realizations of $\bar{\delta}(\vec{k})$. Squaring and using the fact that

$$< \delta(\vec{k})\delta(\vec{k}') >= (2\pi)^3\delta^3(\vec{k}+\vec{k}')P(k) \tag{7.115}$$

leads to

$$\sigma_R^2 = \int \frac{d^3k}{(2\pi)^3}P(k)\left|\bar{W}_R(\vec{k})\right|^2. \tag{7.116}$$

It remains only to compute the Fourier transform of the tophat window function,

$$\begin{aligned}\bar{W}_R(\vec{k}) &= \frac{1}{V_R}\int d^3xW_R(\vec{x})e^{-i\vec{k}\cdot\vec{x}} \\ &= \frac{2\pi}{V_R}\int_0^R dxx^2\int_{-1}^1 d\mu e^{ikx\mu}. \end{aligned} \tag{7.117}$$

Note that I have normalized the window function so that the integral over it is unity; hence the factor of $V_R = 4\pi R^3/3$. Carrying out the remaining angular and radial integrals leads to

$$\begin{aligned}\bar{W}_R(k) &= \frac{3}{kR^3}\int_0^R dxxsin(kx) \\ &= \frac{3}{k^3R^3}\left[-kR\cos(kR)+\sin(kR)\right]. \end{aligned} \tag{7.118}$$

By way of solving Problem (10), note that

$$\Delta^2(k) = \frac{4\pi}{(2\pi)^3}k^3P(k). \tag{7.119}$$

Chapter 7

Problem (3) Assume a solution of the form $x = e^{i\omega t}$. The damping equation then becomes a quadratic equation for ω:

$$\omega^2 - \frac{ib}{m}\omega - \frac{k}{m} = 0. \tag{7.120}$$

Solving with $k/m > \gamma^2 \equiv (b/2m)^2$, leads to

$$\omega = i\gamma \pm \omega_1. \tag{7.121}$$

The frequency is now $\omega_1 \equiv [k/m-\gamma^2]^{1/2}$, smaller than in the undamped case. The amplitude is also damped by $e^{-\gamma t}$.

Problem (10) Use the addition theorem of spherical harmonics to write

$$P_{l'}(\hat{\gamma}\cdot\hat{k}) = \frac{4\pi}{2l+1}\sum_{m'}Y_{l'm'}^*(\hat{\gamma})Y_{l'm'}(\hat{k}). \tag{7.122}$$

Then the angular integral becomes an integral over the product of two spherical harmonics, which – due to orthogonality – is equal to one if $l' = l$ and $m' = m$ and zero otherwise. This leads directly to the desired reult.

Problem (14) The generalization of equation [7.67] to tensors gives

$$C_l^T = \sum_{ll'}(-i)^{l'+l''}(2l'+1)(2l''+1)\int\frac{d^3k}{(2\pi)^3}\Theta_{l'}^T(k)\Theta_{l''}^{T,*}(k)I_{lml'}(k)I_{lml''}^*(k) \tag{7.123}$$

where I have defined

$$I_{lml'}(k) \equiv \sqrt{\frac{8\pi}{15}}\int d\Omega P_{l'}(\hat{k}\cdot\hat{\gamma})Y_{lm}(\Omega)\left[Y_{22}(\Omega)+Y_{2-2}(\Omega)\right]. \tag{7.124}$$

The factor of $[8\pi/15]^{1/2}[Y_{22}+Y_{2-2}]$ is the combination $\sin^2\theta\cos(2\phi)$ which appears in equation [3.110], so this expression is valid only for the + mode. However, the × mode gives exactly the same result.

The integral $I_{lml'}$ is not trivial. By rewriting the Legendre polynomial as $[4\pi/(2l'+1)]^{1/2}Y_{l'0}/i^{l'}$, we can turn $I_{ll'}$ into an integral over the product of three spherical harmonics. Such integrals are well-studied in quantum mechanics and can be expressed in terms of the Wigner 3-j symbols. By the way, my favorite reference for these things – especially useful for this integral – is *Quantum Mechanics* (Landau & Lifshitz), like all the other texts in their Course of Theoretical Physics a wonderful investment. The integral is then

$$I_{lml'} = \sqrt{\frac{32\pi^2}{15(2l'+1)}}\frac{1}{i^{l'}} < lm|Y_{22}+Y_{2-2}|l'0 > \tag{7.125}$$

which vanishes unless $m = 2$ or $m = -2$. When m takes on one of these two values, the matrix element is

$$< l2|Y_{22}+Y_{2-2}|l'0 >= i^{l'-l}\begin{pmatrix} l & 2 & l' \\ 0 & 0 & 0\end{pmatrix}\left[\frac{5(2l'+1)(2l+1)}{4\pi}\right]^{1/2}\begin{pmatrix} l & 2 & l' \\ -2 & 2 & 0\end{pmatrix}. \tag{7.126}$$

The first 3-j symbol here, the one with the bottom row all zero, vanishes unless the sum of the elements in the top row $l+l'+2$ is even. And of course l' cannot differ from l by more than 2 since the combination of Y_{22} $Y_{l'0}$ leads to angular momenta ranging from $l'-2$ to $l'+2$. So the only time the matrix element is non-zero is when $l' = l-2, l, l+2$. Using Table 9 in §106 of *Quantum Mechanics* leads to the final result:

$$I_{lml'} = \sqrt{\frac{8\pi}{3}}\sqrt{2l+1}\,i^{-l}\,\left(\delta_{m,2}+\delta_{m,-2}\right)\left[c_{-2}\delta_{l',l-2}+c_0\delta_{l',l}+c_2\delta_{l',l+2}\right] \tag{7.127}$$

where here $\delta_{m,2}$ (and all other δ's) is the Kronicker delta, equal to one if $m = 2$ and zero otherwise. The coefficients are

$$\begin{aligned} c_{-2} &= \frac{\sqrt{6}}{4}\,\frac{[(l-1)l(l+1)(l+2)]^{1/2}}{(2l-3)(2l-1)(2l+1)} \\ c_0 &= \frac{-2\sqrt{6}}{4}\,\frac{[(l-1)l(l+1)(l+2)]^{1/2}}{(2l-1)(2l+1)(2l+3)} \\ c_2 &= \frac{\sqrt{6}}{4}\,\frac{[(l-1)l(l+1)(l+2)]^{1/2}}{(2l+1)(2l+3)(2l+5)}. \end{aligned} \tag{7.128}$$

The result in equation [7.91] then follows.

Problem (15)

(a) On large scales, we can take the matter dominated solution for h, so

$$\Theta_{l,i} = \frac{-1}{2}\int_{\eta_*}^{\eta_0} d\eta\; j_l[k(\eta_0-\eta)]\frac{d}{d\eta}\left[\frac{3j_1(k\eta)}{k\eta}\right] P_h^{1/2}. \tag{7.129}$$

Here I have used the fact that the initial amplitude of the gravity waves is $P_h^{1/2}$ with the time dependence given in the square brackets. Plug this into equation [7.91] to get

$$\begin{aligned} C_l^T &= 2\frac{9(l-1)l(l+1)(l+2)}{4\pi}\int_0^\infty dk\; k^2 P_h(k) \\ \times &\quad \left|\int_0^{\eta_0} d(k\eta)\frac{j_2(k\eta)}{k\eta}\Big[\frac{j_{l-2}(k[\eta_0-\eta])}{(2l-1)(2l+1)} + 2\frac{j_l(k[\eta_0-\eta])}{(2l-1)(2l+3)} + \frac{j_{l+2}(k[\eta_0-\eta])}{(2l+1)(2l+3)}\Big]\right|^2 \end{aligned} \tag{7.130}$$

where I have set the lower limit on the time integral to zero since $\eta_* << \eta_0$. Also, I have used the identity $(j_1/x)' = -j_2/x$. The factor of two out in front comes from the sum over the $+$ and $\times$ components. Using equation [5.103] for P_h (in the slow roll approximation $\epsilon = 0$ and $\nu = 3/2$) and defining new integration variables $y \equiv k\eta_0$ and $x \equiv k\eta$ leads to

$$\begin{aligned} C_l^T &= 36\left(\frac{H_{\rm inf}}{m_{\rm Pl}}\right)^2 (l-1)l(l+1)(l+2)\int_0^\infty \frac{dy}{y} \\ &\times\left|\int_0^y dx\frac{j_2(x)}{x}\Big[\frac{j_{l-2}(y-x)}{(2l-1)(2l+1)} + 2\frac{j_l(y-x)}{(2l-1)(2l+3)} + \frac{j_{l+2}(y-x)}{(2l+1)(2l+3)}\Big]\right|^2 \end{aligned} \tag{7.131}$$

Here $H_{\rm inf}$ denotes the Hubble rate during inflation, or more precisely the Hubble rate when the modes in question crossed the horizon (when $k\eta = -1$ early on). This expression does well on the low multipoles. To get even better results stick in the transfer function of equation [4.75].

(b) For the $l = 2$ mode, the double integral in equation [7.131] is equal to 2.139×10^{-4}, so $C_2^T = 0.185(H/m_{\rm Pl})^2$. The scalar C_2 is equal to $\pi\delta_H^2/12$. Using equation [5.103] for δ_H leads to

$$r = 13.86\epsilon. \tag{7.132}$$

(c) Combining with equation [5.107], we expect

$$r = -6.93 n_T. \tag{7.133}$$

For many models, the inflationary parameter $\delta = -\epsilon$, so

$$n - 1 = -n_T. \tag{7.134}$$

Bibliography

[1] F. Atrio-Barandela and A. G. Doroshkevich, *Astrophysical Journal* **420**, 26 (1994)

[2] J. M. Bardeen, *Physical Review* **D22**, 1882 (1980)

[3] C. L.Bennett et al., *Astrophysical Journal* **464**, L1 (1996)

[4] J. Bernstein, *Kinetic Theory in the Expanding Universe*, (Cambridge University Press, Cambridge, 1988)

[5] G. R. Blumenthal, S. M. Faber, J. R. Primack, and M. J. Rees, *Nature* **311**, 57 (1984)

[6] J. R. Bond and A. Szalay, *Astrophysical Journal* **274**, 443 (1983)

[7] J. R. Bond et al., *Physical Review Letters* **72**, 13 (1994)

[8] E. F. Bunn and M. White, *Astrophysical Journal* **480**, 6 (1997)

[9] S. Burles, K. M. Nollett, and M. S. Turner, `astro-ph/9903300`

[10] R. R. Caldwell, R. Dave, and P. J. Steinhardt, *Physical Review Letters* **80**, 1582 (1998)

[11] K. A. Coble, S. Dodelson, and J. A. Frieman, *Physical Review* **D55**, 1851 (1997)

[12] R. Crittenden et al., *Physical Review Letters* **71**, 324 (1993)

[13] S. Dodelson and J. M. Jubas, *Astrophysical Journal* **439**, 503 (1995)

[14] S. Dodelson, E. I. Gates, and A. S. Stebbins, *Astrophysical Journal* **467**, 10 (1996)

[15] A. G. Doroshkevich, Ya. B. Zel'dovich, and R. A.Sunyaev, *Soviet Astron.* **22**, 523 (1978).

[16] G. Efstathiou in *Physics of the Early Universe,* ed. J. A. Peacock, A. F. Heavens, and A. T. Davies (Edinburgh University Press, Edinburgh, 1990)

[17] G. Gibbons, S.Hawking, and T. Vachaspati eds., *The Formation and Evolution of Cosmic Strings* (Cambridge University Press, Cambridge 1990)

[18] Gradshteyn and Ryzhik, *Table of Integrals, Series, and Products* (Academic, San Diego 1994)

[19] A. Guth, *Physical Review* **D23**, 347 (1981)

[20] A. Guth, *The Inflationary Universe* (Perseus, 1998)

[21] D. W. Hogg, `astro-ph/9905116`

[22] W. Hu and N. Sugiyama, *Astrophysical Journal* **444**, 489 (1995)

[23] W. Hu and N. Sugiyama, *Astrophysical Journal* **471**, 542 (1996)

[24] W. Hu, M. Fukugita, M. Zaldarriaga, and M. Tegmark, `astro-ph/0006436`

[25] W. Hu and M. White, *Astrophysical Journal* **479**, 568 (1997)

[26] H. E. Jorgenson, E. Kotok, P. Naselsky, and I. Novikov, *Astronomy and Astrophysics* **294**, 639 (1995)

[27] G. Jungman, M. Kamionkowski, A. Kosowsky, and D. N. Spergel, *Physical Review Letters* **76**, 1007 (1995)

[28] M. Kamionkowski, D. N. Spergel, and N. Sugiyama, *Astrophysical Journal* **426**, L57 (1994)

[29] H. Kodama and M. Sasaki, Prog. Theor. Phys. Suppl. **78**, 1 (1984)

[30] A. Kosowsky, *Annals of Physics* **246**, 49 (1996)

[31] L. D. Landau and E. M. Lifshitz, *Quantum Mechanics* (Pergamon, Oxford 1977)

[32] A. R. Liddle and D. H. Lyth, *Cosmological Inflation and Large Scale Structure* (Cambridge University Press, Cambridge, 2000)

[33] E. M. Lifshitz, *J. Phys. USSR* **10**, 116 (1946)

[34] A. Linde, *Inflation and Quantum Cosmology* (Academic Press, Boston 1990)

[35] C. -P. Ma and E. Bertschinger, *Astrophysical Journal* **455**, 7 (1995)

[36] C. Misner, K. S. Thorne, and J. A. Wheeler, *Gravitation* (Freeman, San Francisco 1973)

[37] V.F. Mukhanov, H.A. Feldman , and R.H. Brandenberger, *Physics Reports* **215**, 203 (1992)

[38] T. Padmanabhan, *Structure Formation in the Universe*, (Cambridge Universe, Cambridge 1993)

[39] R. B. Partridge, *3K: The Cosmic Microwave Background*, (Cambridge University, Cambridge 1995)

[40] J. A. Peacock, *Cosmological Physics*, (Cambridge University Press, Cambridge, 1999)

[41] J. A. Peacock and S. J. Dodds, *Monthly Notices of Royal Astronomical Society* **267**, 1020 (1994)

[42] P. .J .E. Peebles, *The Large-Scale Structure of the Universe*, (Princeton University Press, Princeton, 1980)

[43] P. J. E. Peebles, *Astrophysical Journal* **153**, 1 (1968)

[44] P. J. E. Peebles, *Astrophysical Journal* **263**, L1 (1982)

[45] P. J. E. Peebles and J. T. Yu, *Astrophysical Journal* **162**, 815 (1970)

[46] G. B. Rybicki and A. P. Lightman, *Radiative Processes in Astrophysics* (Wiley, New York, 1979)

[47] R. K. Sachs and A. M. Wolfe, *Astrophysical Journal* **147**, 73 (1967)

[48] B. F. Schutz, *A first course in general relativity* (Cambridge University Press, Cambrdige, 1990)

[49] S. Seager, D. Sasselov, and D. Scott, *Astrophysical Journal* **128**, 407 (2000)

[50] U. Seljak and M. Zaldarriaga, *Astrophysical Journal* **469**, 437 (1996)

[51] J. Silk, *Astrophysical Journal* **151**, 459 (1968)

[52] G. Smoot et al., *Astrophysical Journal* **396**, L1 (1992)

[53] M. Tegmark, *Astrophysical Journal* **46**, L35 (1996)

[54] K. S. Thorne, *Black Holes and Time Warps* (Norton, New York 1994)

[55] M. S. Turner, M. White, and J. E. Lidsey, *Physical Review* **D48**, 4613 (1993).

[56] A. Vilenkin and E. P. S. Shellard, *Cosmic Strings and Other Topological Defects* (Cambridge University Press, Cambridge 2000)

[57] R. M. Wald, *General Relativity* (University of Chicago, Chicago 1984)

[58] S. Weinberg, *Gravitation and Cosmology* (Wiley and Sons, New York 1972)

[59] S. Weinberg, *The First Three Minutes* (Perseus Books 1993)

[60] M. Zaldarriaga and D. Harari, *Physical Review* **D52**, 3276 (1995)