NOTES ON THE COSMIC MICROWAVE BACKGROUND

ROBERT CRITTENDEN

1. Introduction

The cosmic microwave background (CMB), discovered in 1965 by Penzias and Wilson, is thought to be relic photon radiation left over from when the universe was in thermal equilibrium and is one of the most convincing pieces of evidence that the universe experienced a hot big bang. With a present temperature of 2.73 K, its frequency spectrum is consistent with a black body spectrum to 50 parts in a million. Such a spectrum is hard to produce unless at one point the photons were in thermal equilibrium as predicted by the big bang theory.

What makes the CMB such a useful cosmological tool is the fact that, for the most part, it has not interacted since the universe was very young – only a few hundred thousand years old. The photons decoupled from the electrons and baryons when the universe cooled sufficiently to allow the formation of hydrogen, a process known as 'recombination.' Recombination occurred at approximately a redshift of $z_{rec} \sim 1100$, at a temperature of $T_{rec} \sim 0.3eV$. At this early time, the fluctuations which would eventually grow into galaxies and the large scale structures we observe today were much smaller and were to a good approximation evolving linearly. Thus the CMB provides a snapshot of the early universe when the physics was much easier to understand than it is presently.

The temperature of the CMB photons on the sky is remarkably uniform. The largest anisotropy is the dipole produced by our motion with respect to the rest frame of the background, $\Delta T = 3.372 \pm 0.007mK$ corresponding with a velocity of $v = \frac{\Delta T}{T_0} c = 360$ km/s. Beyond this, the temperature is uniform to one part in 100,000. For nearly three decades, searches for temperature anisotropies could only place upper bounds on the fluctuations. This changed dramatically in 1993 when the COBE DMR instrument discovered temperature fluctuations of order $30\mu K$ on scales of 7° and larger. Since then, temperature fluctuations have been seen at smaller angular scales by instruments based on the ground and mounted on balloons. This has been capped very dramatically by the recent full sky CMB maps produced by NASA's Wilkinson Microwave Anisotropy Probe (WMAP) in February 2003.

1.1. Quantifying the CMB temperature maps.

In order to compare with theories, it is necessary to quantify the CMB maps. This can be done either by giving a sky temperature map, or equivalently by a Fourier like expansion into orthogonal functions. The temperature fluctuations on the sphere can be expanded in terms of the spherical harmonic functions:

Date: Lent term, 2003.

$$\frac{\delta T(\Omega)}{T} = \sum_{\ell,m} a_{\ell m} Y_{\ell m}(\Omega).$$

Properties of the spherical harmonic functions are given in Appendix A. The orthogonality of the spherical harmonic functions leads to the inverse transform,

$$a_{\ell m} = \int \frac{\delta T(\mathbf{n})}{T} Y^*_{\ell m}(\mathbf{n}) d\Omega_{\mathbf{n}}.$$

The ℓ value corresponds roughly with the angular size of the temperature fluctuations, $\Delta\theta \sim \pi/\ell$. For each ℓ, there are $2\ell+1$ values of $m = -\ell, -\ell+1, \ldots \ell$.

The simplest statistic we can consider is the two-point correlation function, or its Fourier analog, the temperature power spectrum. The power spectrum of the fluctuations is defined as the expectation of the multipole moments: $\langle a_{\ell m} a^*_{\ell' m'} \rangle = C_\ell \delta_{\ell\ell'} \delta_{mm'}$, which follows simply from rotation invariance. The two point correlation function in real space is defined as

$$C(\theta) \equiv \left\langle \frac{\delta T(\mathbf{n})}{T} \frac{\delta T(\mathbf{n}')}{T} \right\rangle = \sum_{\ell,m} C_\ell Y_{\ell m}(\mathbf{n}) Y^*_{\ell m}(\mathbf{n}') = \frac{1}{4\pi} \sum_\ell (2\ell+1) C_\ell P_\ell(\cos\theta),$$

where θ is the angle between $\mathbf{n}$ and $\mathbf{n}'$. The variance is given by the correlation at zero separation, $C(0) = \frac{1}{4\pi}\sum_\ell (2\ell+1) C_\ell$.

If the fluctuations are Gaussianly distributed, the power spectrum contains all the information that the map contains. All higher moments are either zero (for an odd number) or are simply related to the power spectrum. Studies of the non-Gaussian properties of the CMB sky often focus on the lowest order moments, such as the three or four point correlation functions or their Fourier analogs, known as the *bispectrum* or *trispectrum* respectively. However, there are an unlimited numbers of ways that non-Gaussianity may appear. While most studies so far have concluded that the observations are consistent with Gaussianity, there could be other kinds of non-Gaussianity which have not yet been tested.

2. Perturbations

Throughout we will be assuming that the metric fluctuations are in the linear regime, and that we can ignore higher order interactions. This simplifies the analysis considerably, and in particular means that when expanded in Fourier space, the various modes do not couple to each other. The harmonic functions for the Fourier series are solutions to the Helmholtz equation. We will focus on flat cosmologies, where the harmonic functions are simply $e^{i\mathbf{k}\cdot\mathbf{x}}$. If the universe is not spatially flat, the harmonic functions become more complex.

2.1. Synchronous gauge.

Throughout this discussion, we will be working in synchronous gauge with a (-+++) signature, though occasionally we will discuss transforming into Newtonian gauge to make contact with some other treatments. Synchronous gauge is generally the gauge of choice for CMB codes due to the simplicity of the time coordinates. However, it is not a fully fixed gauge and there exist residual degrees of freedom, known as gauge modes, which have to be considered.

The exact form of the metric we will be considering is

$$ds^2 = a^2(\tau)(-d\tau^2 + (\delta_{ij} + h_{ij}) dx^i dx^j).$$

Derivatives with respect to comoving time will be denoted by dots. We will denote sums (0,1,2,3) with Greek indices and sums (1,2,3) with Latin indices.

In Fourier space, the tensor can be broken up into scalar, vector and tensor pieces (see handout / appendix B). Most often we will just be considering the scalar piece, which can be written as

$$h_{ij} = \frac{V}{(2\pi)^3}\int d^3\mathbf{k} e^{i\mathbf{k}\cdot\mathbf{x}}\left[\frac{1}{3}h\delta_{ij} + \left(\hat{k}_i\hat{k}_j - \frac{1}{3}\delta_{ij}\right)h_S\right].$$

Here, h is the trace of h_{ij}, h_S is the traceless scalar piece and $h_- = h - h_S$ measures the anisotropic stress.

2.2. **Einstein equations.** The Einstein equations for the scalar metric degrees of freedom can be written a number of ways. Here we will use the two equations:

$$\ddot{h} + \frac{\dot{a}}{a}\dot{h} + 8\pi G a^2 \sum_N (\delta\rho_N + 3\delta p_N) = 0$$

and

$$\frac{2}{3}ik\dot{h}_- - 16\pi G a^2 \sum_N (\rho_N + p_N)v_N = 0.$$

Here the sum is over all the various particle species. Since we are focusing on the scalar perturbations, the velocities can be written as $\mathbf{v} = v\hat{\mathbf{k}}$.

3. Perfect Fluids

We next need to understand how the fluctuations in the various types of matter evolve. This will allow us to follow their evolution from when the fluctuations were produced in the early universe until the present.

3.1. **Perfect fluids.** In the linear regime, many of the constituent elements in the universe can be treated as perfect fluids. For a perfect fluid, the pressure is isotropic and a function of the density alone, and its stress-energy tensor may be written as,

$$T_{\mu\nu} = (\rho + p)u_\mu u_\nu + p g_{\mu\nu},$$

where $u^\mu = \frac{dx^\mu}{(-ds^2)^{1/2}}$ is the four velocity of the fluid and $u^\mu u_\mu = -1$. This implies that

$$\begin{aligned} T^0_0 &= -(\bar{\rho} + \delta\rho) \\ T^i_j &= (\bar{p} + \delta p)\delta^i_j \\ T^0_i &= (\bar{\rho} + \bar{p})v_i. \end{aligned}$$

Here, the over bar denotes averaged density or pressure, while the $\delta\rho$ denotes differences from the mean values. We can further define the density contrast $\delta \equiv \delta\rho/\bar{\rho}$, as well as the average equation of state $w \equiv \bar{p}/\bar{\rho}$ and the sound speed $c_s^2 \equiv \delta p/\delta\rho$.

Conservation of stress-energy implies then that

$$\begin{aligned} \dot{\delta} + (1+w)(ikv + \dot{h}/2) &= 3\tfrac{\dot{a}}{a}(w - c_s^2)\delta \\ ik\dot{v} + \tfrac{\dot{a}}{a}(1-3w)ikv + \tfrac{\dot{w}}{1+w}ikv &= \tfrac{c_s^2}{1+w}k^2\delta. \end{aligned}$$

The first equation is the usual continuity equation, while the second is a version of the Euler equation. It is useful to consider these equations for some special cases: cold dark matter ($p = 0$) and relativistic fluids ($p = \rho/3$).

3.1.1. *Cold dark matter.* Most leading models for structure formation contain some amount of cold dark matter. The dark matter is needed to form the potential wells that the baryons fall into when they have decoupled from the photons. This is necessary since the original baryon fluctuations will be damped by diffusion effects (i.e. Silk damping.) In purely baryonic models, the amplitude of density fluctuations on small scales tends to be greatly suppressed relative to the CMB anisotropy.

The dark matter evolution is particularly easy to follow in synchronous gauge. The frame defined by the freely falling dark matter defines the synchronous coordinates, and as a result the dark matter has no velocity in this frame, ($\mathbf{v}_c = 0$). Fluctuations in the dark matter density result only from the stretching of space, $\dot{\delta}_c + \dot{h}/2 = 0$.

The same equations also effectively apply to the baryons after they decouple. However, they begin with non-zero velocities resulting from their interactions with the photons. These slowly decay away as the photons fall into the gravitational wells made by the dark matter. The velocity equation, $\dot{v} + \frac{\dot{a}}{a} v = 0$, implies that the velocities decay as a^{-1} .

3.1.2. *Relativistic fluids.* At early times, well before decoupling, the photons and baryons were tightly coupled and acted as a single relativistic fluid. This fluid has a pressure due to the photons ($p = \rho_\gamma/3$) and a density which is the sum of the photon and baryon density ($\rho = \rho_\gamma + \rho_b$). Before matter domination, the baryon density can be ignored, so we find

$$\begin{aligned} \dot{\delta}_\gamma + \tfrac{4}{3}(ikv_\gamma + \dot{h}/2) &= 0 \\ ik\dot{v}_\gamma &= \tfrac{1}{4}k^2\delta_\gamma. \end{aligned}$$

There is a competition between the gravitational potential term, which tends to make the fluctuations collapse, and the pressure term, which would cause the fluctuations to oscillate acoustically. On very large scales, the gravitational term dominates, while on smaller scales the pressure dominates. The scale where they are roughly equal is known as the *Jeans scale,* and it gives an idea of the smallest objects which could have collapsed. For the baryons, this scale drops considerably at recombination when the pressure due to photons nearly disappears, and the Jeans mass drops from supercluster scales to globular star cluster scales.

3.2. **Superhorizon evolution - growing and gauge modes.** In general, if we consider perturbations in a universe dominated by a isotropic single fluid, we have four degrees of freedom: h, $\dot{h}$, δ, v. We can solve to find four separate solutions, two gauge and two physical, known as growing and decaying modes (Press & Vishniac, 1980). Outside the horizon, the gauge modes correspond to a constant shift in h, and one which is proportional to t^{-1}. We can solve for the behavior of the growing solutions easily outside the horizon if we set the velocities to zero by hand (though note that this will eliminate the decaying mode solution, but will leave one gauge mode.)

3.2.1. *Radiation domination.* Outside the horizon (and ignoring the fluid velocity), we can combine the metric evolution and the continuity equation to find a single second order equation for the relativistic fluid density

$$\ddot{\delta}_\gamma + \frac{\dot{a}}{a}\dot{\delta}_\gamma = \frac{32\pi}{3} G a^2 \bar{\rho}_\gamma \delta_\gamma = 4\left(\frac{\dot{a}}{a}\right)^2 \delta_\gamma$$

where the last follows from the definition of the Hubble parameter. Given that the scale factor grows as $a \propto \tau \propto t^{1/2}$ in the radiation dominated era, its easy to show this has the solutions $\delta \propto \tau^2$, $\tau^{-2} \propto t$, t^{-1}. The first solution is the growing mode, while the second corresponds to the gauge mode. A more detailed solution shows that there is also a decaying mode $\delta \propto \tau$ outside the horizon. The growing and decaying modes match onto the two phases of oscillations of the fluid once the modes enter the horizon.

3.2.2. *Matter domination.* A similar treatment of the fluctuations in the matter regime lead to an equation for the CDM density contrast

$$\ddot{\delta}_c + \frac{\dot{a}}{a}\dot{\delta}_c = 4\pi G a^2 \bar{\rho}_c \delta_c = \frac{3}{2}\left(\frac{\dot{a}}{a}\right)^2 \delta_c.$$

With the matter dominated scaling $a \propto \tau^2 \propto t^{2/3}$,we can find the solutions $\delta \propto \tau^2$, $\tau^{-3} \propto t^{2/3}$, t^{-1}. These again correspond to the growing and gauge modes outside the horizon, while the decaying mode can be shown to be $\delta_c \propto \tau^{-1} \propto t^{-1/3}$. Note that the equation above applies even after the perturbations enter the horizon. However the scaling of the gauge and decaying modes swap behaviours.

This equation also has very important consequences for the matter power spectrum. Modes outside the horizon (or those inside after matter domination) grow as τ^2, which corresponds to a gravitational potential which is constant in time. However, the righthand side of the equation is only large when the dark matter density is dominant. If the perturbations enter during the radiation dominated regime, their growth slows. This leads to a turnover in the matter power spectrum (or matter transfer function). The power spectrum ends up with a peak on scales which entered the horizon at the radiation-matter transition. Thus the matter power spectrum becomes a very good probe of the matter-radiation ratio in the universe.

A similar effect happens if the universe becomes dominated by a cosmological constant which is spatially smooth. Again, the growth of the fluctuations slows, affecting the power spectrum near the horizon scale. In doing so, it causes the gravitational potential to decay, which can be observed by the integrated Sachs-Wolfe effect (see below.)

4. Collisionless Boltzmann Equation

We introduce the Boltzmann equation in stages, first considering the case where there are no collisions and examining its consequences for the photon distribution function, in particular how it leads to the Sachs-Wolfe equation. In future sections, we consider the effects of introducing collisions and how they can lead to damping effects. Finally, we consider the full anisotropic Boltzmann equation and its implications for CMB polarization.

4.1. **The relativistic Boltzmann equation.** For many kinds of particles, the perfect fluid approximation is not sufficient. This is particularly true for relativistic species such as neutrinos and photons after last scattering, which have a large mean free path length. The Boltzmann equation keeps track of the number of particles in a particular element of phase space, the number with a particular position and momentum: $dN = f(\mathbf{x}, \mathbf{p}, \tau) d^3\mathbf{x}\, d^3\mathbf{p}$. Here, f is the phase space distribution function. In the context of CMB fluctuations, we want to know the number of

photons coming from a given direction, which is precisely the present distribution function.

We begin by considering simple one component particles with some mass . Our treatment can easily be extended to massless particles by taking the limit as $m \to 0$. By Louiville's theorem, if the Lagrangian includes all the relevant physics, then the distribution function for a given component of phase space is invariant so that $df = 0$. Since we are considering a non-trivial metric, it is essential to distinguish between the canonical momentum, $\widetilde{\mathbf{p}}$, which is intrinsic to Louiville's theorem and the 'proper' momentum, $\mathbf{p}$. The proper momentum is defined in a frame which is locally Minkowski ($p_i = p^i$). The canonical momentum is given by $\tilde{p}_i = mu_i$ is related to the proper momentum by

$$\tilde{p}_i = a(\delta_{ij} + \frac{1}{2}h_{ij})p^j.$$

The proper momentum can be written as $p_i = aq\hat{n}_i$, where $\hat{n}$ is a 3-d unit norm vector.

We can write the Boltzmann equation then in terms of the position and the new momentum parameterization, q and $\hat{n}$. Louiville's theorem implies

$$\frac{df}{d\tau} = \frac{\partial f}{\partial \tau} + \frac{dx^i}{d\tau}\frac{\partial f}{\partial x^i} + \frac{dq}{d\tau}\frac{\partial f}{\partial q} + \frac{d\hat{n}_i}{d\tau}\frac{\partial f}{\partial \hat{n}_i} = 0.$$

This can be solved perturbatively. At zeroth order, the solution should be homogeneous and isotropic. Assuming an initially thermal spectrum, we find that

$$f_0(q) = g_S(e^{-q_0/T_0} \pm 1)^{-1}.$$

Here, g_S is the number of degrees of freedom, T_0 is the present CMB temperature and $q_0 = (q^2 + a^2m^2)^{1/2}$ is the energy measured by a comoving observer. The sign depends on whether the particles are fermions or bosons.

We find the first order solution by expanding in orders $f = f_0(q) + f_1(\tau, q, \hat{\mathbf{n}}, \mathbf{x}) + ...$ and then solving the Boltzmann equation keeping only terms to first order. It is straight froward to show that to zeroth order

$$\frac{dx^i}{d\tau} = \frac{\widetilde{p}^i}{\widetilde{p}^0} \simeq \frac{p^i}{p^0} = \frac{q\hat{n}_i}{q_0}.$$

Since $\frac{\partial f}{\partial q}$ is zeroth order, we can keep $\frac{dq}{d\tau}$ to first order, and the geodesic equation leads to

$$\frac{dq}{d\tau} = -\frac{1}{2}q\dot{h}_{ij}\hat{n}_i\hat{n}_j.$$

In the final term of the Boltzmann equation, $\frac{d\hat{n}_i}{d\tau}$ and $\frac{\partial f}{\partial \hat{n}_i}$ are both first order, so their product can be ignored in the first order equation. Thus the first order Boltzmann equation becomes (for a given Fourier mode, in which the dependence on position becomes a dependence on wavenumber)

$$\frac{\partial f_1}{\partial \tau} + i\mathbf{k}\cdot\hat{\mathbf{n}}\frac{q}{q_0}f_1 - \frac{1}{2}\frac{\partial f_0}{\partial q}q\dot{h}_{ij}\hat{n}_i\hat{n}_j = 0.$$

This is appropriate for either neutrinos or photons, where we can take the massless limit and $q/q_0 = 1$.

Often the Boltzmann equation is expressed in terms of so-called *brightness functions,* where the q dependence of the distribution function has been integrated out,

and its expressed as the ration of first order fluctuations to the mean density:

$$\Delta(\mathbf{k},\mu,\tau) \equiv \frac{\int q^2 dq\, q f_1}{\int q^2 dq\, q f_0} = -4 f_1 \left(q \frac{\partial f_0}{\partial q} \right)^{-1} = 4 f_1 \left(T_0 \frac{\partial f_0}{\partial T_0} \right).$$

Here, $\mu \equiv \hat{\mathbf{k}} \cdot \hat{\mathbf{n}} = \cos\theta$. In terms of the brightness function, the Boltzmann equation is

$$\dot{\Delta} + ik\mu\Delta = -2\dot{h}_{ij}\hat{n}_i\hat{n}_j.$$

The brightness function basically is the ratio of the density of particles *moving in a given direction* to the mean density. Since the density is proportional to T^4, the temperature fluctuation is given by $\frac{\delta T}{T} = \Delta/4$. For a perfect fluid, the pressure is isotropic in its rest frame. In this case, the brightness function is particularly simple, $\Delta = \delta + 4\mu v$.

It is sometimes useful to write the angular dependence of the brightness function in a Legendre series with the argument μ, that is

$$\Delta(\mathbf{k},\tau,\mu) = \sum_\ell (2\ell+1)\Delta_\ell(\mathbf{k},\tau) P_\ell(\mu),$$

where the moments are given by $\Delta_\ell = \frac{1}{4\pi}\int d\Omega \Delta P_\ell(\mu)$. This is useful because the gradient term in the Boltzmann equation couples only Δ_ℓ modes to the $\Delta_{\ell-1}$ and $\Delta_{\ell+1}$ modes. If early on, only the lowest modes are excited, then the evolution of the higher modes are excited slowly. For example, in a perfect fluid only the lowest two moments are non-zero.

4.2. **SW equation.** It is straightforward to solve for the evolution of the brightness function. Multiplying the Boltzmann equation by $e^{ik\mu\tau}$ and integrating over time, one finds

$$\Delta(\tau_f) = \Delta(\tau_i) e^{-ik\mu(\tau_f-\tau_i)} - 2\int_{\tau_i}^{\tau_f} d\tau e^{-ik\mu(\tau_f-\tau)} \dot{h}_{ij}\hat{n}_i\hat{n}_j.$$

Consider the case of instantaneously decoupling at the last scattering surface, τ_{rec}. Prior to decoupling, the fluid is assumed to be a perfect fluid. The solution at the present becomes

$$\Delta(\tau_f) = (\delta + 4\mu v)\, e^{-ik\mu(\tau_f-\tau_{rec})} - 2\int_{\tau_{rec}}^{\tau_f} d\tau e^{-ik\mu(\tau_f-\tau)} \dot{h}_{ij}\hat{n}_i\hat{n}_j.$$

This, divided by a factor of four to convert into temperature fluctuations, is known as the Sachs-Wolfe equation. The physical interpretation of each term is discussed below.

4.3. **Origin of the Primary Anisotropies.** Recombination of electrons and protons to form hydrogen occurs when the photon-baryon fluid reaches a fixed temperature of order a third of an eV. This then occurs at different times in different regions; occurring first in underdense regions and later in overdensities. However, as discussed above, we can consider a fixed time, τ_{rec}, immediately after all the regions recombined as the *last scattering surface.*

There are three main contributions to the temperature anisotropies coming from the last scattering surface: the intrinsic density fluctuation, the Doppler effect and

the redshifting of the photon along the line of sight. In synchronous gauge, these can be written as

$$\frac{\delta T}{T} = \frac{1}{4}\delta_\gamma + \mathbf{v}_r \cdot \mathbf{n} - \frac{1}{2}\int_{\tau_{rec}}^{\tau_f} \dot{h}_{ij} n_i n_j d\tau,$$

where the dots are derivatives with respect to co-moving time and $\mathbf{n}$ is a unit vector along the line of sight. This is known as the Sachs-Wolfe equation. (Note that in the picture where recombination occurs at fixed temperature, there are no density fluctuations, but this term is replaced by a $\frac{\delta a}{a}$ factor since photons from different regions can have more or less time to redshift.)

The contribution from redshifting along the photon path can be understood in a fairly simple way. Consider a small segment of the path that a photon transverses in time $\delta\tau$. This segment has a length given by $\delta L = (g_{ij}dx^i dx^j)^{1/2} = (a^2(\delta_{ij} + h_{ij})n_i n_j)^{1/2}\delta\tau \simeq a(1 + \frac{1}{2}h_{ij}n_i n_j)\delta\tau$. The shift in the photon temperature over this distance is given by the fractional change of the segment length in the period $\delta\tau$,

$$\frac{\delta T}{T}|_{\delta L} = -\frac{1}{\delta L}\frac{\partial(\delta L)}{\partial\tau}\delta\tau = -\left[\frac{\dot{a}}{a} + \frac{1}{2}\dot{h}_{ij}n_i n_j\right]\delta\tau.$$

The first term corresponds to the usual homogeneous redshifting of photons and the second, integrated over the whole photon path, gives the term in Sachs-Wolfe expression.

If we focus on the scalar piece of the metric, there are two degrees of freedom and the metric fluctuation is parameterized in synchronous gauge as

$$h_{ij} = \frac{1}{3}h\delta_{ij} + (\widehat{k}_i\widehat{k}_j - \frac{1}{3}\delta_{ij})h^S.$$

Thus, in Fourier space the path dependent contribution to the temperature fluctuation is

$$\frac{\delta T}{T} = -\frac{1}{2}\int d\tau \sum_{\mathbf{k}} \left(\frac{1}{3}\dot{h}^- + (\widehat{\mathbf{k}}\cdot\mathbf{n})^2 h^S\right) e^{i\mathbf{k}\cdot\mathbf{n}\tau}$$

4.3.1. *Newtonian gauge version.* We can now use this expression to transform the Sachs-Wolfe equation into the Newtonian gauge ($h_{00} = 2\Psi$ and $h_{ij} = 2\Phi\delta_{ij}$). The gauge transformations can be shown to be

$$\begin{aligned}
\Psi &= \frac{1}{2k^2}\left[\ddot{h}^S + \frac{\dot{a}}{a}\dot{h}^S\right] \\
\Phi &= -\frac{1}{6}h^- - \frac{1}{2k^2}\frac{\dot{a}}{a}\dot{h}^S \\
\delta &= \delta_{syn} + \frac{3}{2k^2}\frac{\dot{a}}{a}(1+w)\dot{h}^S \\
\mathbf{v} &= \mathbf{v}_{syn} + \frac{i\mathbf{k}}{2k^2}\dot{h}^S
\end{aligned}$$

where $w \equiv p/\rho = 1/3$ is the equation of state for the fluid in question. Using these relations, the Sachs-Wolfe equation in Newtonian gauge can be shown to be

$$\frac{\delta T}{T} = \frac{1}{4}\delta_\gamma - \Psi + \mathbf{v}_r \cdot \mathbf{n} + \int_{\tau_{rec}}^{\tau_f} (\dot{\Phi} + \dot{\Psi})d\tau.$$

This is similar to the synchronous gauge expression, except the intrinsic piece is modified to include the initial gravitational potential. The final term, which is the contribution along the path of the photon, is known as the *integrated* Sachs-Wolfe contribution. Note that in the absence of anisotropic stress, the Einstein equations imply $\Psi = \Phi$.

4.4. **Neutrinos.** The neutrinos freestream as soon as they enter the horizon. Their density can sometimes be modeled by treating them as a viscous fluid. A more exact treatment requires following the full neutrino phase space density, as with the photons after decoupling.

5. Photon scattering history

Its useful to quantify how often the CMB photons scatter. Here we will focus on Thomson scattering from free electrons, but Raleigh scattering from hydrogen atoms can also be important for some frequencies.

5.1. **Optical depth.** The fundamental quantity to consider is the probability of scattering in a time dt, which is given by $\sigma_T n_e dt$, where σ_T is the Thomson cross section and n_e is the density of free electrons. The integrated probability over some time interval is known as the *optical depth,* $\kappa = \int dt \sigma_T n_e$. (Note, this is also denoted as τ in the literature.) This is related to the mean free time for scattering, which is $t_c = 1/\sigma_T n_e$.

Next consider the probability that a photon has traveled freely from some early time to the present without scattering, $\mathcal{P}_{\bar{s}}(t)$. The probability at the photon traveled from a slightly earlier time, $t - dt$, is given by $\mathcal{P}_{\bar{s}}(t - dt) = \mathcal{P}_{\bar{s}}(t)(1 - \sigma_T n_e dt)$, where the second factor is just the probability that the photon traveled the interval dt without scattering. This implies that $\mathcal{P}_{\bar{s}}(t) = e^{-\kappa(t)}$, where the integral for the optical depth is evaluated from t until the present.

Finally, we can calculate the *visibility function,* $g(t)$, which is the probability that a photon was last scattered at a time t. This is given by $g(t)dt = \mathcal{P}_{\bar{s}}(t)\sigma_T n_e dt$, that is the probability that it did not scatter after time t multiplied by the probability that it did scatter during the time interval $[t, t + dt]$. Thus, one finds that $g(t) = \sigma_T n_e e^{-\kappa(t)}$. Generally, around the time of last scattering, the density of free electrons is very quickly decreasing, while $\mathcal{P}_{\bar{s}}(t)$ is growing quickly, approaching unity at late times. Thus the visibility function is peaked at last scattering and is often approximated by a simple Gaussian. (Normalized, as $\int dt g(t) = 1$.)

5.2. **Recombination.** A detailed knowledge of the scattering history requires understanding precisely how and when the electrons and protons combined to form hydrogen. This occurred roughly at a redshift of $z \sim 1100$, when the temperature of the universe was about a third of an eV. (I.e., somewhat less than the electron binding energy.)

At early times, when energetic photons were plentiful, its a good approximation to assume the reaction

$$e + p \longleftrightarrow H + \gamma$$

was in thermal equilibrium. In this case, the chemical potentials for the various species balance, so that $\mu_H = \mu_e + \mu_p$. Combining this with the expression for the equilibrium density of non-relativistic species, we can derive the Saha equation:

$$\frac{1 - x_e}{x_e^2} = \frac{4\sqrt{2}\zeta(3)}{\sqrt{\pi}} \eta \left(\frac{T}{m_e}\right)^{3/2} e^{B/T}.$$

Here $x_e = n_e/n_b$ is the fraction of free electrons, the hydrogen binding energy $B = 13.6\, eV$, $\eta = n_b/n_\gamma$ and is constrained by big bang nucleosynthesis, and $\zeta(3) = 1.20206$.

Eventually the density of energetic photons decreases and the ground state of hydrogen falls out of equilibrium. Any combination into the ground state produces a photon which excites the next ground state it sees, leading to no net change in the fraction of free electrons. However, ground state hydrogen can still be produced via transitions through the higher excited states of hydrogen.

To follow the recombination process in detail, we must keep track of not only the number of ground state atoms, but also the number of atoms in each excited state. Such a calculation is now in principle manageable (for a finite number of excited levels), but it is complex. However, much can be learned from a very accurate approximation introduced by Peebles, which models the hydrogen atom as having effectively two states. This is possible because the higher excited states remain in equilibrium longer than the ground state, and while in equilibrium their occupancy ratios are a fixed function of the temperature. The excited states, while remaining in thermal equilibrium, act as a conduit for free electrons to move into the ground state.

There are two principle routes to producing the ground state electrons. Direct decay from the 2s to the 1s state produces a Lyman α photon which excites another atom. However, if the photon produced is redshifted out of the Lyman α resonance by the cosmic expansion before it hits another ground state atom, then a net ground state will be produced. Alternatively, if the 2s state decays with two photons rather than a single photon, neither of the photons will be able to excite another ground state.

Following even this simple approximation requires evolving a stiff differential equation. The basic result is that there remains a small fraction of free electrons that is frozen in at a level of $10^{-3} - 10^{-4}$.

Its also worth noting that the temperature of the electrons, while initially coupled to the photon temperature, eventually begins to drop faster after equilibrium ends ($z < 400$). While the photon temperature drops inversely with the scale factor, its the velocity of the electrons which drops as the inverse scale factor. The electron temperature is proportional to the energy, or the velocity squared.

See also Peebles (1993) sections 6 & 23. Seager, Sasselov and Scott (1999).

5.3. **Re-ionization.** We believe the electrons and protons combined to form hydrogen at a redshift of $z_{rec} \sim 1100$. However, there is good evidence that this is no longer the case today. Were there large amounts of neutral hydrogen, it would be observed in the spectra of high redshift quasars by Lyman α absorption (the Gunn-Peterson effect.) Studies of quasars show most of the universe was reionized by $z_{rei} > 5$, except for clouds of gas dense enough to be self shielded (Lyman α clouds.) There are tentative observations of the Gunn-Peterson effect in quasars at $z \simeq 6$, which indicates we may be seeing the tail end of reionization.

Reionization most probably occurred when the first generation of stars formed and produced a large flux of ionizing radiation ($E_\gamma > 13.6$ eV.) Precisely when this occurred depends on how much small scale power there is. In typical CDM cosmologies, reionization is not expected to happen until a redshift of $z_{rei} < 20$. However, it might happen much earlier if the initial perturbations were non-Gaussian or were isocurvature in nature.

Reionization is important for the CMB because if it happens early enough, the photons CMB photons can be rescattered, which would erase whatever initial fluctuations previously existed and create new ones. To evaluate whether reionization

is important, we must look at the optical depth

$$\kappa = \int_{t_{rei}}^{t_0} \sigma_T n_e c dt = \frac{0.88 c \sigma_T H_0 \Omega_B}{4\pi G m_p} \Omega^{-2} \left[(2 - 3\Omega) - (2 - 3\Omega - \Omega z_{rei})(1 + \Omega z_{rei})^{\frac{1}{2}} \right],$$

which is related to the probability that a photon is rescattered. If $\kappa > 1$, then it is quite likely that photons will scatter off of the ionized medium. This will depend on the baryon density, but if the it is consistent with nucleosynthesis constraints ($\Omega_b \sim 0.05$), reionization would have to occur at a redshift of $z_{rei} \sim 60$ in order to have a significant impact on the CMB. A significant optical depth will suppress structure on scales smaller than $\theta \sim 2^\circ (1000/z_{rei})^{1/2}$ and will also create large scale polarization.

For typical CDM models where reionization takes place at $z_{rei} \sim 10$, the optical depth is of order a few percent. This tends to give a suppression to the small scale anisotropies. The recent WMAP polarization results imply $\kappa = 0.17 \pm 0.04$, suggesting that reionization might have occurred even earlier, at a redshift of $z_{rei} = 20 \pm 10$.

Finally, another potentially important issue is how quickly reionization takes place. If it is not instantaneous, but is 'patchy', then this can also result in temperature anisotropies being created.

6. Collisional Boltzmann equation

6.1. **Isotropic scattering.** Now that we understand when scattering is likely to occur, we can try to introduce it into Boltzmann equation to find a more realistic evolution of the distribution function. Here we will examine isotropic scattering, and in the next section move on to realistic Thomson scattering taking into account possible polarization of the CMB. We will see that the main implication of a more realistic scattering history is to damp the smaller scale modes.

The effect of scattering is to locally redistribute particles in phase space. In a given element of momentum space, some particles are scattered out of it, while at the same time particles with other momenta are scattered into it. Both processes happen with a rate proportional to the probability of scattering. The Boltzmann equation becomes

$$\frac{df}{d\tau} = \frac{d\kappa}{d\tau} (f_{in} - f),$$

where $f_{in} = \frac{1}{4\pi} \int d\Omega P(\Omega, \Omega') f(\Omega')$ is the number of particles scattered into the phase space of f, and $P(\Omega, \Omega')$ represents a possible angular dependent cross section. The last term represents the scattering out of the phase space of f.

In terms of the photon brightness function, we find the Boltzmann equation is

$$\dot{\Delta} + ik\mu\Delta + 2\dot{h}_{ij}\hat{n}_i\hat{n}_j = \sigma_T n_e a (\delta_\gamma + 4\mu v_b - \Delta)$$

Here the form of the incoming radiation assumes that scattering is isotropic in the rest frame of the scatterer. At early times, $\sigma_T n_e a$ is large and scattering fixes the photon and baryon velocities to be the same. Nearer to decoupling, we can solve this by expanding in powers of the mean free scattering time. Once the mean scattering time grows to be very large, we can use the collisionless Boltzmann equation.

6.2. **Silk Damping.** Just prior to recombination, the photons and baryons are not perfectly coupled. In fact, the photons generally can travel some distance between scatterings from electrons. Very roughly, their mean free path length is $\lambda_c = (\sigma_T n_b x_e)^{-1}$, where x_e is the fraction of free electrons and n_b is the baryon density. This mean free path has two implications for the CMB anisotropies.

First, this allows the photons to diffuse out of the potential wells they had been trapped in, causing the perturbations to become exponentially damped at small scales. At recombination, each photon has had a number of interactions, given by $N \sim c\tau_{rec}/\lambda_c$. The photon diffusion length is thus approximately $\lambda_D \sim \sqrt{N}\lambda_c = \sqrt{c\tau_{rec}\lambda_c}$. This leads to a damping of the shorter wavelength power given by $T(k) \simeq e^{-k^2/k_D^2}$. This is known as diffusion damping, or Silk damping.

A more exact result follows from expanding the collisional Boltzmann equation to second order and solving for how the density evolves in time. (See Efstathiou (1990) or Peebles (1980) for more details.)

6.3. **Thickness of the last scattering surface.** The second effect of the non-negligible photon mean free path is that the recombination is not instantaneous, and the light we see originated from a range of redshifts ($\Delta z \simeq 80$.) This will also smear the small scale microwave anisotropies. If we assume the time of last scattering was Gaussianly distributed, with a standard deviation given by $\sigma\tau_{rec}$, then the small scale anisotropies will be suppressed by $T(k) \sim e^{-k^2\sigma^2\tau_{rec}^2 c^2/2}$. Typically, the width is $\sigma \sim 0.03$.

7. Polarization

7.1. **Origin of CMB polarization.** Polarization arises from Thomson scattering off of electrons in the presence of a quadrupole anisotropy in the incident radiation. The radiation that arises is linearly polarized, not circular, and is typically 5-10% of the level of the anisotropies, though its precise level depends on the thickness of the last scattering surface.

The incident quadrupole anisotropy is generally due to the gradient of the photon velocity. The power spectrum of polarization fluctuations also reflects the oscillations of the photon-baryon fluid. However, peaks in the polarization occur when the velocity was maximum, and are thus out of phase with the Doppler peak structure in the anisotropy spectrum. There is generally no power on modes that were outside the horizon at last scattering, though large scale polarization can result if reionization occurred sufficiently early.

Polarization can provide independent confirmation of the anisotropy measurements and also offer a way for breaking parameter degeneracies that occur with the anisotropies alone. In particular, it is quite useful for constraining the reionization history of the universe and for detecting the presence of tensor fluctuations.

In some ways, measuring the polarization is simpler than measuring the temperature anisotropies themselves, at least from the ground. This is because measurements can be made by finding the difference in two polarizations at a fixed point in the sky. By switching quickly between two polarizers, any contribution from atmospheric effects can be canceled off precisely.

A number of new experiments are now trying to search for CMB polarization and two have detected evidence for it. The DASI interferometer at the South pole detected polarization directly at sub-degree scales. In addition, WMAP has

detected the temperature-polarization cross correlation, primarily on large scales and used it to constrain the scattering optical depth.

7.2. **Stokes parameters.** Let us now consider the physics of polarization in more detail. Consider radiation of frequency ω traveling in the $\widehat{z}$ direction. Its electric field can be written as

$$\begin{aligned} E_x &= a_x \cos(\omega t - \theta_x) \\ E_y &= a_y \cos(\omega t - \theta_y). \end{aligned}$$

We can parameterize the degree of polarization in terms of four Stokes parameters:

$$\begin{aligned} I &= \left\langle a_x^2 \right\rangle + \left\langle a_y^2 \right\rangle \\ Q &= \left\langle a_x^2 \right\rangle - \left\langle a_y^2 \right\rangle \\ U &= \langle 2 a_x a_y \cos(\theta_x - \theta_y) \rangle \\ V &= \langle 2 a_x a_y \sin(\theta_x - \theta_y) \rangle \end{aligned}$$

Here, the averages are over times long compared to the period of the wave. I measures the total intensity, while Q and U parameterize the magnitude and direction of the linear polarization. V measures the level of circular polarization, which is not usually excited in the CMB context.

If we consider rotating the coordinate basis by some angle ϕ, the measures of linear polarization transform as

$$\begin{aligned} Q' &= Q \cos 2\phi + U \sin 2\phi \\ U' &= U \cos 2\phi - Q \sin 2\phi. \end{aligned}$$

Rotating by 45° takes $Q \to U$ and $U \to -Q$, while a rotation of 180° returns to the same values. This is because the polarization is represented by its amplitude and its plane rather than by an ordinary vector. For this reason, polarization is a spin-2 field and it is often represented by a 2×2 symmetric traceless matrix:

$$\widetilde{P} = \begin{pmatrix} Q & U \\ U & -Q \end{pmatrix}.$$

7.3. **The polarized Boltzmann equation.** To follow the creation and evolution of polarization fluctuations, we must modify the Boltzmann equations to follow the polarization degrees of freedom and also include the anisotropic nature of Thomson scattering. Thomson scattering has the differential cross section given by,

$$\frac{d\sigma}{d\Omega} = \frac{3\sigma_T}{8\pi} \left| \epsilon' \cdot \epsilon^* \right|^2,$$

where ϵ and ϵ' are the polarization vectors of the incoming and outgoing radiation, respectively. Even if the incoming radiation is unpolarized, radiation scattered at 90° will be leave completely polarized.

This is included in the Boltzmann equation by treating the distribution function as a vector, with each element representing a different Stokes parameter. For the case of scalar fluctuations, no U modes are generated and the distribution functions can be represented by two brightness functions, Δ_T and Δ_P. The anisotropy of Thomson scattering couples these equations, with the polarization being sourced by the quadrupole of temperature distribution, Δ_{T2}. The polarization feeds back into the evolution of the temperature fluctuations, damping them at smaller scales compared to the isotropic scattering results.

In the tightly coupled limit, the photon-electron fluid acts as an ideal fluid described by its density and velocity alone. The polarization and temperature

quadrupole are zero. As the time of last scattering approaches, we can solve for the polarization by expanding in powers of the conformal mean scattering time, τ_c. To first order, we find $\Delta_{T2}, \Delta_P \propto ik\mu v\tau_c$; that is, we find the quadrupole moment and the induced polarization are proportional to the gradient of the velocity times the mean free scattering time. If we integrate this through the last scattering surface, we find $\Delta_{T2}, \Delta_P \propto ik\mu v\,\Delta\tau_{rec}$, where the width of the visibility function has replaced the mean scattering time. This assumes the velocity is fairly constant over the surface of last scattering; for small scales this is not the case and the polarization becomes suppressed by the thickness of the last scattering surface.

Thus the polarization traces the velocity field on the surface of last scattering. If the velocities are inwardly converging into a point, then it will be surrounded by a radial polarization pattern. If instead the velocities are diverging, there will be a tangential polarization pattern. The polarization power spectra will be greatest on scales where the velocities were biggest on the last scattering surface. Thus the peaks will be out of phase with the intensity peaks which follow the density of the fluid (see below.)

7.4. **Correlation functions.** In general, we can look for correlations between any of the Stokes parameters: I, Q or U. However, certain correlations must be zero if the universe respects parity invariance. Under a reflection about the $\widehat{y}-$axis, the total intensity and Q polarization of a point remains unchanged, but $U \rightarrow -U$. Two point correlations between things of opposite parity, such as $\langle TU\rangle$ or $\langle QU\rangle$ must be zero if the universe is parity invariant. Note also that $\langle TQ\rangle$ must approach zero at small separations where the polarization coordinate basis becomes undefined.

In addition to temperature and polarization auto-correlations, one can also search for correlations of the polarization with the temperature anisotropy. These can be positive or negative and oscillate with the frequency of the photon fluid, but are out of phase with either the temperature or the polarization correlation functions. In real space, these will look like net radial or tangential polarization patterns around hot or cold spots in the temperature distribution.

Observation strategies for detecting $\langle QQ\rangle$ and $\langle QT\rangle$ differ significantly. The former require noise levels to be of order the size of the polarization signal, so observations must focus on a fairly small section of the sky to reduce the noise. The cross correlation can be detected with a larger noise level, as long as the temperature fluctuations are measured reliably. The noise in polarization cancels out when sufficiently large areas of the sky are observed. Polarization may well be first detected by its cross correlation with temperature - in particular the MAP satellite will have noise levels too high to detect the polarization auto-correlation, but it will survey the entire sky and thus will be able to detect the cross correlation fairly easily.

7.5. **Electric and magnetic modes.** At any point on the sky, the polarization has two degrees of freedom, its amplitude and inclination angle. In general, the polarization can be decomposed into two kinds of modes: electric (E), curl-free modes and magnetic (M), gradient-free modes. Locally, E-modes are either tangential or radial configurations of the polarization pattern, while B-modes have a either a clockwise or counterclockwise handedness. (Figure?) A pure E-mode configuration can be transformed into a pure B-mode by rotating each polarization plane by $\pm 45^\circ$ (and vice versa.)

The polarization matrix can be expanded in orthogonal functions just like the temperature,

$$\widetilde{P} = \sum_{l,m} \left(a^E_{lm} \widetilde{Y}^E_{lm} + a^B_{lm} \widetilde{Y}^B_{lm} \right),$$

where the $\widetilde{Y}^{E,B}_{lm}$ are an orthogonal basis of symmetric, traceless 2×2 matrices. These can get a bit complex when dealing with large scale polarization.

7.6. **Cosmological constraints from polarization.** Perhaps the most important thing to be learned from polarization is the time of last scattering. Polarization is a causal process and we do not expect correlations in it on scales larger than the horizon at last scattering. Thus, polarization can be used to infer the time of last scattering. Polarization correlations on very large scales, like those seen in WMAP , indicate that the universe was reionized early enough to achieve a significant optical depth for the photons to rescatter.

In addition, the E-B decomposition of the polarization will be important. Linear scalar fluctuations create only E modes. However, gravitational waves can create both E and B modes. Thus a detection of intrinsic B-modes would be good evidence for significant tensor fluctuations. The expected levels of B modes from gravitational waves, though, are expected to be small, making the observations quite challenging.

8. Initial conditions

Now we have a complete set of coupled differential equations which we can evolve from early times, when the modes of interest were far outside the horizon, until the present. These include the Einstein equations, fluid equations for cold dark matter and baryons, Boltzmann equations for photons (temperature and polarization) and neutrinos. Finally we must also consider the evolution of what ever dark energy component is making up 70% of the matter in the universe, which is particularly easy in the case of a cosmological constant.

All of these equations must be supplemented by the initial conditions of the perturbations, which are determined by a model of the early universe. Here we discuss the various possibilities for initial conditions.

8.1. **Active vs. passive.** Some models for the origin of structure include sources which still exist today, whose behaviour must also be followed along side the other perturbations. Such sources are known as *active sources.* Other models, such as inflation, where only the initial conditions are changed early in the universe, are known as *passive sources.*

The best known examples of active sources are cosmic defect theories, such as cosmic strings and cosmic texture (e.g. Shellard & Vilenkin). The defect evolution is non-linear and occurring constantly as the defects attempt to unwind themselves as they enter the horizon. Typically, the defects interact only gravitationally through their influence on the Einstein equations. Because of their non-linear evolution, the predictions for defects are considerably harder to determine. At present, though, they appear to be an unlikely candidate for the primary source of anisotropies.

8.2. **Casual vs. acausal.** Cosmic defects are also a very good example of a causal mechanism for structure formation, because they do not require correlations on scales larger than the horizon. Inflation naturally produces correlations on super-horizon scales, which in the standard big bang context would appear to be acausal.

8.3. **Adiabatic.** The simplest kinds of initial conditions are where the number densities of all the constituent particles fluctuate identically. These are known as *adiabatic* fluctuations and are necessarily curvature fluctuations. For adiabatic fluctuations, $\frac{\delta n_\gamma}{n_\gamma} = \frac{\delta n_c}{n_c} = \frac{\delta n_b}{n_b}$ and for all the particle species. This implies that $\delta_c = \delta_b = \frac{3}{4}\delta_\gamma$, since the photons are relativistic and $n_\gamma \propto \rho_\gamma^{3/4}$. In addition, since the perturbations were produced in the early universe, it is assumed that only the growing mode of the perturbations is important today, so that $\delta \propto \tau^2$.

It is worth describing a bit more about the statistics of the fluctuations. Throughout we have been considering the evolution of the Fourier transformed amplitudes defined by

$$\delta(\mathbf{x}) = \frac{V}{(2\pi)^3}\int d^3\mathbf{k}\, e^{i\mathbf{k}\cdot\mathbf{x}}\delta_\mathbf{k}.$$

The Fourier modes are characterised by an amplitude and a phase. In the linear regime the amplitudes evolve independently of each other, while the phases remain preserved. For Gaussian fields, the phases are random. The linear evolution preserves any non-Gaussian information which is in the phase correlations.

For Gaussian random fields, all the information is contained in the two point correlation function, or in Fourier space in the power spectrum, $P(k)$. The ensemble average of two Fourier modes is given by

$$\langle \delta_\mathbf{k}\delta_{\mathbf{k}'}\rangle = \frac{V}{(2\pi)^3}P(k)\,\delta_D(\mathbf{k}-\mathbf{k}'),$$

where the Dirac delta function is the consequence of homogeneity. In practice, we usually assume that the ensemble averages are equivalent to averages over large enough volumes (i.e. the spatial ergodic theorem.) The real space two point function can be shown to be the Fourier transform of the power spectrum,

$$\langle \delta(\mathbf{x})\delta(\mathbf{x}')\rangle = \frac{V}{(2\pi)^3}\int d^3\mathbf{k}P(k)e^{i\mathbf{k}\cdot\mathbf{x}}.$$

The variance, or the correlation function at 0°, is given by $\langle\delta^2\rangle = \frac{V}{(2\pi)^3}4\pi\int k^2dkP(k)$.

8.3.1. *Scale independent and invariant spectra.* The most widely considered class of spectra are known as *scale independent*, and are characterized by a simple power law $P(k) \propto k^n$, where n is known as the spectral index.

A special subclass of these is the *scale invariant* spectrum, which is also known as the Peebles-Harrison-Zeldovich spectrum. Here, the contribution to the variance is the same for all modes entering the horizon. Outside the horizon, $\delta \propto \tau^2$, so $P(k) \propto \tau^4 k^n$. From the expression for the variance, the contribution per log interval is $k^3P(k) \propto \tau^4 k^{3+n}$ which should be invariant at horizon crossing $k\tau \sim 1$. This implies that a scale invariant spectrum has $n = 1$.

Inflation usually predicts a nearly scale invariant spectrum, with $n \simeq 0.92 - 1.0$. Current observational constraints place the spectral index in the range $n = 1 \pm 0.1$.

8.4. **Isocurvature modes.** The alternative to adiabatic fluctuations are known as *isocurvature* fluctuations, or perhaps more accurately as *entropy* fluctuations. These have initial conditions where the ratios of various particle species vary from place to place. However, they may be initially compensated such that the density fluctuations cancel. But the modes that begin as isocurvature may not remain so, particularly after matter domination.

There are as many different kinds of isocurvature fluctuations as there are pairs of particle species. For example, often the number of photons per dark matter particle is assumed to vary, but one could also consider variations in the number of photons per baryon, or baryons per dark matter particle. There are many ways of exciting isocurvature modes, from multifield inflationary models to axions produced in the early universe. These modes may also co-exist with adiabatic modes and there may be correlations between them.

The evolution of the modes is most easily followed by defining an analog to the Bardeen variable for each particle species. That is, we define

$$\zeta_N = \frac{1}{3}\left(\frac{1}{2}h_- + \frac{\delta\rho_N}{\rho_N + p_N}\right),$$

which is the curvature fluctuation on surfaces of constant density in a given particle species. These can be shown to be conserved outside the horizon for most kinds of particles by stress-energy conservation. Isocurvature perturbations are usually quantified by

$$S_{ij} = \frac{\delta n_i}{n_i} - \frac{\delta n_j}{n_j} = 3\left(\zeta_i - \zeta_j\right),$$

which is also conserved outside the horizon. However, the full Bardeen variable, which is a weighted sum of the individual terms,

$$\zeta_B = \sum_N \zeta_N(\rho_N + p_N)/\sum_N(\rho_N + p_N)$$

is not necessarily conserved outside the horizon. (Note that for adiabatic fluctuations, $\zeta_B = \zeta_c = \zeta_\gamma \ldots$ and all are conserved outside the horizon.) See Wands, Malik, Lyth & Liddle (2000) for further discussion.

9. Reading the power spectrum

9.1. **SW effect - Superhorizon evolution.** Let us consider the particular case of superhorizon modes in the matter dominated regime. Recall that there is a growing mode solution $\delta_m = Ak^{n/2}\tau^2$. The fluid and Einstein's equations imply, $-2\dot{\delta}_m = \dot{h} = \dot{h_S}$ when the velocities are negligible, so that $h_S = -2Ak^{n/2}\tau^2 + C$. Next, Poisson's equation, $\nabla^2\Phi = 4\pi G\delta\rho$, implies the gravitational potential is constant: $\Phi = -\frac{3}{2}\left(\frac{2}{\tau}\right)^2\delta_m/k^2 = -6Ak^{n/2-2}$. (Note this is the same as one finds by transforming to Newtonian gauge.) On superhorizon scales, the Sachs-Wolfe expression is dominated by the $\ddot{h}_S/2k^2$ term, so the temperature anisotropy is effectively

$$\frac{\delta T}{T} \simeq \ddot{h}_S/2k^2 = -2Ak^{n/2-2} = \frac{1}{3}\Phi.$$

This expression is easier to understand in the Newtonian gauge. Naively, one expects the redshifting of photons to be given by the depth of the potential well, $\delta T/T = \Phi$. However, in the cosmological setting one must also take into account the expansion. For adiabatic fluctuations, recombination occurs later in potential

wells because the photon temperature is higher there. Thus, these photons have had less time to expand. The delay in time is given by the depth of the potential, $\delta t/t = \Phi$. Since the scale factor grows as $t^{\frac{2}{3}}$ in the matter dominated regime, the fluctuation in scale factor is given by $\delta a/a = 2\Phi/3$. This shift cancels part of the naive contribution, leaving $\delta T/T = \Phi/3$.

In terms of the matter fluctuation, the temperature anisotropy can be written as

$$\frac{\delta T(\mathbf{n})}{T} = \sum_{\mathbf{k}} \frac{2}{k^2\tau^2}\delta_{\mathbf{k}} e^{i\mathbf{k}\cdot\mathbf{n}\tau}.$$

Using the expansion of planes waves in spherical harmonic functions, the multipole moments can be shown to be,

$$a_{\ell m} = \int \frac{\delta T(\mathbf{n})}{T} Y^*_{\ell m}(\mathbf{n}) d\Omega = 8\pi \sum_{\mathbf{k}} \frac{1}{k^2\tau^2}\delta_{\mathbf{k}} \mathbf{i}^\ell j_\ell(k\tau) Y^*_{\ell m}(\widehat{\mathbf{k}}).$$

For a scale invariant spectrum, we can write the power spectrum as $P_k \equiv |\delta_{\mathbf{k}}|^2 = Bk\tau^4/L^3$, where have factored out the comoving time and wavenumber dependence and the L^3 factor assures that B is dimensionless. The CMB anisotropy spectrum then is

$$C_\ell = |a_{\ell m}|^2 = \frac{64\pi^2}{L^3} B \sum_{\mathbf{k}} k^{-3} j^2_\ell(k\tau) Y_{\ell m}(\widehat{\mathbf{k}}) Y^*_{\ell m}(\widehat{\mathbf{k}}).$$

Taking the limit as $L \Rightarrow \infty$, the sum over modes becomes an integral, $\sum_{\mathbf{k}} \Rightarrow (L/2\pi)^3 \int d^3\mathbf{k}$. The angular part of this integral is trivial by the orthogonality of the spherical harmonics, which leaves,

$$C_\ell = \frac{8B}{\pi} \int \frac{dk}{k} j^2_\ell(k\tau) = \frac{4B}{\pi\ell(\ell+1)}.$$

This is the reason CMB spectra are usually plotted as $\ell(\ell+1)C_\ell$ versus ℓ. More generally, for a power law density spectrum $P_k \propto k^n$, one can show

$$C_\ell \propto \frac{\Gamma\left(\ell + \frac{n-1}{2}\right)}{\Gamma\left(\ell + \frac{n-5}{2}\right)}.$$

9.2. **Baryon-photon oscillations.** On large angular scales, greater than a degree or two, the fluctuations leading to temperature anisotropies are outside the horizon at the time of recombination. As described above, their resulting spectrum is fairly featureless if the underlying initial density spectrum can be described as a power law. On smaller scales, however, the modes have had time to evolve within the horizon prior to last scattering. On these scales, the anisotropies are dominated by the intrinsic anisotropy and to a lesser extent by the Doppler effect. As a result, the oscillations of the photon-baryon fluid take on a key roll.

Prior to recombination, the photons and baryons are strongly coupled and thus act as a single fluid. They then share the same velocity $\mathbf{v}_b = \mathbf{v}_\gamma$ and their densities are related by the adiabatic condition that the number of photons per baryon is fixed, $\delta_\gamma = 4\delta_b/3$. This fluid acts as one with pressure $\frac{1}{3}\overline{\rho}_\gamma$, but which has density $\overline{\rho}_\gamma + \overline{\rho}_b$,

$$\frac{3}{4}\dot{\delta}_\gamma = \dot{\delta}_b = -\frac{1}{2}\dot{h} - ikv_\gamma$$

$$(\overline{\rho}_b + \frac{4}{3}\overline{\rho}_\gamma)\dot{v}_\gamma + \overline{\rho}_b \frac{\dot{a}}{a} v_\gamma = -\frac{ik\overline{\rho}_\gamma}{3}\delta_\gamma$$

The speed of sound in this fluid is

$$c_s^2 = \frac{\delta P}{\delta \rho} = \frac{\frac{1}{3}\bar{\rho}_\gamma}{\frac{3}{4}\bar{\rho}_b + \bar{\rho}_\gamma} = \frac{1}{3(1+R)},$$

where $R \equiv 3\bar{\rho}_b/4\bar{\rho}_\gamma$.

The velocity equation can be rewritten using in terms of the sound speed as

$$\dot{\mathbf{v}}_\gamma + \frac{\dot{a}}{a}(1-3c_s^2)\mathbf{v}_\gamma + \frac{3}{4}c_s^2 i\mathbf{k}\delta_\gamma = 0.$$

Combining this with the density evolution equation, we find that the density acts as a forced, damped harmonic oscillator:

$$\ddot{\delta}_\gamma + c_s^2 k^2 \delta_\gamma = -\frac{2}{3}\ddot{h}.$$

Here for simplicity we have dropped the damping terms, which is appropriate in the radiation dominated regime ($c_s^2 \simeq 1/3$.) Following decoupling, the photons freestream and the baryons fall into the potential wells generated by the dark matter.

9.2.1. *The Doppler peaks.* Oscillations for a given wavenumber begin when $k\tau c_s \simeq 1$ and their subsequent evolution behaves roughly as $\delta \sim \cos(k\tau c_s + \psi)$. The phase offset depends on the initial conditions and on how the perturbations are forced, but is fairly independent of the wavenumber.

The CMB anisotropies probe the the density fluctuations at a fixed time, τ_{rec}. Each mode at this time will be at a different point in its oscillation, depending on how long that mode has been inside the horizon. Modes just entering the horizon are beginning their first collapse due to gravitational attraction. Slightly smaller modes will have had time to collapse to a point where their pressure was sufficient to force a bounce, and will be expanding back again. Much smaller wavelengths will have had time for multiple compressions and rarefactions.

The peaks in the CMB spectrum on scales smaller than a degree are associated with those wavenumbers that have just reached the peak of their compression or rarefaction at τ_{rec}. The term Doppler peaks is a bit of misnomer, since they are created at minima and maxima of the density fluctuations, where the velocities are actually zero. The true Doppler effect resulting from the velocities does contribute, but it is smaller than the intrinsic piece and is 90° out of phase. It has the effect of filling up the space between the Doppler peaks, making the dips shallower than they would be otherwise.

Note that there is a further blurring of the peak structure resulting from the projection of the modes from three dimensions down to two. Each ℓ mode receives contributions from a range of wavenumbers, depending on the angle at which the mode intersects the sphere of the sky.

9.3. **How the spectrum depends on parameters.** The physical evolution of CMB anisotropies depends sensitively on a number of cosmological parameters. If one assumes a simple form for the initial perturbations, then measurements of the CMB anisotropies can be used to constrain the cosmological model.

9.3.1. *Angle-distance relation.* The photon-baryon oscillations provide a useful physical yardstick for the early universe. By knowing when last scattering occurred, we can infer the total curvature of the universe from the angle-distance relation. The comoving time of recombination, assuming it happens sufficiently after the matter-radiation equality, is given by $\tau_{rec} = 2H_0^{-1}\sqrt{\Omega_m^{-1}(a_{rec}/a_0)}$. In a flat, matter dominated universe, the comoving distance to the last scattering surface is given by $r_{rec} \simeq \tau_0 = 2H_0^{-1}\sqrt{\Omega_m^{-1}}$, so the angular scale of the sound horizon at last scattering is $\theta = c_s\tau_{rec}/r_{dec} = c_s\sqrt{a_{rec}/a_0} \simeq 1^\circ$.

If the universe is open, the comoving distance to the last scattering surface is no longer equal to the comoving time since last scattering, but is given by $r_{rec} = 2H_0^{-1}\Omega_m^{-1}$. Then the relevant angular scale is $\theta = c_s\sqrt{\Omega_m(a_{rec}/a_0)}$. (If the matter density is sufficiently low that the matter-radiation transition is concurrent with recombination, then this will shift the coming time of last scattering somewhat. We have neglected this effect here.)

Finally, for universes with a cosmological constant, $r_{rec} \simeq \tau_0 = 2H_0^{-1}\Omega_m^{-1/2}F(\Omega_m)$, where the hyper-geometric function $F(\Omega_m)$ is a slowly (logarithmically) varying function of the matter density, ranging from 1 ($\Omega_m = 1$) to 0.8 ($\Omega_m = 0.1$). Thus the angular scale of the Doppler peaks is fairly independent of the matter density for flat universes, but it can be increased somewhat for large values of the cosmological constant.

9.3.2. *Matter density.* Another important factor in generating the CMB anisotropies is the present matter density. Lowering the matter density makes the time of radiation-matter equality later, $z_{eq} = 2.3 \times 10^4\Omega_m h^2$. This effects both the forcing term for the photon-baryon oscillations and the contributions from the integrated Sachs-Wolfe effect. The forcing term is stronger near radiation-matter equality, so the heights of the Doppler peaks are higher when the matter density is lower.

9.3.3. *Baryon density.* Finally, the physical baryon density, $\Omega_b h^2$, is also critical to the evolution of the microwave anisotropies. Recall that the photon-baryon fluid acts as a fluid with pressure $p = \rho_\gamma/3$ and density $\rho = \rho_b + \rho_\gamma$, so that increasing the baryon fraction at recombination makes the fluid heavier. This causes the fluid to compress more and bounce back less, shifting the effective zero point of the oscillations so that $\delta_{compression} > \delta_{expansion}$. Thus the (odd) Doppler peaks associated with the compression will be higher than those associated with the rarefaction (the even numbered peaks).

Another effect of increasing the baryon density is that the sound speed is lowered, so that the peak structure can be shifted. However, this effect is fairly small as $c_s^2 = R/3(1+R)$, where the ratio is typically $R = 4\bar{\rho}_\gamma/3\bar{\rho}_b \sim 3-4$ at recombination if nucleosynthesis is correct. Such a high value means that the baryon density needs to be changed a great deal to effect the speed of sound.

9.4. **Damping effects.** As discussed above, diffusion effects damp out the smallest scale modes. In addition, the observed temperature anisotropies are also blurred by the finite thickness of the last scattering surface. These effects lead to an exponential suppression of the high ℓ modes, for which $k\Delta\tau_{rec} \geq 1$.

9.5. **The integrated Sachs-Wolfe effect.** The integrated Sachs-Wolfe contribution arises along the photon path from last scatter to today, when it passes through

a time varying gravitational potential. Effectively, the photon receives a shift because the potential it falls into is different from the potential it must climb out of.

For the growing mode during the matter dominated epoch, the gravitational potential is actually constant because the growth of perturbations precisely compensates for the decreasing mean energy density, so that $\delta\rho = \bar{\rho}\delta$ is constant. Thus, the ISW effect contribution is most important before the matter-radiation transition or after matter domination ends (if there is a cosmological constant or a quintessence field.) The former effect influences the heights of the first few Doppler peaks, while the latter can create additional fluctuations on very large scales (i.e. at low ℓ.)

The large scale excess power resulting from a cosmological constant is hard to verify directly due too cosmic variance. However, it may be measured indirectly via correlations of the CMB with the local matter distribution. That is, matter with a mean redshift of $z \sim 1$. Possible probes for this structure include the radio and x-ray backgrounds.

9.6. **Vector and tensor modes.** Vector modes and gravitational radiation can also contribute to the significantly to the CMB anisotropies via the integrated Sachs-Wolfe effect. In the absence of sources for these modes, vector perturbations decay away quickly, $h^v \propto t^{-1}$. These are not expected to be important for inflationary scenarios, but could prove important if there are active sources for perturbations such as cosmic defects.

Gravitational waves obey the Einstein equation,

$$\ddot{h}^T_{ij} + 2\frac{\dot{a}}{a}\dot{h}^T_{ij} + k^2 h^T_{ij} = -16\pi G\Theta^T_{ij}.$$

In the absence of sources, the solutions to this are given by spherical Bessel functions, i.e. $h^T_{ij} \propto j_1(k\tau)/k\tau$ in the matter dominated regime. These tensor modes are frozen in outside the horizon, but decay away quickly upon entering. Thus, the temperature anisotropies from gravitational radiation are predominately on scales that were outside the horizon at decoupling. If the relative contribution of tensor fluctuations to the anisotropies ($T/S \equiv C_2^T/C_2^S$) is small, then these can be difficult to detect because of cosmic variance.

9.7. **Isocurvature and defect models.** Thus far we have been focusing primarily on adiabatic curvature fluctuations, which are fluctuations in the overall density, but where the relative numbers of the different particle species are uniform throughout space (e.g. $\frac{3}{4}\delta_\gamma = \delta_b = \delta_c$.) We can also consider isocurvature, or entropy, perturbations, in which the density of matter on superhorizon scales is uniform, but the relative densities of particle species fluctuate.

The relation between the temperature anisotropies and the amplitude of large scale structure is significantly different for isocurvature fluctuations. Recall that on large scales, adiabatic perturbations predict $\delta T/T = \Phi/3$. However, for isocurvature modes the corresponding temperature fluctuation is higher, $\delta T/T \sim 2\Phi$. Thus for a given amplitude of CMB anisotropies, the large scale structure predicted is much lower in isocurvature theories, and so large bias factors are generally required to explain the observations.

The phase of the photon-baryon oscillations in isocurvature models is shifted relative to that in adiabatic models. This is because the oscillations are initially

not offset outside the horizon, but are started when the modes enter the horizon due to the isocurvature forcing term. Thus the modes are approximately 90° out of phase with the adiabatic result.

9.8. **Cosmic defect anisotropies.** Cosmic defects are effectively isocurvature perturbations with an active source whose evolution is non-linear. The source is most important just after the modes enter the horizon and the defects begin to straighten themselves. The non-linear evolution of this process means that the source is not coherent, and the defects will source modes in some regions before others. As a result, the Doppler peak structure of the scalar spectrum can be smoothed out in defect models.

Probably the most important effect of the active defect sources is that their vector and tensor anisotropies are typically comparable to the scalar anisotropies and cannot be neglected. These will greatly increase the large scale power in the CMB spectrum, but will not affect the small scale structure. Thus, the Doppler peak structures, if they do exist, will be suppressed relative to the COBE scale anisotropies.

10. Observing the CMB fluctuations

10.1. **Cosmic variance.** If the fluctuations are Gaussian, the multipole moments are all independent. Measurement of the correlation function is fundamentally limited by the fact that we observe only one sky. The power spectrum C_ℓ is discrete, and each ℓ value is sampled only $2\ell+1$ times. This results in an unresolvable uncertainty in the spectrum measurement of $\delta C_\ell \simeq \sqrt{2}C_\ell/(2\ell+1)^{\frac{1}{2}}$. This is known as *cosmic variance.*

10.2. **Higher order effects.** Non-linear effects can also produce anisotropies in the microwave background and these are usually important at smaller scales (< 0.1 degrees.) However, if nonlinear effects lead to the universe being ionized at high enough redshifts, the entire CMB spectrum can be dramatically altered.

Higher order effects can produce anisotropies at late times, once the perturbations become large. These can come from the CMB photons being rescattered via Compton scattering, once the electrons are reionized, or when gravitational non-linearities change the energy or direction of the photons.

10.2.1. *The Sunyaev-Zeldovich effect.* One notable mechanism which creates temperature fluctuations and affects the CMB spectrum is inverse Compton scattering of photons off of hot electrons in clusters, known as the Sunyaev-Zeldovich effect. This effect gives a way to measure directly the baryonic mass in clusters at high redshifts. Scattering off of hotter electrons causes a simple energy shift parameterized by the Compton $y-$parameter,

$$y = \frac{\delta\nu}{\nu} = \int \sigma_T n_e \frac{kT_e}{m_e c^2} dl$$

where σ_T is the Thomson scattering cross section, n_e is the free electron density and the integral is along the photon path through the cluster.

The photons are upscattered, causing the low energy tail of the spectrum to be shifted to higher frequencies,

$$\frac{\delta n_\nu}{n_\nu} = -y \frac{x e^x}{e^x - 1} \left[4 - x \coth(x/2)\right],$$

where $x \equiv h\nu/kT_\gamma$. This results in a temperature deficit at frequencies below the blackbody peak $\Delta T/T = -2y$, and a temperature increase at higher frequencies, with the sign changing at $\nu \sim 220$ GHz. Typical clusters have temperatures of order $10^7 - 10^8 K$, and yield Compton-y parameters of order $y \sim 10^{-4}$.

10.2.2. *The Doppler effect.* Another important effect arises if there is significant bulk motion in the ionized medium, which induces a temperature fluctuation via the Doppler effect:

$$\frac{\Delta T}{T} = \int v_\parallel n_e \sigma_T dl$$

Unlike the S-Z effect, the Doppler shift does not change the spectrum of the anisotropies and so is harder to distinguish from the primary anisotropies.

This effect is second order in linear theory and can be larger when non-linear effects are included. The second order effect is usually known as the Ostriker-Vishniac effect. It can also occur for clusters with proper motions along the line of sight, where it is known as the kinetic S-Z effect.

10.2.3. *The Rees-Sciama effect.* The Rees-Sciama effect is related to the integrated Sachs-Wolfe effect and results from the non-linear evolution of the gravitational potential in clusters. The integrated S-W effect usually concerns the large scale linear evolution of the potential, while the Rees-Sciama effect depends on its non-linear evolution inside a cluster.

10.2.4. *Lensing by large scale structure.* Clumping matter between us and the last scattering surface bends the CMB light rays and changes the CMB patterns. While it cannot generate anisotropies itself, it can transfer power from large angular scales to shorter ones. It may be an important source of non-Gaussianity in the microwave background and could be used to probe the large scale matter power spectrum. Lensing can also induce anisotropies from the motion of clusters transverse to the line of sight.

10.3. **Foregrounds.** There are a variety of foreground sources that can contaminate our observations of the microwave background. The most important foreground contaminant is our own galaxy, which emits from a number of different kinds of sources. These foregrounds are so strong that observations are only possible outside the galactic plane.

What makes observations of the CMB possible at all is that its frequency dependence is generally greatly different than the frequency dependence of the foregrounds. Some sources dominate at high frequencies, and others at low frequencies, but there exits a frequency window around 50-100 GHz where the CMB is the dominant source. The behavior of foregrounds can be studied at higher and lower frequencies and these observations can be used to help remove their contamination from the CMB maps.

10.3.1. *Thermal dust.* One foreground which is most important at high frequencies is that from thermal emission from dust. The primary source of this dust is in our own galaxy, where it is typically of order 20K, heated by optical or UV radiation. The dust is typically carbon or silicate in composition.The frequency spectrum depends on the size of the dust grains (typically 0.01-0.1 μm,) but it scales as a power law in the microwave frequencies.

This dust is distributed on the sky more smoothly distributed than we expect the CMB to be. Thermal emission from high redshift star forming galaxies can also be important.

10.3.2. *Synchrotron emission.* Both magnetic fields and energetic electrons are required in order to generate synchrotron emission. Synchrotron contamination thus tends to come from the disk of our galaxy where the magnetic fields are highest. Its tends to dominate at low frequencies, below 30 GHz or so, and has a spectrum $T \propto \lambda^{2.4\pm0.3}$.

Synchrotron emission can also come from radio galaxies. The contribution from radio galaxies will have a Poisson distribution on the sky and is an important foreground source on small angular scales, as it cannot be avoided by simply looking away from the galactic plane. Unfortunately, the frequency dependence for these sources changes from source to source, and the sources may also vary in time, so it is necessary to observe them at the same time as the CMB observations.

10.3.3. *Bremsstrahlung radiation.* Bremsstrahlung radiation, also known as free-free radiation, results from electrons scattering off of ionized hydrogen or helium atoms. Like synchrotron, it primarily originates from the disk of our galaxy, where the densities of ionized particles are highest. It has a frequency dependence very close to that of synchrotron radiation, $T \propto \lambda^{2.1}$, and has a similar intensity.

10.3.4. *Atmospheric thermal fluctuations.* The major problem with observing CMB anisotropy from the ground is thermal emission from water and oxygen molecules in the atmosphere. Water vapor is particularly a problem because it varies in time and space. One way to counter this problem is to make measurements of temperature differences along two lines of sight, where the atmospheric contributions should cancel statistically. To reduce the atmospheric contamination, CMB measurements are generally taken at sites which are as dry and cool as possible, such as at the South pole or on mountain tops.

10.3.5. *Instrumental effects.* While each experiment offers its own set of problems, there are a couple of generic problems in detecting CMB anisotropy. The first issue is to do with sidelobes, or sensitivity to sources in a direction other than that of the central beam. The horn system is designed to minimize the system's response in the sidelobes, but even so, bright sources (such as the earth or sun) can pass through them and get confused with the signal.

The other generic issue is amplifier drift. The detector signals must be amplified by many orders of magnitude and there is unavoidable slowly varying drift in the amount of amplification (which is usually assumed to have a 'one over f' noise spectrum.) This drift, if not corrected for, can result in 'striping' of the inferred temperature maps which reflects the observing pattern of the instrument.

10.4. **Kinds of detectors and platforms.**

10.4.1. *Detectors.* There are three types of detectors in common use today: interferometers, bolometers and waveguides with HEMT (high electron mobility transistor) amplifiers. HEMT detectors are generally used at lower frequencies (below 100 GHz) and bolometers, with their broader band powers, are used at higher frequencies.

10.4.2. *Platforms.* The CMB anisotropy may be observed from the earth, from high altitude balloons or from satellites. There are advantages and disadvantages to each platform. Ground based observations must contend with contamination from the atmosphere which leads to much higher noise levels. The advantages of ground based detectors are that you can integrate for much longer and are able to fix any problems that arise. In addition, instruments observing at very small angular scales or with interferometers are too massive to be used anywhere than from the ground.

Balloon based detectors do not have to deal with the atmospheric foregrounds, but have much less time to observe and cannot be fixed in flight if a problem arises. Typically, balloons can observe for only a few hours, though long duration balloon flights around the south pole have been able to extend this to over a week.

Satellite missions generally have plenty of integration time and have less potential for systematic errors than ground or balloon based detectors. However, the cost of satellites can be prohibitive and such missions take many more years of planning to carry out. As a result, satellites generally cannot use state of the art technology.

10.5. **Window functions.** Observations do not measure the sky temperature directly, but measure it convolved with an experimental beam shape,

$$T_E(\mathbf{x}) = \int W(\mathbf{x}-\mathbf{y})T(\mathbf{y})d^2\mathbf{y}.$$

This convolution is easiest to evolve in Fourier space, where it is simply a product, i.e. $T_E(\mathbf{k}) = W(\mathbf{k})T(\mathbf{k})$. In general, the beam shape can be quite complicated, including being anisotropic and having significant significant sidelobes. It is sometimes a useful approximation to treat the central beam as a Gaussian function, with $W(\theta) = e^{-\theta^2/2\sigma^2}/2\pi\sigma^2$. This has a simple Fourier representation, $W(k) = e^{-k^2\sigma^2/2}$. The variance angle is related to the full width-half maximum size of the beam by $\theta^2_{FWHM} = 8\sigma^2 \ln 2$. We have treated the sky as if it were effectively flat, which is a reasonable approximation at small scales. There, we can make the substitution $kr \to (\ell + \frac{1}{2})\theta$.

For ground based instruments it is often useful to measure temperature differences in order to reduce contamination from atmospheric effects. In the past, both two and three beam difference experiments were performed. For example, a two beam experiment measures the variance of the temperature difference of two patches separated by an throw angle, θ_T. It is simple to show that

$$\langle (T_1 - T_2)^2 \rangle = 2C(0) - 2C(\theta_T) = \frac{1}{4\pi}\sum_\ell (2\ell+1)W_\ell(2 - 2P_\ell(\cos\theta_T))C_\ell.$$

Where W_ℓ is the single beam window function. A three beam experiment can be evaluated similarly.

10.6. **Future experiments.** A number of balloon based experiments are underway. Boomerang and Maxima have already provided important constraints on the position of the Doppler peaks. More will come from the BEAST and TopHat detectors, as well as from further Boomerang and Maxima flights.

The most important upcoming ground based detectors are interferometers such as the Very Small Array (VSA) based in Tenerife, DASI based at the south pole,

and CBI based in the mountains of Chile. These are all underway and all have reported results.

Two important satellite missions are planned or in progress. NASA's Microwave Anisotropy Probe (MAP) was launched in April 2001 and has produced its first results. The ESA Planck satellite is due to be launched in 2007. These will measure the whole microwave sky with resolution of approximately 0.15 degrees. Planck will have greater frequency coverage and should also provide useful measurements of the polarization.

11. The CMB frequency spectrum

The microwave background has a blackbody spectrum with a temperature of $T_0 = 2.728 \pm 0.004$. The COBE FIRAS instrument verified that the CMB spectrum was consistent with a blackbody within 50 parts per million in the millimeter wavelengths. Deviations from blackbody can be parameterized in many different ways, the most common being either the Compton $y-$parameter or in terms of a chemical potential, μ. The motivation for these parameterizations will be explained shortly. The FIRAS instrument placed upper limits on these parameters as $|y| < 1.5 \times 10^{-5}$ and $\mu < 9 \times 10^{-5}$.

The form that the spectral distortions are likely to take depend on when and how they are produced. Possible sources of distortions at very early times are inconsequential before a redshift of $z_{th} \sim 10^6$ since the universe remains in thermal equilibrium until this time via free-free and radiative Compton scattering, as these processes can change the number of photons. After this epoch, only ordinary Compton scattering is important, at least until a redshift $z_c \sim 10^4$. Since this process conserves photon number, the background remains in kinetic, but not thermal, equilibrium. While in kinetic equilibrium, injected energy can distort the spectrum. However, the spectrum has the form of a Bose-Einstein spectrum, parameterized by a chemical potential, μ.

The effect of energy injected after z_c depends on its production mechanism. There are a variety of possible sources, including the decay of massive particles or superconducting cosmic strings, stellar radiation, supernovae explosions, black hole accretion or evaporation, or by dust. Another important mechanism is inverse Compton scattering from hotter free electrons, which upscatters photons and redistributes them from lower frequencies to higher ones. This is observed to occur for clusters where it is known as the Sunyaev-Zeldovich effect, but it also could occur if the ionized inter-galactic medium is sufficiently hot. This leads to a Compton-y distortion of the spectrum.

12. Statistics and the CMB

There are a number of statistical challenges in going from observations of the CMB to understanding their implications for cosmological parameters. One often goes through many layers of data compression: from the raw time stream and pointing data, to maps at many different frequencies, to reconstructing a single map of the underlying CMB anisotropies, to its binned power spectrum and then finally to constraints on cosmological parameters. Each step comprises a difficult statistical problem.

Here I will discuss some of the general tools required to address these problems and many other problems in observational cosmology.

12.1. **Basic rules of probability.** The rules for manipulating probabilities are quite simple and intuitive. The first, the so called *sum rule* , requires the sum of the probabilities for the possible states of a variable be unity. That is,

$$\sum_{X_i} P(X_i|I) = 1,$$

where the sum is over all possible states of X. Here $P(X_i|I)$ corresponds to the probability that X in the state X_i given some background information I. This basically says that the variable must be in one of its possible states.

The second rule is known as the *product rule,* which states the probability of two things being true, $P(X, Y|I)$, is given by the probability of the first being true, $P(X|I)$, times the probability of the second is true *given that the first is true,* $P(Y|X, I)$. That is

$$\begin{aligned} P(X, Y|I) &= P(X|I)P(Y|X, I) \\ &= P(Y|I)P(X|Y, I), \end{aligned}$$

where the second relation follows from symmetry.

Written in this way, the product rule naturally leads into what is known as Bayes' Theorem:

$$P(X|Y, I) = \frac{P(Y|X, I)P(X|I)}{P(Y|I)}.$$

Usually this is written so that X represents the parameters of some model and Y represents some data or observation. $P(X|I)$ is called the *prior* distribution for the parameter X, while $P(X|Y, I)$ is known as its *posterior* distribution. The probability of the data given some model, $P(Y|X, I)$, is known as its *likelihood.* Finally the normalizing factor in the denominator, the probability of data $P(Y|I)$, is sometimes called the *evidence.* This is found by marginalizing the numerator over the p[possible theories,

$$P(Y|I) = \sum_{X_i} P(Y|X_i, I)P(X_i|I).$$

We often deal with continuous variables, where these probabilities are written instead in terms of a probability distribution function, e.g. the probability that the variable is in some range can be written as $P([X, X + dX]|I) = P(X|I)dX$. In this case, marginalizations like the one above can be written as

$$P(Y, I) = \int dX P(Y|X, I)P(X|I).$$

12.2. **Bayesian vs. Frequentist approaches.** There are two basic approaches to interpreting the probability distributions. In the Bayesian approach, probability distribution functions represent a degree of belief or plausibility. Bayes theorem then gives a mechanism of learning, or modifying your beliefs based on new data. The frequentist approach imagines that the observations are random draws from some ensemble of possible observations. Probability of a observation is then the long time frequency that it is observed with repeated trials.

Frequentists often criticize the Bayesian interpretation and priors as too vague and subjective, and they tend to focus primarily on the likelihood. Bayesians counter that frequentists are using priors implicitly, e.g. a flat prior, and they are just making their assumptions explicit.

In effect, Bayesians and frequentists are asking different questions. A Bayesian asks, 'how likely is a given parameter value given the data?', while a frequentist asks, 'how probable is the data, given certain parameters?' In certain circumstances, these approaches give the same results, particularly if the data at hand is very informative.

12.3. **Priors.** Since the biggest difference between the approaches is their treatment of priors, it is useful to discuss these here. One's priors are meant to reflect all previous information, excluding the data at hand. In physical situations, we generally have a theoretical expectation or some other weaker data which should determine the shape of ones priors.

In the absence of such observations and theoretical expectations, one is forced to make some choice of prior, generally based on the symmetry of the problem. Often there are some constraints which the probability distribution must obey.

Information theory suggests a way to choose a prior based on the principle of Maximum Entropy. One chooses a prior which maximizes the entropy of the distribution, where the entropy is defined as

$$S = -\int dx p(x) \log\left(\frac{p(x)}{m(x)}\right),$$

where $m(x)$ is the measure.

A number of possible distribution functions may be chosen:

- $P(X) \propto constant$ - the flat prior
- $P(X) \propto 1/X$ - the Jeffries prior
- $P(X) \propto e^{-(X-\overline{X})^2/2\sigma^2}$ - the Gaussian prior
- $P(N) = \frac{\mu^N}{N!}e^{-\mu}$ -the Poisson prior for discrete variables

The flat prior is usually used for location parameters, where we have no information to prefer one value over another. The Jeffries prior is basically uniform in the logarithm of the parameter and is used for scale parameters, where we have no information as to the over all scale of the parameter. Both the flat and the Jeffries priors are improper, as they diverge when integrated over their possible ranges. As a result, one needs to define upper and lower bounds for these distributions which requires some physical input.

The Gaussian and Poissonian priors are proper and integrable, but they require extra input. Either a variance σ^2 in the case of a Gaussian distribution, or a mean occupation $\mu = \langle N \rangle$ in the Poisson case. The flat prior may be taken as the limit of a Gaussian, with the variance going to infinity. The Gaussian is also the limit of a Poissonian as μ gets very large.

12.4. **Gaussian distributions.** In studying likelihood functions, it is useful to understand the multivariate Gaussian distribution, since observational errors are often taken to be Gaussian. In addition, non-Gaussian distributions can often be approximated near their peaks by a Gaussian approximation.

The most general multivariate Gaussian distribution for a n dimensional data vector $\mathbf{D}$ may be written as

$$P(\mathbf{D}|\mathbf{X}, I) = (2\pi)^{-\frac{n}{2}} |\det \widetilde{C}|^{-\frac{1}{2}} \exp\left[-\frac{1}{2}(\mathbf{D} - \overline{\mathbf{D}})^T \widetilde{C}^{-1} (\mathbf{D} - \overline{\mathbf{D}})\right].$$

Here $\overline{\mathbf{D}}$ is the mean value of the vector $\langle \mathbf{D} \rangle$, while $\widetilde{C}$ is its covariance matrix, $C_{ij} = \langle (D_i - \overline{D}_i)(D_j - \overline{D}_j) \rangle$. The brackets denote marginalization over the probability function, i.e. $\langle F(\mathbf{D}) \rangle \equiv \int d^n \mathbf{D} F(\mathbf{D}) P(\mathbf{D}|\mathbf{X}, I)$. In general, either the mean value of the data or its covariance matrix may be a function of the theory.

Even if the data are Gaussian distributed, the implied parameter distribution will not be unless the predictions are linear in the theory parameters. However, for the moment lets assume that the theory parameters are Gaussian distributed:

$$P(\mathbf{X}|\mathbf{D}, I) = (2\pi)^{-\frac{m}{2}} |\det \widetilde{T}|^{-\frac{1}{2}} \exp\left[-\frac{1}{2} (\mathbf{X} - \overline{\mathbf{X}})^T \widetilde{T}^{-1} (\mathbf{X} - \overline{\mathbf{X}}) \right].$$

This probability peaks for the parameters $\overline{\mathbf{X}}$, which is the maximum likelihood value.

The uncertainty in a parameter depends on whether one considers the other parameters as fixed or if one chooses to marginalize over them. If we hold the other parameters fixed, then the variance of a parameter X_i is given by $\sigma_i^2 = 1/(T^{-1})_{ii}$. On the other hand, if one decides to marginalize over the other parameters, then the variance grows due to the covariance between the data points. In the Gaussian case, this marginalization is performed simply by completing the square, yielding $\sigma_i^2 = T_{ii}$. These are related by $(T^{-1})_{ii} = 1/(T_{ii} - \widetilde{T}_{ij}(\widetilde{T}_{jk})^{-1}\widetilde{T}_{ki})$, where j, k run over all indices besides i. The errors can only increase when the marginalization is performed.

12.5. **Non-Gaussian distributions.** In general, the posterior distributions will likely be non-Gaussian. In this case, it may be difficult both to find the peak of the likelihood and to marginalize to find individual distributions. These operations usually must be performed numerically, either directly or by Monte Carlo methods. One popular Monte Carlo method, called the Markov Chain method, samples the likelihood surface via a random walk. Depending on the shape of the likelihood surface, such methods can save considerable computational time compared to direct integration.

It can still be useful to make a Gaussian approximation to the likelihood surface by expanding the log of the likelihood in a Taylor series,

$$\ln L(\mathbf{X} + \delta\mathbf{X}) = \ln L(\mathbf{X}) + \sum_i \frac{\partial \ln L(\mathbf{X})}{\partial X_i} \delta X_i + \frac{1}{2} \sum_{ij} \frac{\partial^2 \ln L(\mathbf{X})}{\partial X_i \partial X_j} \delta X_i \delta X_j + \cdots$$

Cutting off at the third term leads to a Gaussian approximation. The first term is effectively $-\frac{1}{2}\chi^2$ for the model. If we evaluate about the peak of the likelihood, then the linear terms are zero. Finally, $-\frac{\partial^2 \ln L(\mathbf{X})}{\partial X_i \partial X_j}$ is sometimes called the curvature matrix for the parameters or the Hessian. Flat directions of the curvature matrix correspond to parameter degeneracies.

Finally, let us define the Fisher matrix which is useful for making predictions for what a proposed experiment will tell us about the underlying theory. The Fisher matrix is simply the expectation of the curvature matrix,

$$F_{ij}(\mathbf{X}) = -\left\langle \frac{\partial^2 \ln L(\mathbf{X})}{\partial X_i \partial X_j} \right\rangle,$$

where the expectation denotes an averaging over data consistent with a given model. Thus, the Fisher matrix is defined for a particular theory and is independent of any actual data.

The Fisher matrix is useful because it allows us to place a lower bound on the derived parameter errors resulting from an experiment. In particular, if all the parameters are fixed but one, then its error is bounded by $\sigma_i^2 \geq 1/F_{ii}$. If we marginalize over the other parameters, $\sigma_i^2 \geq (F^{-1})_{ii}$. These bounds are the same as what we found earlier for the Gaussian case, where the bound is saturated. However, it must be realized that the Fisher matrix approach only gives lower bounds. Real experiments must deal with possible systematic errors, so the resulting error bounds will rarely achieve the predictions of the Fisher matrix approach.

12.6. **Comparing models.** While the peak of the likelihood provides a mechanism to choose between different models with the same parameterization, it is not adequate for choosing between models which have different variables or even different numbers of variables. For this, one usually looks at the ratio of posterior probabilities. Suppose you have two hypotheses, A and B, each with their own set of parameters (λ_A and λ_B) and you wish to discover which is most consistent with the priors and observations. Consider the ratio

$$R_{AB} = \frac{P(A|D,I)}{P(B|D,I)} = \frac{P(A|I)}{P(B|I)} \times \frac{P(D|A,I)}{P(D|B,I)}.$$

When R_{AB} is much larger than one, hypothesis A is favored, while hypothesis B is preferred if R_{AB} is small. If the ratio is of order 1, then one cannot make a clear preference. The first term on the right represents the ratio of ones prior beliefs in the two hypotheses, while the second represents the ratio of the evidences for each of the hypotheses. Recall that $P(D|A,I)$ is the likelihood of the data, marginalized over the possible values of the parameters λ_A.

It is helpful to see this in action on a very simple example. Assume that you measure some data and find $D = \overline{D} \pm \sigma$. You wish to compare two hypotheses: A which predicts that $D = 0$, and B which predicts that $D = \lambda$, where $\lambda_{min} > \lambda > \lambda_{max}$. Hypothesis A has no parameters, so the evidence is simply

$$P(D|A,I) = (2\pi)^{-\frac{1}{2}}\sigma^{-1}e^{-\overline{D}^2/2\sigma^2},$$

while the evidence for hypothesis B requires an integration over possible values of λ:

$$P(D|B,I) = \int_{\lambda_{min}}^{\lambda_{max}} d\lambda P(\lambda|B,I)(2\pi)^{-\frac{1}{2}}\sigma^{-1}e^{-(\lambda-\overline{D})^2/2\sigma^2}.$$

Let us assume that the prior for λ is flat over the range, so that $P(\lambda|B.I) = (\lambda_{max} - \lambda_{min})^{-1}$. Further, if we assume that the observation is well within the allowed range and that $\sigma \ll \lambda_{max} - \lambda_{min}$, then we can do the integration exactly to find $P(D|B,I) = P(\lambda = \overline{D}|B,I) = (\lambda_{max} - \lambda_{min})^{-1}$.

Thus we find

$$R_{AB} = \frac{P(A|I)}{P(B|I)} \times \frac{\lambda_{max} - \lambda_{min}}{(2\pi)^{\frac{1}{2}}\sigma} \times e^{-\overline{D}^2/2\sigma^2}.$$

Again, the first factor is the ratio of the priors, while the exponential factor is effectively the ratio of the best fit likelihood for Hypothesis A to that for the best fit model of Hypothesis B (i.e. $\lambda = \overline{D}$.) The second factor is sometimes called

Occam's factor, which effectively penalizes Hypothesis B because it has a larger parameter space. Given two models which fit the data equally well, this factor will tell us to prefer the simpler model, the one with a smaller parameter space. In this case, if the initial parameter space for Hypothesis B is small ($\lambda_{max} - \lambda_{min} \sim 20\sigma$), then Hypothesis A is preferred only if $\overline{D} < 2\sigma$. However, this cutoff could grow significantly if the B parameter space were increased exponentially.

References

[1] J.M. Bardeen, 'Cosmological Perturbations From Quantum Fluctuations To Large Scale Structure,' lectures given at 2nd Guo Shou-jing Summer School on Particle Physics and Cosmology, Nanjing, China, Jul 1988.

[2] J. R. Bond, in *Cosmology and Large-scale Structure, Proc. 60th Les Houches School,* eds. R. Schaeffer, et al., p. 469, Elsevier (1997).

[3] G. Efstathiou, in *Physics of the Early Universe, Proc. 36th Scottish Summer School*, eds. J. Peacock, et al., p. 361, Adam Hilger (1990).

[4] M. Hindmarsh & T. Kibble, *Rep. Progress. Phys.* **58**, 477 (1995).

[5] E. W. Kolb & M.S. Turner, *The Early Universe*, Addison-Wesley (1990).

[6] A. Liddle & D. Lyth, *Cosmological Inflation and Large Scale Structure*, Cambridge (2000).

[7] A. Linde, *Particle Physics and Inflationary Cosmology*, Harwood (1990).

[8] V. Mukhanov, H. Feldman & R. Brandenberger, *Phys. Reports* **205**, 203 (1992).

[9] T. Padmanabhan, *Structure Formation in the Universe*, Cambridge (1993).

[10] R. B. Partridge, *3K: The Cosmic Microwave Background Radiation*, Cambridge (1995).

[11] J. A. Peacock, *Cosmological Physics,* Cambridge (1999).

[12] P. J. E. Peebles, *The Large Scale Structure of the Universe,* Princeton (1980).

[13] P. J. E. Peebles, *Principles of Physical Cosmology*, Princeton (1993).

[14] D.S. Silvia, *Data Analysis: A Bayesian Tutorial*, Oxford University Press (1996).

[15] A. Vilenkin & E.P.S. Shellard, *Cosmic Defects*, Cambridge (1994).

Appendix A: Useful Relations of Special Functions

Spherical harmonic relations:

- Definition

$$Y_{lm}(\theta,\phi) = \left(\frac{2l+1}{4\pi}\frac{(l-m)!}{(l+m)!}\right)^{1/2} P_l^m(\cos\theta)e^{im\phi}$$

- Orthogonality -

$$\int Y_{\ell m}(\mathbf{n})Y^*_{\ell' m'}(\mathbf{n})d\Omega_{\mathbf{n}} = \delta_{\ell\ell'}\delta_{mm'}$$

- Completeness -

$$\sum_{\ell m} Y_{\ell m}(\mathbf{n})Y^*_{\ell m}(\mathbf{n}') = \delta(\phi-\phi')\delta(\cos\theta - \cos\theta')$$

- Plane wave expansion -

$$e^{i\mathbf{k}\cdot\mathbf{r}} = 4\pi\sum_{\ell m} i^\ell j_\ell(kr)Y_{\ell m}(\widehat{\mathbf{k}})Y^*_{\ell m}(\widehat{\mathbf{r}}) = \sum_\ell (2\ell+1)i^\ell j_\ell(kr)P_\ell(\cos\theta),$$

where $j_\ell(x)$ are the spherical Bessel functions.

Legendre relations (note $P_l(x) = P_l^0(x)$):

- Legendre equation

$$(1-x^2)\frac{d^2}{dx^2}P_l^m(x) - 2x\frac{d}{dx}P_l^m(x) + \left[l(l+1) - \frac{m^2}{1-x^2}\right]P_l^m(x) = 0$$

- Orthogonality
$$\int_{-1}^{1} P_l^m(x) P_{l'}^m(x) dx = \frac{2}{2l+1} \frac{(l+m)!}{(l-m)!} \delta_{ll'}$$
- Rodrigues' formula
$$P_l^m(x) = \frac{(-1)^m}{2^l l!} (x^2-1)^{m/2} \frac{d^{l+m}}{dx^{l+m}} (x^2-1)^l$$
- Derivative relation
$$(1-x^2) \frac{d}{dx} P_l^m(x) = (l-m+1) P_l^m(x) - (l+1) x P_l^m(x)$$
- Recurrence relation
$$(2l+1) x P_l^m(x) = (l-m+1) P_{l+1}^m(x) + (l+m) P_{l-1}^m(x)$$
- Relation to spherical harmonics
$$(2l+1) P_l(\hat{\mathbf{k}} \cdot \hat{\mathbf{r}}) = 4\pi \sum_m Y_{lm}(\hat{\mathbf{k}}) Y_{lm}^*(\hat{\mathbf{r}})$$

Bessel relations (useful for 2-d relations):

- Bessel equation
$$\frac{d^2}{dx^2} J_m(x) + \frac{1}{x} \frac{d}{dx} J_m(x) + \left(1 - \frac{m^2}{x^2}\right) J_m(x) = 0$$
- Orthogonality
$$\int k dk J_m(k\rho) J_m(k\rho') = \frac{1}{\rho} \delta(\rho - \rho')$$
- Plane wave expansion
$$e^{ik\rho\cos\phi} = \sum_m i^m e^{im\phi} J_m(k\rho)$$
- Integral representation
$$J_m(x) = \frac{1}{2\pi i^m} \int_0^{2\pi} d\phi e^{ix\cos\phi - im\phi}$$
- Derivative relation
$$\frac{d}{dx} J_m(x) = \frac{1}{2} \left[J_{m-1}(x) - J_{m+1}(x) \right]$$
- Recurrence relation
$$J_{m+1}(x) = \frac{2m}{x} J_m(x) - J_{m-1}(x)$$
- Spherical Bessel functions
$$j_m(x) \equiv \left(\frac{\pi}{2x}\right)^{1/2} J_{m+\frac{1}{2}}(x)$$

Appendix B: Scalar-Vector-Tensor Decomposition