SAPIENZA
UNIVERSITÀ DI ROMA

Facoltà di Scienze Matematiche, Fisiche e Naturali
Corso di Dottorato in Fisica

Cosmological Constraints on Light Particles

Relatore
Alessandro Melchiorri

Studente
Ludovico Mark Capparelli

XXXII Ciclo

Contents

Part I

Observations and open problems

The Large Hadron Collider has recently discovered[51, 54] and measured several properties of the Higgs Boson, completing the experimental verification of the Standard Model[7, 31, 50, 63, 64, 113, 118, 122, 212]. The Standard Model is perhaps the most successful predictive theory in the history of physics. The simplest extensions to the standard model, such as low scale supersymmetry, have been ruled out by the continuing ATLAS and CMS collaborations at the Large Hadron Collider[55, 56]. If it were not for cosmological and astrophysical observations, there would be little experimental data to cast doubt on the validity of the standard model.

The discovery of *dark matter*[24, 68, 224] and *dark energy*[58, 91, 96] are phenomena not explainable by the standard model and which have not been observed in any Earth based experiment. This is a strong indication that the theory is not a complete description of all interactions, with the exception of gravity. Cosmological observations are explained through the standard ΛCDM *model*[59]. It remains a theoretical challenge to link this model to fundamental particle physics. Moreover, other problems linking the theories have been identified. The origin of the matter and anti-matter asymmetry[79] is not known. The flatness and horizon problems are possibly hinting to a theory of *cosmic inflation*, generated by physics well out of reach of Earth based colliders[120].

Cosmological observations are becoming more abundant and precise, with claims that we are now in the era of *precision cosmology*[124]. No longer are we going to use fundamental physics to predict cosmological observables; we will use the precise data to probe fundamental physics. We have already done so, as evidenced by the discovery of dark matter and dark energy. The Cosmic Microwave Background radiation is sensitive to physics in the earliest moments of the universe, in a condition ofextremely high energies. It is sensitive to both physics at very high and very small energy scales.

1 Dark Matter

All the experimental evidence so far demonstrating the existence of Dark Matter involve its gravitational effects on baryonic matter and photons[10, 20, 23, 48, 58, 74, 92, 97, 209, 224]. These effects are ubiquitous, having been observed at many astrophysical scales, from galaxy rotation curves to clusters to the large scale structure of the universe. Consistently, there appears to be a large excess of non-interacting matter. This matter, dubbed dark, is stable over cosmological time scales, collisionless and non-relativistic, hence the use of the term *cold[129, 157].* Explanations from particle physics involve the proposal of *candidate particles*[33, 69, 81, 146, 182]. Many candidates have been proposed including, but not limited to, WIMPs, axions.

1.1 Astrophysical evidence

The most compelling evidence of the existence of an invisible but gravitationally interacting component of matter in galaxies is the flatness of galactic rotation curves*[68, 74]*. From Newtonian gravity it is known that a stars circular velocity around the center of a galaxy should be

$$v_c = \sqrt{\frac{GM(r)}{r}} \tag{1.1}$$

where $M(r)$ is the enclosed mass up until r. For a star far enough out of the galactic disk, $r \gtrsim R_{\text{disk}}$, we expect that mass $M(r) \simeq$ const and so that velocities of stars as a function of radial distance fall as $v_c \propto r^{-1/2}$. Instead, velocity measurements have consistently shown that, at large r, $v_c \simeq$ const, which implies a mass distribution $M(r) \simeq r$ at large distances. At those distances, typical baryonic matter (stars and dust) is absent. We refer to this extra component of matter as the *dark matter halo*. Stellar dynamics and N-body simulations can be used to show that the approximate shape of the halo is spherical[1]. This has a strong implication on the interactions of dark matter with itself or baryonic matter. Indeed, baryonic matter interacts strongly with itself and, with time, its distribution should flatten to the plane perpendicular to the total angular momentum vector. Baryonic matter tends to become shaped as disks. Since the dark matter halo is very weakly interacting, the halo remains much more spherical in shape.

The local dark matter density can be inferred from the velocity curves. At the location of the sun it is expected to be[9]

$$\rho_{DM} = 0.3 \frac{GeV}{cm^3} \tag{1.2}$$

We caution that this is an indirect measurement, averaging galactic measurements at our distance from the center, and may be affected by local peculiarities in the distribution of dark matter.

Evidence for the existence of large quantities of dark matter is found in clusters of galaxies. Historically, dark matter was discovered by Fritz Zwicky when analyzing the Coma Galaxy Cluster[224]. Applying the virial theorem to the distribution of galaxies

$$< \sum_i m_i v_i^2 > = < \sum_{ij} \frac{G m_i m_j}{2 r_{ij}^2} > \tag{1.3}$$

it is possible to measure the amount of mass present. Dividing by the number of galaxies, $N_{\text{galaxy}} \sim 1000$, yielded an average mass per galaxy of about $\sim 10^{11} M_\odot$, while the average luminosity is $\sim 10^8$ solar luminosities. This strongly implied that a large fraction of mass does not emit electromagnetic radiation. Although some mass is found in the form X-ray emitting intracluster gas[32], the rest is interpreted as the presence of a massive dark matter halo. The discovered X-ray emitting gas actually helps the case for dark matter. Indeed, the high measured temperatures of the gas must be explained by the presence of a gravitational potential well inside which the gas falls and heats up[105].

The total mass in clusters and galaxies can also be measured by astronomical lensing of distant objects[209]. The Sloan Digital Sky Survey used a statistical study of weak lensing, measuring the shear and distortion of distant images, to conclude that galaxies, including Milky Way, must have a dark matter halo extending at least an order of magnitude further

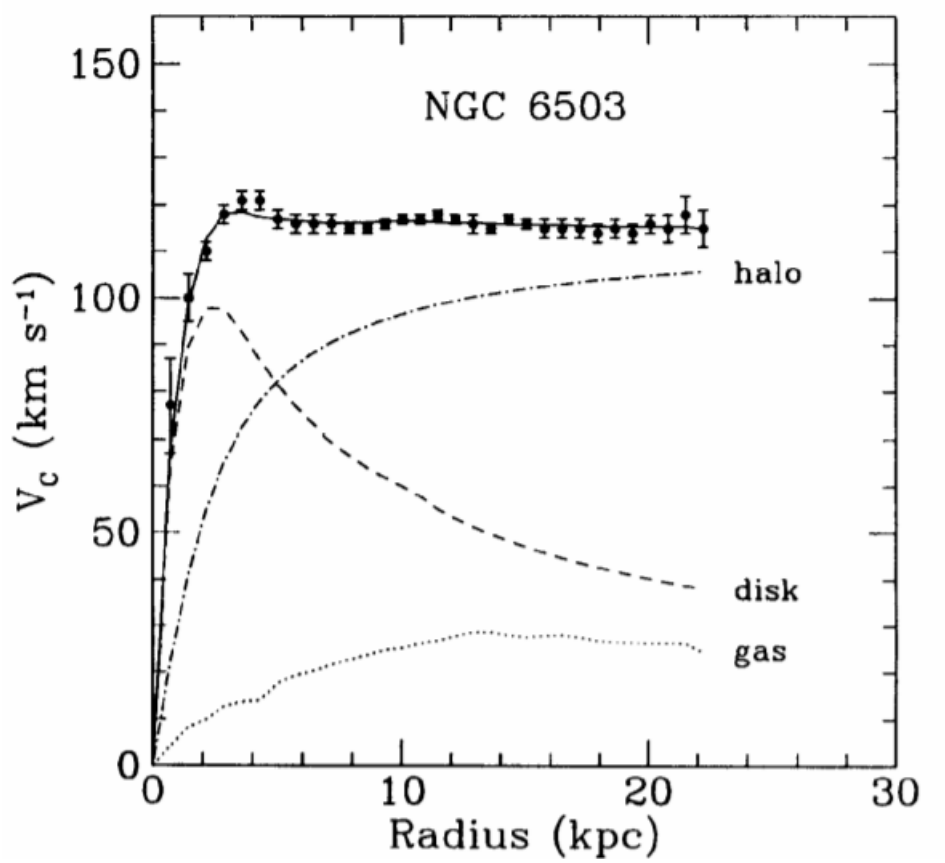

Figure 1.1: Typical rotation curve of a galaxy. Here NGC 6503. The contribution expected from visible matter in the disk and the interstellar gas is included, as well as the necessary contribution from the dark matter halo to fit the data. From ref [20].

than stellar matter[62].

From a cosmological perspective, the most compelling evidence for the existence of dark matter is the detection of old galaxies, at high redshifts $z \sim 10$[200, 125]. Cosmic Microwave Background radiation measurements show that at $z \sim 1100$ the matter over-densities, with respect to the universe averages, were $\delta\rho/\rho \sim 10^{-5}$[59, 92]. Matter over-density only grows during the matter-dominated epoch of the universe and in order for the structures to have formed by $z \sim 10$ the universe must have entered this epoch early enough[80, 129]. Increasing the early density of baryons would spoil measurements of the acoustic oscillations of the CMB and Big Bang Nucleosynthesis predictions[161]. The extra matter must negligibly interact non-gravitationally.

Studies of structure formation have also shown how the dark matter over-densities must be the seeds for the large scale structure we observe today. The dark matter halos form first, then attract baryonic matter onto their over-densities[80, 163].

1.2 Candidates

The most compelling candidates for a dark matter particle from theoretical fundamental physics are the axion[84, 182, 197] and the lightest supersymmetric particle[24] (LSP) from a theory of supersymmetry[99]. The latter can be referred to as a Weakly Interacting Mas-

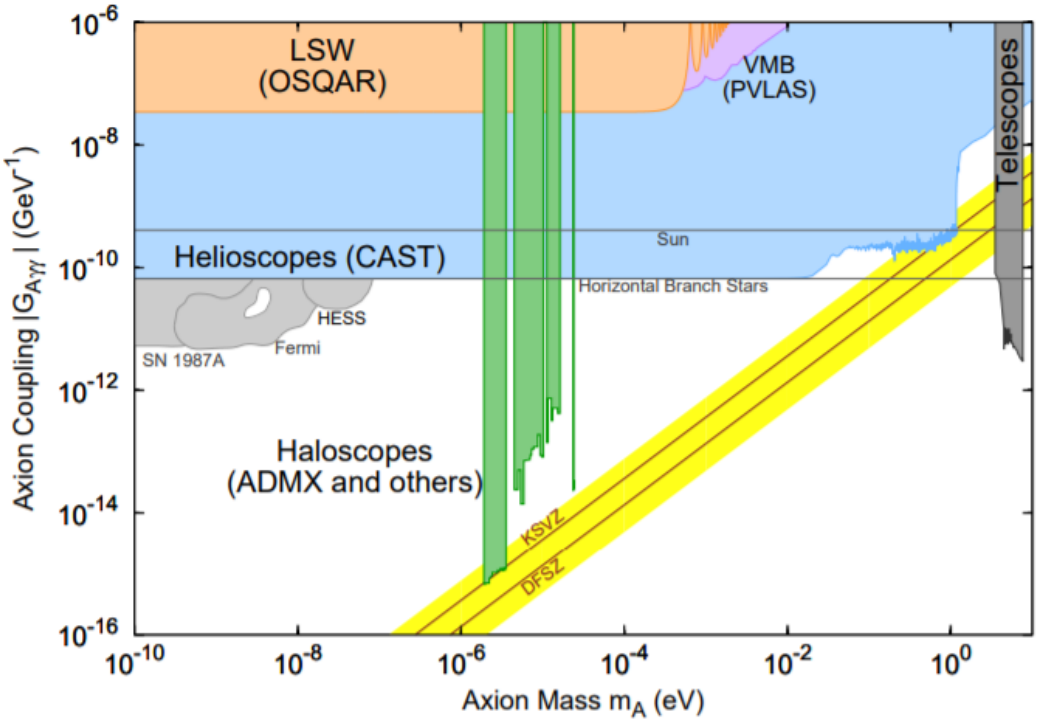

Figure 1.2: Exclusion plot from various axion searches. The yellow region is the parameter region for a QCD axion that would solve the strong CP problem. Axion masses of $m_a \sim 10^{-6} \div 10^{-5}$ are a cosmologically viable dark matter candidate. From ref [99].

sive Particle (WIMP). The existence of such particles was postulated outside of the context of astrophysics to solve different theoretical problems[11, 164, 165, 205, 206, 213, 217]. With time, it was realized that they make viable dark matter candidates as, in certain parameter ranges, they can be produced primordially with the correct abundances. Other possibilities have been proposed. Sterile neutrinos[30], MACHOs[179] and Q-balls[146], among others[24].

The idea for a Weakly Interacting Massive Particle (WIMP) was motivated by the fact that calculations of the relic abundance for a primordially produced massive particle $M \gtrsim 10GeV$ with a cross section typical of weak interactions would produce the required dark matter[143]. This was known as the WIMP "miracle". Recent direct detection experiments have ruled out these cross sections with ordinary matter[65], casting some doubt on the WIMP as a viable solution. Nonetheless, it remains a very promising candidate since a light and weakly interacting particle is naturally arising in a theory of supersymmetry. A theory of supersymmetry posits the existence of more massive *superpartners* of the usual standard model particles, with a spin differing by $\frac{1}{2}$[214]. In many supersymmetric theory a new multiplicative quantum number R is assigned to particles. $R = 1$ to the usual standard model particles and $R = -1$ to the superpartners. R-parity is a symmetry of the theory which means the total number is conserved during any interactions. With this idea, a cosmological picture emerges. Any population of superpartners which formed in the early universe has decayed to standard model particles *and to the LSP*, which is stable by the conservation of R-parity. There is no lower mass $R = -1$ particle. If the LSP has a negligible standard model cross section as well, it is a natural dark matter candidate.

Recent limits by the Large Hadron Collider on supersymmetric theories have constrained well motivated examples[55, 56].

The axion is the Nambu-Goldstone Boson of the broken Peccei-Quinn symmetry which was proposed to solve the QCD strong CP problem[164, 165, 213, 217]. The strong CP problem requires a solution on its own and so the axion may be an "economic" solution to both problems. It was realized that an axion may be produced non-thermally in the correct abundance if its mass and coupling strength were of the right values, $m_a \sim 10^{-6} eV$[84]. Such values are consistent with well motivated axion models[201]. Currently, the experiment ADMX is closing in on the very small coupling constant to the photon and may detect axion dark matter[49], or else severely constrain this possibility. Unlike WIMPs, axions have a very small mass. Pseudo-scalar bosons like the axions with much smaller masses may arise through symmetry breaking of many high energy symmetries. Indeed there is an abundance of axion particles predicted from string theory[204]. If the axion mass is very small, they would form a field with an astrophysical coherence scale[130]. This may alleviate the known cusp/core problem[156]. This is a discrepancy between measurements of dark matter halos which show a flat density profile at the center of galaxies, compared to N-body simulations of cold dark matter which predict a steeply rising profile, the cusp, at the center[73].

Sterile neutrinos with a keV mass may also solve the dark matter problem if they are produced non-thermally through mixing with standard model neutrinos[81]. The sterile neutrino should decay into a photon and a standard neutrino with a lifetime comparable to the age of the universe. The possible detection of a line at $3.5keV$ in the spectrum of galaxy clusters is tantalizing[34].

2 Dark Energy

First evidence that the expansion of the universe is accelerating rather than decelerating came from the distance vs redshift measurements of type 1a supernovae (SNe)[47, 91, 96, 154]. These measurements favor models of expansion of the universe with a very large cosmological constant $\Omega_\Lambda \sim 0.7$. Compared to the matter density, $\Omega_m \sim 0.3$, it seems most of the universe is made up by *dark energy[131].* This measurement implies we are at the early stage of a dark energy dominated expansion of the universe.

The dark energy problem refers to giving a fundamental physics explanation to the present value of this energy density. The cosmological constant was famously posited by Einstein in a static universe model but then disregarded[183]. However, as any interaction in particle physics, if there is no symmetry to eliminate it, it must there. Indeed it is[170]. At first, we may accept the cosmological constant as simply a constant of nature but, when quantum effects are considered, many problems arise.

The simplest explanation for Ω_Λ would be that of a vacuum energy. Indeed, the cosmological constant is none other that a term proportional to the metric itself in the Einstein equations[80]. In quantum mechanics, only the relative energies between two states are important[194]. We may add or subtract any value to make the energy of the vacuum zero. In general relativity, absolute values of the energy are important. The quantum mechanical zero point energy becomes relevant. Any calculation of the zero point energy

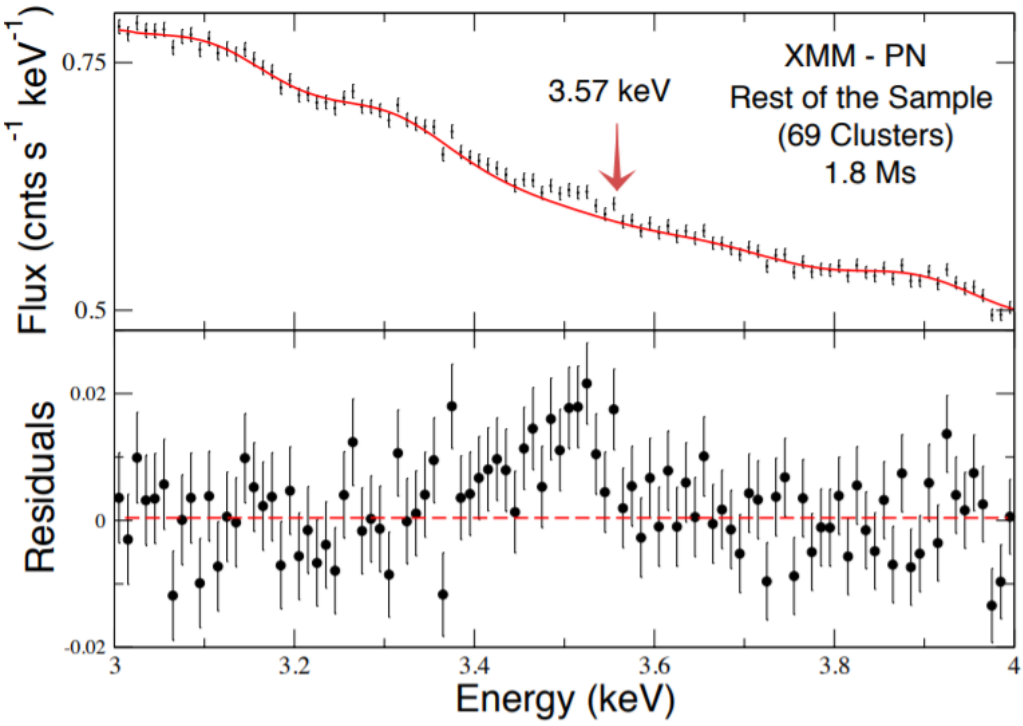

Figure 1.3: Unidentified emission line in the stacked X-Ray spectrum of galaxy cluster at $E = 3.57keV$. This may be the decay of a sterile neutrino into a photon and a lighter neutrino. Figure from ref [34]

requires taking into account renormalization effects, which drive up the value of the constant. Precise calculation yields an energy density which is about 10^{55} times larger than the measured value[195]. If we discard a miraculous fine-tuning between the bare parameter and the renormalized value, this is very wrong and implies some new physics at or below the Planck scale. It may that the solution can be found in a theory of quantum gravity. It has recently been conjectured that string theory is incompatible with a vacuum energy explanation[75].

Other explanations involve the existence of a very light scalar field, known as quintessence [130, 203]. A scalar field frozen in the minimum of potential, or which is still in the process of rolling down its potential, would be nearly indistinguishable from a vacuum energy. An example of such a scalar field is a Pseudo Nambu-Goldstone Boson, much like the axion, arising after some high energy symmetry is spontaneously broken.

3 The horizon and flatness problems

The Cosmic Microwave Background is nearly perfectly homogeneous[58, 92]. This homogeneity was the reason its detection conclusively proved the correctness of the Big Bang model, as opposed to steady state models of the universe[76, 168]. Today, the consistency of this homogeneity with the Big Bang is being called into question. Within the standard ΛCDM model, the CMB may only be uniform if the various locations of the universe at which it formed at $z \sim 1100$ had time to reach thermal equilibrium. A precise calculation, within the framework of ΛCDM, shows that not only could they not have thermalized, but

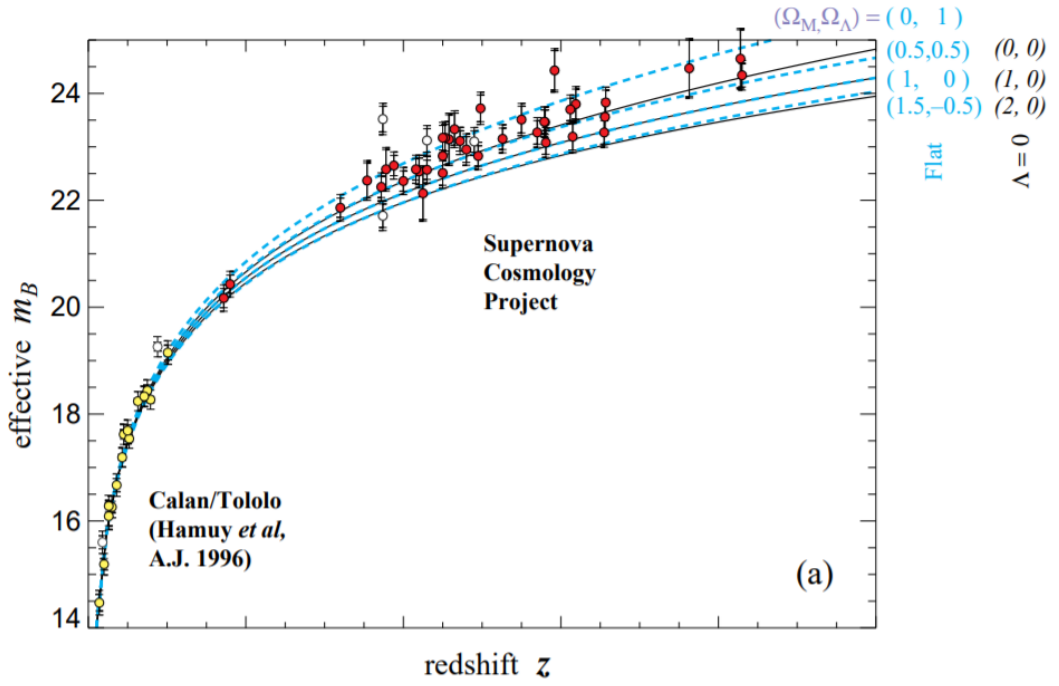

Figure 2.1: Hubble diagram (magnitude vs redshift) of Type Ia supernovae from the Supernova Cosmology Project, and 18 low-redshift Type Ia supernovae from the Calán/Tololo Supernova Survey. Curves for various cosmological models are shown. At high redshifts a non-zero cosmological constant is favored. Figure from ref. [96].

indeed there were regions which could not have been in causal contact with each other since the Big Bang. This is known as the horizon problem[66, 80]. How could very distant regions of the universe have the same primordial temperature if they had never been in causal contact? One may just argue by symmetry, as a cosmological principle, that this is the case. However, tiny perturbations $\delta T/T \sim 10^{-5}$ in the temperature of the CMB are measured across the sky[48, 59, 92]. Which means that the symmetry was broken. Furthermore, it is not even clear if, from a philosophical perspective, it is *scientific* to appeal to the initial conditions. This is known as the *horizon problem.*

An apparently unrelated problem is the flatness problem[80]. The universe today appears to be very close to flat, as indicated by measurements of the CMB anisotropies. In the dynamics of general relativity, a flat universe is an unstable configuration[171]. Any initial deviation of flatness would have grown quickly. If the universe is very close to being flat today, it must have been extremely so in the earliest moments. What caused the universe to be so flat ab-initio? Again, one may appeal to the initial conditions and to some "beautiful" symmetry which constrains the universe, amongst many possible geometries, to have a Euclidean one.

A third cosmological problem is the absence of magnetic monopoles[178]. Many Grand Unified Theories (GUTs) provide an extended gauge group, for example $SU(5)$, which spontaneously breaks into the standard model gauge group $SU(3) \otimes SU(2) \otimes U(1)$[110]. A by-product of the symmetry breaking is a copious production of magnetic monopoles in the early universe which would have obvious cosmological signatures. These signatures are not detected. Any GUT, which of course is still hypothetical at this stage, must be able

to explain the absence of a cosmological population of magnetic monopoles. We mention this problem because it is this context that solution to the three problems we have described was originally proposed.

A very early phase of *cosmological inflation* can solve all three problems[120]. This is the idea that, before the standard ΛCDM evolution of the universe, there was a very early and exponentially fast expansion[153]. To solve the above problems, this period should last until about $\sim 10^{-32}s$ after the initial singularity. This exponential expansion may solve the horizon problem as it may take regions of the universe which were very close together initially, and therefore in causal contact, and push them away at cosmologically relevant distances. Then, although they would no longer come in contact until billions of years later, they would have had the same common temperature anyways, having been in thermal equilibrium at the start[80]. The monopole problem is solved in the same way, so long as the temperature after inflation doesn't become that of a GUT scale, expected to be $E_{GUT} \sim 10^{16}GeV$. Any population of monopoles is diluted exponentially so that today we could expect to have $\lesssim 1$ magnetic monopole per observable universe. The flatness problem is also solved: during the exponential expansion a flat universe is actually a dynamical attractor of the equations of general relativity[171]. As a bonus, inflation may also give an explanation for the form of the initial perturbations of the universe around the cosmic background, solving yet another initial conditions problem.

There is no standard model mechanism to produce inflation. The most accredited theory is that inflation is produced by a scalar field which slowly rolls down its potential, having started from an equilibrium very far from its minimum[181]. While the *inflaton* is rolling, it acts as a vacuum energy in terms of coupling to gravity and produces the exponential expansion. Inflation ends when the field reaches the bottom of the potential, where it is assumed to decay into standard model particles. The only scalar field in the standard model would be the Higgs field, but its potential is not of the right form at inflation[36, 99].

4 The Hubble Tension

The Hubble constant H_0 describes the rate of expansion of the universe. It may be measured in two, very different and independent ways. The first is by the redshit versus distance of bodies in our cosmological neighborhood at low redshifts $z \lesssim 0.5$. This provides a direct measurement of H_0[43, 47, 154]. At the other end of the spectrum we may measure it by analyzing the anisotropies in the CMB, in particular by the angle subtended of the sound horizon at last scattering[80]. The value inferred is then dependent of the correctness of the underlying model, typically assumed to be ΛCDM.

The most precise local estimates of H_0 are obtained by measurement of distances two type 1a Supernovae (SNe). In this type of Supernova, the nova star is in a binary system. This produces a consistent peak luminosity which may used as a *standard candle*. Determination of the intrinsic luminosity requires the construction of a *distance ladder* up to local galaxies containing type 1a SNe. The current best constraint on the distance to local type 1a SNe is given by the observation of Cepheid variables in the host galaxies. It is understood that the systematic error in the construction of the distance ladder is a challenge in determination of the local value of H_0.

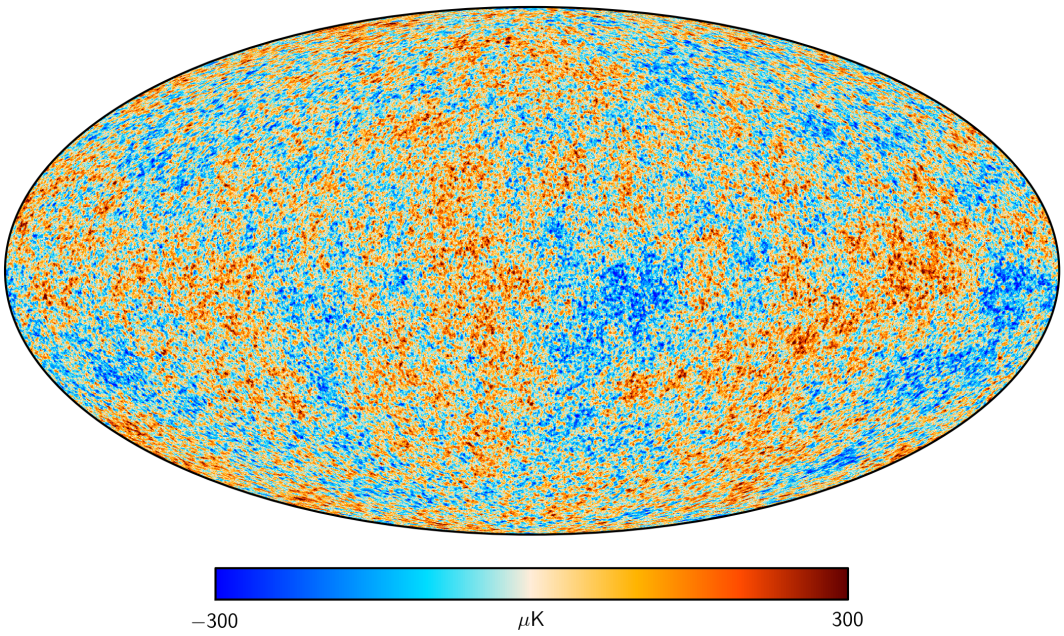

Figure 3.1: Reconstructed CMB temperature map from the PLANCK mission (2015). The anisotropies over the full sky are small $\delta T \sim 10^{-4} K$ compared to the uniform temperature $T = 2.7K$. The image is formed at the last scattering surface around $380,000$ years after the Big Bang. How can there be so much uniformity when different patches of the sky could never have been in causal contact before then? What genereates the primordial fluctuatuions which evolve with time to form this structure? A theory of cosmic inflation may be the answer. Figure from [57].

Assuming the systematic errors in constructing the distance ladder are under control, 1a SNe are measured at redshfits up until $z \simeq 0.15$. Any higher and the nonlinearities in the evolution of the universe, given by the complete cosmological model, would creep in. The maximum redshift must not be chosen too small, in order for the effect of peculiar velocities to be negligible.

The most recent measurement of the geometrical distance to the Large Magellanic Cloud (LMC) has allowed a reconstruction of H_0 from the SNe 1a redshifts[180]. This provides a value of

$$H_0 = 74.22 \pm 1.82 \; km \; s^{-1} \; Mpc^{-1} \tag{4.1}$$

On the opposite side of the universe, the CMB anisotropies as measured by the PLANCK collaboration are well fit by[61]

$$H_0 = 67.36 \pm 0.54 \; km \; s^{-1} \; Mpc^{-1} \tag{4.2}$$

Thus there is a $\sim 4\sigma$ discrepancy in the two observed values.

There is no commonly accepted solution to this discrepancy[104, 160]. Either there is some uncontrolled systematic, or there is new physics. Distance measurements rely on determination of geometric variables and local universe evolution. They are not affected by any change in the cosmological model. On the other hand the inferred values of the CMB are intrinsically dependent on the cosmological model one is using to fit the data. Recently it has been proposed that an *early dark energy* may push the expansion of the primordial

universe anomalously fast until the era of matter radiation equality. This would increase the inferred value of H_0 from CMB anisotropies, while being consistent with cosmological observables.

Part II

Homogeneous universe

5 Assumptions and principles

The Greek philosopher Anaximander (c. 610 - c.546 B.C.) formulated one of the first models of cosmology and of celestial mechanics which was independent of mythology[100]. One of the tenets of this theory was that the Earth did not "stand" on anything, there was no proverbial "stack of turtles", as referenced by B. Russell centuries later[184]. Rather, all directions in space where supposed to be equal and the Earth, being special, occupied a unique position in the universe: the center. Its movement was *forbidden* because that would *identify a special direction* in the universe.

In modern terms we would say this was a model of the universe with a *spherical symmetry* or, in other terms, an *isotropic universe*. Human thought has progressed much since ancient times and our understanding of the universe is more complete. Yet, physicists attempt to reduce the bulk of our knowledge to some elementary principles. The principles are often connected to some idea of symmetry, just as Anaximander proposed. Perhaps, the beauty of the physical nature is that symmetry is everywhere, once one looks for it.

In the centuries since Anaximander's models, we have come to the conclusion that *first principles must be inferred from experimental data*. This is seemingly a contradiction in terms: axioms are absolute. It is the nature of modern science that a well conducted and repeatable experiment is the basis for our model of the physical universe. What we may call first principles about the universe are in fact *hypotheses.* Luckily, nature appears to be logical, in that, at the very least, it seems to follow some form of the law of non-contradiction throughout space and time. This means that, from the first principles we inferred, we may proceed deductively to the prediction and, later, to empirical validation of new phenomena. Thus first principles, whether they are logical principles or not, are a *tool* to condense the vast experimental data which we acquire continuously.

In the study of cosmology there is one extra problem, compared to other branches of physics: we only have one universe on which to experiment. This can put in doubt the concept of *repeatability* of the experiment. We cannot set up a different universe, like we'd set up a different proton-proton collision, and analyze similarities and differences. Therefore we will observe features in our universe which may or may not arise out of necessity from the laws of physics, but may be due to some stochastic, or quantum, fluctuations of the initial conditions. Thus, there will also be a few *assumptions* we must make about what is just a *peculiarity of our universe* and what is signalling hitherto not understood physics.

The standard cosmological model relies on a few first principles/hypotheses and assumptions. [171]

1. *Gravitation is well described by General Relativity.* General relativity has been well measured on scales as large as the solar system, but it is difficult to obtain direct experimental evidence on galactic and extra-galactic scales. In fact, already on galactic scales there are problems which are generally ascribed to the presence of *dark matter*.

Since there is consistent evidence for dark matter, this is not usually seen as disproving general relativity, although one should certainly keep in mind the alternatives. General relativity is reliant on the *equivalence principle.* We may describe this as the existence of a locally flat frame at any space-time point, which would mean that the laws of physics, including those relating to non-gravitational interactions, can be extrapolated everywhere and to very early times, at least until quantum mechanics and general relativity don't begin to clash.

2. *Cosmological principle.* The cosmological principle is the assumption that the universe is homogeneous and isotropic. That is, it looks the same, in terms of *statistical distribution* of galaxies and radiation, as viewed from any point and there are no special directions. The cosmological principle may be assumed to hold for the part of the universe which is not, and may never be, observable. In this sense, it is a very strong assumption. An obvious, but very meaningful fact, must be pointed out. The observable universe *is not homogeneous* at small scales, where we find galaxies, stars and planets. Thus when speaking of homogeneity we are implicitly defining some length scale above which the universe is *uniform in a statistical sense.*

3. *Globally hyperbolic.* This last hypothesis involves the global topology of the universe. In the context of general relativity, given the cosmological principle, there are many possible pseudo-Riemannian 4-manifolds. A spacetime is globally hyperbolic if a hyper-surface, with an everywhere timelike normal vector, exists such that every timelike or null geodesic crosses the hyper-surface once and only once. Equivalently this means that a Cauchy surface exists, one where we can impose *initial conditions* and solve for the metric in all of space-time. In practice, this means we can slice space-time in hypersurfaces parametrized by a time parameter.

In addition, the overall evolution of the universe will depend on the properties of matter and their interactions. Historically, one would study these properties in a lab, or accelerator, on Earth and apply the discovered laws of physics to cosmology. In recent times astrophysical and cosmological observations have given the clearest hint of physics beyond the standard model. The dark matter and dark energy problems can be addressed by understanding how their properties affect the large scale structure of the universe. The universe itself can then be used as a laboratory to study the physics of these not well understood components.

6 The metric of the universe

6.1 Maximally symmetric spaces

Let's apply the principles described in section 5 to find a metric which may describe the universe at large. The assumption that space-time is globally hyperbolic implies we can foliate the space-time with hyper-surfaces Σ_t , parametrized by t, whose normal vectors are everywhere timelike. There are an infinite possible ways to foliate space-time, but we choose the manner in which each Σ_t is homogeneous and isotropic. Let's look at these

spatial hyper-surfaces. A space which is homogeneous and isotropic is also known as a *maximally symmetric* space[27].

Technically, a space is homogeneous if given any pair P, Q of points on Σ_t there exists an isometry taking P into Q. A transformation taking P into Q which does not change a metric. In general relativity, this is usually expressed with the existence of Killing vectors. A Killing vector field exists connecting P and Q such that if we move along its tangent lines the metric does not change.

The property of isotropy around a point $P \in \Sigma_t$ is taken to mean that a rotation of the vectors in the tangent space T_P does not alter the metric at that point. If we restrict ourselves to study a three dimensional spatial manifold we can conclude that the Riemann tensor, which encodes curvature, at some point x^p cannot depend on the direction we move in. The Riemann tensor must be invariant under rotations in the tangent space. The group of tangent space rotations is, of course, $SO(3)$. Let's work in a locally flat frame, where the metric is Euclidean. The only tensors which are invariant under $SO(3)$ are the flat metric η_{ij} and the Levi-Civita tensor ϵ_{ijk} . The most general Riemann tensor must be of the form[1]

$$R_{ijkl} = a\eta_{ij}\eta_{kl} + b\eta_{ik}\eta_{jl} + c\eta_{il}\eta_{jk} \tag{6.1}$$

The Riemann tensor is anti-symmetric under the exchanges $i \leftrightarrow j$ and $k \leftrightarrow l$. Because η_{ij} is symmetric, this implies that $a = 0$ and $b = -c \equiv k$. Since we have written a tensor equality, we can pass to a generic coordinate system. We conclude that in a maximally symmetric space the Riemann tensor is of the form

$$R_{ijkl} = k(x^p)(g_{ik}g_{jl} - g_{il}g_{jk}) \tag{6.2}$$

with g_{ij} being the *metric of this three-dimensional space*. For now, we are allowing k to depend on the point x_p. From this expression we deduce the Ricci tensor and scalar

$$R_{ij} = (n-1)k(x^p)g_{ij} \tag{6.3}$$

$$R = n(n-1)k(x^p) \tag{6.4}$$

where n is the number of dimensions we are working in. We have made this explicit since the above arguments can be repeated in any dimensionality. The Einstein tensor is

$$G_{ij} = k(x^\alpha)(n-1)(1 - \frac{n}{2})g_{ij} \tag{6.5}$$

Note that in $n = 2$ the Einstein tensor would be identically zero. This is a peculiarity of $2D$ manifolds and can be shown to be true regardless of the symmetry properties of the space. For $n > 2$ the Bianchi identities are meaningful and imply that the Einstein tensor is divergenceless

$$\nabla_i G^{ij} = 0 \tag{6.6}$$

[1]The same argument can be repeated in four dimensions, where the Levi-Civita tensor has four components and may appear as an extra term. However, it cannot contribute to the Riemann tensor due to it having different symmetry properties in its indexes.

Since g_{ij} is covariantly constant ($\nabla_i g_{jk} = 0$) and using the Leibnitz rule we obtain that

$$\frac{\partial}{\partial x^i} k(x^j) = 0 \tag{6.7}$$

Or that k is a constant. We could have guessed this by symmetry, assuming the space was homogeneous. We have proved it assuming only that it was everywhere isotropic. In this context the two ideas have the same implication. Thus we obtain the *constant curvature condition*

$$R_{ij} = k(n-1)g_{ij} \tag{6.8}$$

We now ask ourselves, in 3 dimensions, what spaces satisfy this condition. We can intuit that the metric of this space must be spherically symmetric, or that it can be sliced into 2-spheres, parametrized by "radial" coordinate r:

$$ds^2 = A(r)dr^2 + r^2 d\Omega^2 \tag{6.9}$$

where $d\Omega^2 = d\theta^2 + \sin^2\theta d\phi^2$. Obviously $A(r) = 1$ gives us the Euclidean space, which we already know is a solution. Are there others? The Ricci tensor components of such a space are

$$R_{rr} = \frac{A'}{rA} \tag{6.10}$$

$$R_{\theta\theta} = 1 - \frac{1}{A} + \frac{rA'}{2A^2} \tag{6.11}$$

$$R_{\phi\phi} = \sin^2\theta R_{\theta\theta} \tag{6.12}$$

where $A' = \frac{dA}{dr}$. Solving the constant curvature condition, from the rr equation we obtain

$$A' = 2krA^2 \tag{6.13}$$

which can be plugged in the $\theta\theta$ equation (the $\phi\phi$ equation being proportional) to arrive at an algebraic equation for A:

$$2kr^2 = 1 - \frac{1}{A} + kr^2 \tag{6.14}$$

which is solved by

$$A = \frac{1}{1-kr^2} \tag{6.15}$$

Thus a three-dimensional maximally symmetric metric is of the form

$$ds^2 = \frac{dr^2}{1-kr^2} + r^2 d\Omega^2 \tag{6.16}$$

The factor k is the curvature of the space, as seen in the Ricci Scalar (6.4).The shape of the manifold depends on the sign of k, while its absolute value is simply a change in scale. In fact, consider a metric with $\frac{k}{R^2}$ for some R. With a change of coordinates $r = \tilde{r} \cdot R$, the metric becomes

$$ds^2 = R^2\left(\frac{d\tilde{r}^2}{1-k\tilde{r}^2} + \tilde{r}^2 d\Omega^2\right) \tag{6.17}$$

so changing k is equivalent to a rescaling of the metric. So we reduce our discussion to considering the cases $k = -1, 0, +1$. The case $k = 0$ takes us to the flat, Euclidean, space.

In the case $k = 1$ we make the coordinate transformation

$$r = \sin\psi \tag{6.18}$$

to obtain the metric

$$ds^2 = d\psi^2 + \sin^2\psi d\theta^2 + \sin^2\psi \sin^2\theta d\phi^2 \tag{6.19}$$

which is the metric of a 3-sphere in polar coordinates. The fact that a 3-sphere is a maximally symmetric space was to be expected. We also note that in these coordinates there are no problems with the points at $\psi = \frac{\pi}{2}$, which correspond to $r = 1$, where the metric in the previous coordinates had a singularity. So the divergence at $r = 1$ is a coordinate singularity and not one of the space. This could have also been deduced by the fact that the Ricci scalar is smooth.

On the other hand the case $k = -1$ represents the hyperboloid. A more explicit metric can be obtained by the coordinate transformation

$$r = \sinh\psi \tag{6.20}$$

to arrive at the metric

$$ds^2 = d\psi^2 + \sinh^2\psi d\Omega^2 \tag{6.21}$$

A hyperboloid is a manifold of constant negative curvature.

6.2 Introducing time

Now that we have established the metric on each hypersurface Σ_t , we will deduce the full four-dimensional metric of the universe. The first realization is that the time-space components g_{0i} of the metric must be zero. If they were not, since g_{0i} transforms as a vector under rotations, they would identify a privileged direction violating the hypothesis of isotropy.

The space-space components of the metric g_{ij} must be that of a maximally symmetric space with the new feature that the curvature may now depend on the time coordinate t. As we have seen in (6.17), changing the curvature amounts to an overall scale changing of the metric. All the time dependence can be thus contained in this scale which we shall call the *scale factor* $a(t)$. Henceforth, we will also adopt the convention that $a(t)$ is a pure number, which means that the coordinate $\tilde{r}$ will be treated as a length and k is a constant parameter which has dimensions of inverse length squared. Other conventions are possible, of course.

Finally we consider the g_{00} component of the metric. This could be any smooth function of the time coordinate t, but it may always be eliminated and set to -1 by a transformation on t only. With these considerations we can finally write the metric of the universe given by the three principles in section 5[106].

$$ds^2 = -dt^2 + a^2(t)\left(\frac{dr^2}{1-kr^2} + r^2 d\Omega^2\right) \tag{6.22}$$

This is known as the Friedmann-Robertson-Walker metric (or simply FRW metric). The metric represents a universe whose maximally symmetric spatial slices expand or contract with time. The proper distance between any two points with fixed spatial coordinates must change with time proportionally to $a(t)$. The exact form of the scale factor must be determined by solving the Einstein equations, given the matter content of the universe.

We may recast the FRW metric in a more useful form by introducing *conformal time*

$$\tau = \int_0^t \frac{dt'}{a(t')} \tag{6.23}$$

so that $dt = a(t)d\tau$ and the metric

$$ds^2 = a^2(\tau)\left(-d\tau^2 + \frac{dr^2}{1-kr^2} + r^2 d\Omega^2\right) \tag{6.24}$$

For $k = 0$, the FRW metric is conformally flat. We will return to the significance of conformal time later.

It turns out that the t coordinate has a physical significance in terms of proper time of a certain class of observers. A *comoving observer* is one with fixed spatial coordinate values, thus its world-line is orthogonal to the spatial slices. These observers are also special in that they actually observe an isotropic universe. If they had a world-line with changing spatial coordinates they would see a boosted universe, with length contracted in one direction. Comoving observers must exist by hypothesis. Once we realize that $d\vec{x} = 0$, we realize that t is the proper time of such observers. *Coordinate time is the proper time of comoving observers.*

6.3 Christoffel symbols and curvature tensors

The Christoffel symbols, Ricci tensor, and the Ricci scalar computation is straightforward and not too illuminating so we just give the results here. We introduce the Hubble factor

$$H \equiv \frac{a'}{a} \tag{6.25}$$

where we use primes to denote derivatives with respect to time.

Using the metric (6.22), the non-zero Christoffel symbols are

$$\Gamma^t_{ij} = H g_{ij} \tag{6.26}$$

$$\Gamma^i_{tj} = \Gamma^i_{jt} = H\delta^i_j \tag{6.27}$$

$$\Gamma^r_{rr} = \frac{kr}{1-kr^2} \qquad \Gamma^r_{\theta\theta} = r(-1+kr^2) \qquad \Gamma^r_{\phi\phi} = r\sin^2\theta(-1+kr^2) \tag{6.28}$$

$$\Gamma^\theta_{r\theta} = \Gamma^\theta_{\theta r} = \Gamma^\phi_{\phi r} = \Gamma^\phi_{r\phi} = \frac{1}{r} \tag{6.29}$$

$$\Gamma^{\theta}_{\phi\phi} = -\sin\theta\cos\theta \qquad \Gamma^{\phi}_{\theta\phi} = \Gamma^{\phi}_{\phi\theta} = \cot\theta \tag{6.30}$$

The non-zero components of the Riemann tensor are

$$R^{i}_{tjt} = -\frac{a''}{a}\delta^{i}_{j} \tag{6.31}$$

$$R^{k}_{ilj} = (\frac{k}{a^2} + H^2)(\delta^{k}_{l}g_{ij} - \delta^{k}_{j}g_{il}) \tag{6.32}$$

The non-zero Ricci tensor components are

$$R_{tt} = -3\frac{a''}{a} \tag{6.33}$$

$$R_{ij} = g_{ij}(2\frac{k}{a^2} + 2H^2 + \frac{a^{''}}{a}) \tag{6.34}$$

The Ricci scalar is

$$R = 6(H^2 + \frac{a^{''}}{a} + \frac{k}{a^2}) \tag{6.35}$$

And finally the non-vanishing components of the Einstein tensor are

$$G_{tt} = 3(H^2 + \frac{k}{a^2}) \tag{6.36}$$

$$G_{ij} = -(H^2a^2 + 2a^{''}a + k)g_{ij} \tag{6.37}$$

The terms with upper and lower indexes can come in handy

$$G^{t}_{t} = -3(H^2 + \frac{k}{a^2}) \tag{6.38}$$

$$G^{i}_{j} = -\delta^{i}_{j}(H^2 + 2\frac{a''}{a} + \frac{k}{a^2}) \tag{6.39}$$

On the other hand it can be useful to express the same quantities in the metric written with conformal time (6.24). We denote derivatives with respect to conformal time $\frac{d}{d\tau}$ with dots and introduce the conformal Hubble factor

$$\mathcal{H} = \frac{\dot{a}}{a} = Ha \tag{6.40}$$

The Christoffel symbols are

$$\Gamma^{\tau}_{\tau\tau} = \mathcal{H} \tag{6.41}$$

$$\Gamma^{\tau}_{ij} = \frac{\mathcal{H}}{a^2}g_{ij} \tag{6.42}$$

$$\Gamma^i_{\tau j} = \Gamma^j_{\tau i} = \mathcal{H}\delta^i_j \tag{6.43}$$

with the Christoffel symbols with only spatial indexes being the same as above.

The non-zero components of the Ricci tensor are

$$R_{\tau\tau} = 3\mathcal{H}^2 - 3\frac{\ddot{a}}{a} \tag{6.44}$$

$$R_{ij} = \frac{g_{ij}}{a^2}(-\mathcal{H}^2 + 2k + 2\frac{\ddot{a}}{a}) \tag{6.45}$$

The Ricci scalar is

$$R = 6\frac{k}{a^2} + 6\frac{\ddot{a}}{a^3} \tag{6.46}$$

And the non-vanishing components of the Einstein tensor are

$$G_{\tau\tau} = 3(k + \mathcal{H}^2) \tag{6.47}$$

$$G_{ij} = \frac{g_{ij}}{a^2}(\mathcal{H}^2 - k - 2\frac{\ddot{a}}{a}) \tag{6.48}$$

And the upper-lower index terms are

$$G^\tau_\tau = -3(\frac{k}{a^2} + \frac{\mathcal{H}^2}{a^2}) \tag{6.49}$$

$$G^i_j = \delta^i_j(\frac{\mathcal{H}^2}{a^2} - \frac{k}{a^2} - 2\frac{\ddot{a}}{a^3}) \tag{6.50}$$

6.4 Conformal Lie vectors and redshift

By construction, the FRW metric has six spacelike Killing vectors which correspond to rotation and translation symmetry on the constant time surfaces. These can be tied to conserved quantities along the geodesic. Also of interest, is the fact the metric turns out to be conformal to other, time independent metrics, including the Minkowski space in the $k = 0$ case. With conformal time the metric can be written as (6.24), so

$$g_{\mu\nu} = a^2(\tau)\tilde{g}_{\mu\nu} \tag{6.51}$$

with $\tilde{g}_{\mu\nu}$ time-independent[27]. In a general case one considers two metrics conformal to each other

$$g_{\mu\nu} = e^{2\chi(x^\rho)}\tilde{g}_{\mu\nu} \tag{6.52}$$

where $\chi(x^\rho)$ is some function of space-time. Suppose C^μ is a Killing-vector of the metric $\tilde{g}_{\mu\nu}$. It satisfies the Killing equation

$$\tilde{\nabla}_\mu C_\nu + \tilde{\nabla}_\nu C_\mu = 0 \tag{6.53}$$

where $\tilde{\nabla}_\mu$ are the covariant derivatives with respect to the metric $\tilde{g}_{\mu\nu}$. We use the not explicitly covariant form of the Killing equation [2]

$$\tilde{g}_{\mu\nu,\rho}C^\rho + \tilde{g}_{\rho\nu}C^\rho_{,\mu} + \tilde{g}_{\mu\rho}C^\rho_{,\nu} = 0 \tag{6.54}$$

Then since

$$\tilde{g}_{\mu\nu,\rho} = e^{-2\chi}g_{\mu\nu,\rho} - 2\chi_{,\rho}g_{\mu\nu}e^{-2\chi} \tag{6.55}$$

we obtain

$$g_{\mu\nu,\rho}C^\rho + g_{\rho\nu}C^\rho_{,\mu} + g_{\mu\rho}C^\rho_{,\nu} = L_C g_{\mu\nu} = 2\chi_{,\rho}C^\rho g_{\mu\nu} \tag{6.56}$$

where L_C is the Lie-Derivative with respect to the vector field C. Thus, if C^μ is a Killing vector of the conformally related metric, then it satisfies the *conformal Killing equation* for the new metric:

$$\nabla_\mu C_\nu + \nabla_\nu C_\mu = 2\omega(x^\rho)g_{\mu\nu} \tag{6.57}$$

In this case $\omega(x^\rho) = 2\chi_{,\rho}C^\rho$. In the FRW metric $\chi = \ln a(\tau)$. $C^\mu = \delta^\mu_0$ is obviously a Killing vector of $\tilde{g}_{\mu\nu}$. We could verify explicitly by plugging in the Christoffel symbols that C^μ satisfies the above equation. In the Minkowski case, $k = 0$, it is the generator of time translations, related to conservation of energy. The presence of a Killing vector field implies a conserved quantity along a geodesic:

$$Q_C = C_\mu \dot{x}^\mu \tag{6.58}$$

where the dot represents a derivative with respect to the affine parameter of the curve. To prove this is conserved, take the derivative of this quantity along the affine parameter λ is

$$\frac{dQ_C}{d\lambda} = \dot{C}_\mu\dot{x}^\mu + C_\mu\ddot{x}^\mu = C_{\mu,\nu}\dot{x}^\mu\dot{x}^\nu + C_\mu\ddot{x}^\mu \tag{6.59}$$

where we have used the chain rule in the first term on the RHS. Using the geodesic equation

$$\begin{aligned}\frac{dQ_C}{d\lambda} &= C_{\mu,\nu}\dot{x}^\mu\dot{x}^\nu - C_\mu\Gamma^\mu_{\nu\rho}\dot{x}^\nu\dot{x}^\rho \\ &= \frac{1}{2}(\nabla_\mu C_\nu + \nabla_\nu C_\mu)\dot{x}^\mu\dot{x}^\nu\end{aligned} \tag{6.60}$$

where we used the symmetry of $\dot{x}^\mu\dot{x}^\nu$ on the second line. Now we can use (6.57). Immediately, we notice that if $\omega = 0$, so that C^μ is a Killing vector, the quantity Q_C is conserved. In the case $\omega \neq 0$ we can write

$$\frac{dQ_C}{d\lambda} = 2\omega g_{\mu\nu}\dot{x}^\mu\dot{x}^\nu \tag{6.61}$$

so Q_C is not generically conserved along the equations of motion. *It is however conserved along null, or light-like, geodesics* $g_{\mu\nu}\dot{x}^\mu\dot{x}^\nu = 0$.

This fact leads to an interesting results for cosmology. In this case $Q_C = C_\mu\dot{x}^\mu = -a^2\dot{\tau}$, so that $\dot{\tau} \propto a^{-2}$. By changing back to regular time, $\dot{t} = a\dot{\tau}$, and noticing that $\dot{t}$ is proportional

[2] To get this, pass to the equivalent equation $g_{\rho\nu}\nabla_\mu C^\rho + g_{\rho\mu}\nabla_\nu C^\rho = 0$ and write out explicitly all the Christoffel symbols. After the contractions there are some cancellations.

to the energy E of the particle, we obtain that

$$E \propto a^{-1} \tag{6.62}$$

for lightlike geodesics. For photons in particular $E = \frac{hc}{\lambda}$, λ being the wavelength, we obtain that

$$\lambda \propto a \tag{6.63}$$

This result shows that the wavelength of a photon is redshifted with the expanding universe, a fact which is the basis for observational cosmology.

There are other ways to derive this result, and certainly this method may seem more obscure, but we wish to present this both to remark on the usefulness of conformal Killing vectors and to demonstrate the connection of this well known phenomenon to the underlying symmetries of the metric.

The redshift factor z is defined by the relation

$$1 + z \equiv \frac{\lambda(\text{observation})}{\lambda(\text{emission})} = \frac{a(\text{observation})}{a(\text{emission})} \tag{6.64}$$

7 Dynamics of the universe

7.1 Energy-momentum tensor from a phase space distribution

In general relativity, the energy momentum tensor encodes the energy density, pressure and stresses of the matter content throughout spacetime. Understanding its form is as necessary as understanding the form of the metric. In cosmology we will often deal with statistical mechanics and the properties of matter as encoded in a distribution function. Thus it seems appropriate from the start by relating the energy-momentum tensor to the underlying probability density function of matter.

We recall that the action for a free massive particle in curved spacetime is given by

$$S = -m \int ds = -m \int \sqrt{-g_{\mu\nu} dx^\mu dx^\nu} \tag{7.1}$$

where s is the proper time of the particle. The four-momenta is $p^\mu = m\dot{x}^\mu$. In a Hamiltonian formulation the variables p_μ, with a lower covariant index, are the conjugate momenta to x^μ, since $p_\mu = \frac{\partial L}{\partial \dot{x}^\mu}$, where the dot indicates a derivative with respect to the proper time. In special relativity it is not really important if we study momenta with lower or upper indexes, but in general relativity it makes a difference. By defining phase space as the space through conjugate momenta, we can be sure that the usual results in statistical mechanics, for example Liouville's theorem, hold: the volume occupied by nearby trajectories in phase space is constant.

Given a point in space x^μ and a momentum p_μ we define the distribution function through the number of particles contained within the infinitesimal element of phase space $d^3x = dx^1 dx^2 dx^3$, $d_3p = dp_1 dp_2 dp_3$

$$dN = f(x^0, x^i, p_j)\frac{d^3x d_3p}{(2\pi)^3} \tag{7.2}$$

The factor $(2\pi)^3$, actually $(2\pi\hbar)^3$, is the usual size of a fundamental element of phase space. We will adopt the convention to denote the volume element around covariant components with the number of dimension indicated at the bottom. dN is by no means an invariant quantity: under coordinate changes it must transform in the same way as the product $d^3x d_3p$, since the total number of particles is a scalar. We deduce that the distribution function must be a scalar under coordinate changes.

The energy-momentum tensor can be written as [155]

$$T^{\mu\nu} = \int \frac{d_3p}{(2\pi)^3}\frac{1}{\sqrt{-g}}\frac{p^\mu p^\nu}{p^0} f(x^0, x^i, p_j) \tag{7.3}$$

Note the momenta which appear explicitly have upper indexes, while the integration momenta has lower indexes. Let's check that this expression has the right properties for being the energy-momentum tensor. Working in a locally inertial frame at some space-time point we note that for $\mu = \nu = 0$ we simply obtain the average of the energy over the momenta, which corresponds to the energy density. This, appropriately, is the interpretation of the time-time component of the energy momentum tensor. Similarly, by looking at the spatial components we find the usual expressions for the pressure of a gas.

$$T^i_i = 3P = \int \frac{d_3p}{(2\pi)^3}\frac{p^i p_i}{p^0} f(x^0, x^i, p_j) \tag{7.4}$$

The factor $(\sqrt{-g})^{-1}$ appears when not working a locally inertial frame. In fact, by transforming between coordinate systems we note that $d_3p\frac{1}{\sqrt{-g}}\frac{1}{2p^0}$ is an invariant measure. This can be shown by noticing that

$$\begin{aligned}
\delta(p_\mu p^\mu + m^2)\theta(p^0) &= \delta(g^{00}p_0^2 + 2g^{0i}p_0p_i + p^2 + m^2)\theta(p^0) \\
&= \frac{\delta(p_0 - p_{0+})}{|2g^{00}p_0 + 2g^{0i}p_i|} \\
&= \frac{1}{|2p^0|}\delta(p_0 - p_{0+})
\end{aligned}$$

where we have intermediately used $p^2 = g^{ij}p_ip_j$. p_{0+} is one of the solutions of the Dirac delta's argument:

$$p_{0\pm} = -\frac{g^{0i}p_i}{g^{00}} \pm \frac{1}{g^{00}}\sqrt{(g^{0i}p_i)^2 - 4g^{00}(p^2 + m^2)} \tag{7.5}$$

and we can drop the term with $\delta(p_0 - p_{0-})$ because of the Heavyside theta function on p^0 (with an upper index). In fact it can be readily seen that

$$p^{0\pm} = g^{00}p_{0\pm} + g^{0i}p_i = \pm\sqrt{(g^{0i}p_i)^2 - g^{00}(p^2 + m^2)} \tag{7.6}$$

Using this result we finally get that

$$d_3p\frac{1}{\sqrt{-g}}\frac{1}{2p^0} = d_4p\frac{1}{\sqrt{-g}}\delta(p_\mu p^\mu + m^2)\theta(p^0) \tag{7.7}$$

The measure $\frac{d_4p}{\sqrt{-g}}$ is invariant, the argument of the Dirac delta is scalar and so the Dirac delta is invariant as well. What remains to be seen is if the theta function is also a scalar. Either the coordinate transformation changes the sign of p^0 or it does not. In case it doesn't there is no problem, as the above derivation holds the same. If it changes sign there is also no problem, when multiplying by $\delta(p_\mu p^\mu + m^2)$. In fact if we were to have $\theta(-p^0)$ we would select the solution p_{0-} of the argument, but $|p^{0-}| = |p^{0+}|$ so there is no problem again.

Therefore we have shown that under a coordinate transformation the energy-momentum tensor (7.3) transforms only as the four-momenta $p^\mu p^\nu$ carrying free indexes in the integral.

7.2 Energy-Momentum tensor from the action

For more complicated situations, such as in the presence of a scalar or electromagnetic field, it is not straightforward to know the form of the energy-momentum tensor. Indeed the Einstein equations, through Bianchi's identities, require $T^{\mu\nu}$ to be covariantly conserved:

$$\nabla_\mu T^{\mu\nu} = 0 \tag{7.8}$$

Of course, such a conserved tensor can be built from a Lagrangian via Noether's theorem, assuming the Lagrangian itself is translation invariant. This construction of the energy-momentum tensor is known as the *canonical* energy momentum tensor. Unfortunately it is, in the general case, not suitable for general relativity. The problem arises from the fact that the tensor constructed with Noether's theorem is usually not symmetric, whereas Einstein's tensor $G_{\mu\nu}$ is. This is not an impassable hurdle. Starting from the canonical energy-momentum tensor, one can apply the *Belinfante-Rosenfeld* procedure. We will not go into it here, but it involves adding divergenceless quantities to the tensor in order to make it symmetric while keeping it conserved.

A more straightforward way of obtaining the correct energy-momentum tensor is via an action principle. The purely gravitational part of the action is the Einstein-Hilbert action

$$S_{EH} = \frac{1}{16\pi G}\int R\sqrt{-g}d^4x \tag{7.9}$$

where R is, of course, the Ricci scalar. The gravitational part of the Einstein equations, the Einstein tensor $G_{\mu\nu}$ term, is found by extremizing this action with respect to $g_{\mu\nu}$, the fundamental degree of freedom of general relativity. If this is the only action term we will obtain the Einstein equations in a vacuum. However we live in a universe with matter, so we add the action of matter there will be extra terms involving the $g_{\mu\nu}$

$$S_M(\text{Matter fields}, g_{\mu\nu}) = \int d^4x\sqrt{-g}\mathcal{L}_M \tag{7.10}$$

The variation of these terms with respect to the metric will give us the right hand side of

the Einstein equations, namely the term proportional to the energy-momentum tensor.

By deriving the Einstein-Hilbert action with respect to the metric it can be shown that

$$\partial S_{EH} = \frac{1}{16\pi G}\int \sqrt{-g}G_{\mu\nu}\delta g^{\mu\nu}d^4x \tag{7.11}$$

where $\delta g^{\mu\nu}$ is the variation of the metric and $G_{\mu\nu}$is the Einstein Tensor. When one has the action $S = S_{EH} + S_M$, the total variation would be

$$\partial S = \int d^4x\delta g^{\mu\nu}\left(\frac{\sqrt{-g}}{16\pi G}G_{\mu\nu} + \frac{\partial(\sqrt{-g}\mathcal{L}_M)}{\partial g^{\mu\nu}}\right) \tag{7.12}$$

So in order to get the Einstein equations it is clear that the energy-momentum tensor of the matter fields must be

$$T_{\mu\nu} = \frac{-2}{\sqrt{-g}}\frac{\partial(\sqrt{-g}\mathcal{L}_M)}{\partial g^{\mu\nu}} \tag{7.13}$$

By definition, this is the part of the energy and momentum which couples to gravity. It can be shown that this is the same tensor that can be obtained via Noether's theorem and then the Belinfante-Rosenfeld procedure. Compared to that, it has the advantage of being straightforward to calculate.

7.3 Energy-momentum tensor of a perfect fluid

We have seen that understanding the homogeneous universe reduces to determining the form of the scale factor $a(t)$. This can be done by solving the Einstein equations, but first we must specify the matter content of the universe. This is done by specifying the form of the energy-momentum tensor $T_{\mu\nu}$. This must be consistent with the symmetry properties of the metric. By looking at the explicit form of the Einstein tensor for a FRW metric (6.36), and (6.37), and recalling that $T_{\mu\nu}$ is proportional to $G_{\mu\nu}$ via the Einstein equations, we intuit that the only possible form for the energy momentum tensor is that of a perfect fluid[80, 171, 159]:

$$T_{00} = \rho(t) \tag{7.14}$$

$$T_{0i} = 0 \tag{7.15}$$

$$T_{ij} = g_{ij}P(t) \tag{7.16}$$

where $\rho(t)$ is the proper density of the fluid and $P(t)$ is the pressure in a frame locally at rest with it. They are scalars by construction. By the assumptions presented in section 5, this is the only consistent possibility for the matter content of a homogeneous universe. What remains to be specified is the equation of state, or the relationship between ρ and P.

We can write the energy-momentum tensor in a more explicitly covariant form by introducing the four-velocity $u^\mu = (1, 0, 0, 0)$ of a comoving observer.

$$T_{\mu\nu} = u_\mu u_\nu(\rho + P) + g_{\mu\nu}P \tag{7.17}$$

Useful formulas for the perfect fluid tensors are with upper/lower indexes

$$T_0^0 = -\rho \tag{7.18}$$

$$T_j^i = \delta_j^i P \tag{7.19}$$

so the tensor is diagonal. We can and should ask ourselves how the conservation equation relates the pressure and density. We take the time component of the conservation equation, since for a constant u^μ and homogeneous P, ρ the others are trivially satisfied,

$$\nabla_\mu T_0^\mu = 0 \tag{7.20}$$

From this we simply obtain the continuity equation

$$\rho' + 3H(\rho + P) = 0 \tag{7.21}$$

which can be rearranged, by recalling that $H = \frac{a'}{a}$ as

$$\frac{d}{dt}(\rho a^3) = -3Ha^3 P = -P\frac{d}{dt}a^3 \tag{7.22}$$

This equation is very simple to understand. In fact, considering that any volume V of space expands as a^3, we may simply think as a^3 as volume of space. Therefore ρa^3 is the internal energy of the fluid in such a volume, which means this equations is none other than the first law of thermodynamics $dU = -PdV$. We can note that there is no heat flow which changes the internal energy. This was implicit in our hypothesis for the form of the energy-momentum tensor. Heat flow is in fact encoded in T^{0i} as the energy per unit-time and surface which flows through the surface orthogonal to x^i. We can then understand at this stage that the *evolution of a perfect fluid is adiabatic*. We will explore this idea later, when we discuss the existence of different components of the fluid.

As we pointed out, we must specify the equation of state. This is usually given in the form $P = P(\rho)$. If we have more than one specie of particles, each specie will have its own equation of state. In a cosmological setting we usually take the equation of state to be

$$P = w\rho \tag{7.23}$$

for some parameter w. Usually w is considered to be fixed, but there is no a-priori reason for this to be. The content of the universe is normally separated in non-relativistic species which we generally refer to as *matter*, and relativistic species which we usually refer to as *radiation*. The pressure of matter is proportional to the mean velocity $\frac{v}{c}$, which is small by assumption. Therefore, for matter we have $w_{\text{matter}} \simeq 0$. For radiation, regardless of the specie being a boson or a fermion, we shall see that $w_{\text{radiation}} = \frac{1}{3}$.

During the relevant history of the universe electrons and baryons are assumed to be non-relativistic. Photons are, of course, always relativistic and we shall also consider neutrinos to be always relativistic, even if their small but unknown mass may have an effect in the late universe. In addition to these standard model components, the universe also contains

dark matter and *dark energy*. The properties of these components are deduced from astrophysical and cosmological observations. Dark matter is thought to be *cold*, non-relativistic, so we classify it as matter. Dark energy is on the other hand very odd, in that it is a component completely alien to anything else known. A cosmological constant may explain dark energy, but it may also be some component, such as a scalar field which mimics a cosmological constant. The crucial fact of a cosmological constant is that its density does not dilute with the expansion of the universe. In order for a component to have this property we will shortly see that $w_{\text{dark energy}} = -1$. Dark energy has negative pressure!

Plugging the equation of state (7.23) in the continuity equation (7.21) and integrating one finds that

$$\rho \propto a^{-3(1+w)} \tag{7.24}$$

Thus, for matter the density decreases with the inverse cube of the scale factor. Having understood that volumes scale with a^3, the meaning of this dilution of matter is simply that the number of particles, therefore the total energy $E \sim mN$, remains constant, while the volume increases.

For radiation the density decreases faster than the volume of the universe, as a^{-4}. The extra a factor can be understood from the energy loss of every single photon due to cosmological redshift (6.62).

Finally, we note that indeed for $w = -1$ the energy density remains constant as the universe expands.

7.4 Friedmann Equations

We can now write the equations which govern the evolution of the scale factor $a(t)$. Assuming the metric form (6.22), which has Einstein tensors (6.36) and (6.37), in presence of a perfect fluid (7.17) we write explicitly the Einstein equations with a cosmological constant. The time-time equation gives

$$H^2 = \frac{8\pi G}{3}\rho + \frac{\Lambda}{3} - \frac{k}{a^2} \tag{7.25}$$

which is known as the *First Friedmann equation*. The space-space equations are all the equivalent. After using the first equation in place of H^2 , we obtain the second Friedmann equation

$$\frac{a''}{a} = -\frac{4\pi G}{3}(\rho + 3P) + \frac{\Lambda}{3} \tag{7.26}$$

The first equation determines how the rate of expansion depends on the matter content of the universe, as the Hubble factor contains the derivative of the scale factor. The second equation governs the acceleration, or deceleration, of the expansion. We note that if $\Lambda = 0$, for regular matter $P \geq 0$ and the acceleration must be negative. The fact that experimental data indicates the expansion is accelerating implies either that the cosmological constant is non-zero, or there is a large component of the universe for which $P < 0$.

Since the first equation (7.25) is a first order equation in a, we may expect the second equation (7.26) to be redundant. Indeed by deriving the first equation with respect to time, and using the same to substitute an H^2 term, we obtain

$$\frac{a''}{a} = \frac{4\pi G}{3}(\frac{\rho'}{H} + 2\rho) + \frac{\Lambda}{3} \tag{7.27}$$

Now we can plug in the continuity equation for a perfect fluid (7.21) and find the second equation again.

We can understand that given the variables a, ρ and P there can be at most three independent equations to solve. We can choose two between the Friedmann equations and the continuity equations and add an equation of state $P(\rho)$ for the fluid. Usually, when solving the problem numerically, one chooses the first Friedmann equation and the continuity equation.

Of course, until now we are have only considered a single component of matter. In the presence of many fluids, each with densities ρ_i we may write

$$H^2 = \frac{8\pi G}{3}\sum_i \rho_i + \frac{\Lambda}{3} - \frac{k}{a^2} \tag{7.28}$$

where we assumed that the total density is equal to the sum of the densities of every single species, meaning that interaction energy is negligible. Assuming the universe is in chemical equilibrium, the continuity equation holds *separately for every specie*.

It is useful to see the different forms in which the first Friedmann equation (7.25) is cast. We define the critical density of the universe as

$$\rho_c \equiv \frac{3H^2}{8\pi G} \tag{7.29}$$

We point out that this quantity is time dependent. Then, dividing equation (7.25) by H^2 and replacing H^2 with ρ_c one obtains

$$1 = \sum_i \frac{\rho_i}{\rho_c} + \frac{\Lambda}{8\pi G\rho_c} - \frac{3k}{8\pi G a^2 \rho_c} \tag{7.30}$$

We can define the cosmological constant "density" as

$$\rho_\Lambda = \frac{\Lambda}{8\pi G} \tag{7.31}$$

and the curvature "density" as

$$\rho_k = \frac{3k}{8\pi G a^2} \tag{7.32}$$

which may be positive or negative, and scales as a^{-2}, in order to write

$$1 = \sum_X \frac{\rho_X}{\rho_c} - \frac{\rho_k}{\rho_c} \tag{7.33}$$

where the summation index X now extends to the cosmological constant as well. Finally, we introduce the density parameters

$$\Omega_X \equiv \frac{\rho_X}{\rho_c} \tag{7.34}$$

The first Friedmann equation is now in the form

$$1 + \Omega_k = \sum_X \Omega_X \tag{7.35}$$

which looks really simple, but is telling us a very crucial fact. In a universe with no curvature, $k = 0$, as *experimentally ours seems to be*, the sum over all density parameters is 1. Equivalently the total density of the universe, including the cosmological constant, is equal to the critical density. In this case the density parameter also happens to tell us what fraction of the universe is made up by each component.

In case $k > 0$, where the spatial geometry is a sphere, the total density of the universe would be larger than the critical density. We also say the universe is closed. In the case $k < 0$, where the spatial geometry is a hyperboloid, the total density is less than ρ_c. Our universe seems to be flat.

The first Friedmann equation can also be rewritten to make the scaling of the various components more explicit. We indicate quantities measured today with a subscript or superscript 0, such as a_0 or Ω_X^0. Then

$$\frac{H^2}{H_0^2} + \Omega_k^0 (\frac{a}{a_0})^{-2} = \sum_X \Omega_X^0 (\frac{a}{a_0})^{-3(1+w_X)} \tag{7.36}$$

7.5 Solutions to the Friedmann equations

In the most general case, when we track every component of the standard $\Lambda\mathrm{CDM}$ model, the Friedmann equation, with the continuity equations, must be solved numerically. However we can gain much understanding by considering simplified versions of the content of the universe. In the following we will adopt the convention that the scale factor *today* $a_0 = 1$.

7.5.1 One component universe, $k = 0$, $\Lambda = 0, w > -1$

Consider a universe with one component of a perfect fluid with equation of state given by $P = w\rho$. This corresponds to an idealized version where the density of one component of the universe is much larger than the others. Our universe has indeed passed through several eras where this was a good approximation. We call these the *radiation era*, the *matter era*, and the *dark energy era,* with obvious meaning. In this moment, we are transitioning between a matter dominated era and a dark energy dominated one. Here we will deal with the Friedmann equation during the radiation or matter eras, postponing the discussion of the dark energy era, where $w = -1$.

As we have shown in (7.24), the continuity equation tells us how the density changes with scale. In this case proportionally to $a^{-3(1+w)}$. Using the definition of density parameter (7.34) and critical density (7.29) the right hand side of the Friedmann equation can be written through $\frac{8\pi G}{3}\rho_0 = \Omega_0 H_0^2$. Noticing that for one component in a flat universe $\Omega_0 = 1$ by the closure condition (7.35), we write

$$\frac{a'^2}{a^2} = H_0^2 a^{-3(1+w)} \tag{7.37}$$

We can integrate the differential equation as

$$da \cdot a^{\frac{1}{2}+\frac{3}{2}w} = dtH_0 \tag{7.38}$$

So long as $w > -1$ we can integrate both sides and set $a = 0$ at $t = 0$, which corresponds to the "Big Bang". The solution is

$$a(t) = \left(\frac{3}{2}(1+w)H_0\right)^{\frac{2}{3(1+w)}} t^{\frac{2}{3}\frac{1}{1+w}} \tag{7.39}$$

Since we are in an idealized universe with one component, the overall constant is not that important. What is important is the dependence of the scale factor on time. $a(t) \propto t^{\frac{2}{3}\frac{1}{1+w}}$. In the special cases of matter ($w = 0$) and radiation ($w = \frac{1}{3}$)

$$\begin{cases} w = 0 & a \propto t^{\frac{2}{3}} \\ w = \frac{1}{3} & a \propto t^{\frac{1}{2}} \end{cases} \tag{7.40}$$

As a useful reference we also establish the relationship between conformal time τ and coordinate time. Through the definition

$$\tau = \int_0^t \frac{dt'}{a(t')} = \left(\frac{1 - \frac{2}{3(1+w)}}{(\frac{3}{2}(1+w)H_0)^{\frac{2}{3(1+w)}}}\right) t^{1-\frac{2}{3}\frac{1}{1+w}} \tag{7.41}$$

Again, we report the overall constant for completeness but it is the proportionality $\tau \propto t^{1-\frac{2}{3}\frac{1}{1+w}}$ which is important. In the usual interesting cases we have

$$\begin{cases} w = 0 & \tau \propto t^{\frac{1}{3}} \\ w = \frac{1}{3} & \tau \propto t^{\frac{1}{2}} \end{cases} \tag{7.42}$$

Note that the integration $\frac{dt}{a(t)}$ is only convergent at $t = 0$ for $w > -\frac{1}{3}$ so we restrict ourselves to that case when discussing conformal time. We will discuss the other case when we study the situation with $w = -1$ in section 7.5.2. What is important to take notice at this point is that the conformal time is increasing with coordinate time, we will discuss the significance of this in section 7.8. Using some straightforward algebra we can express the scale factor in terms of conformal time, a relationship which will often come in handy

$$\begin{cases} w = 0 & a = \frac{H_0^2}{4}\tau^2 \\ w = \frac{1}{3} & a = H_0\tau \end{cases} \tag{7.43}$$

Finally we can write down the Hubble factor of this solution

$$H = \frac{2}{3}\frac{1}{1+w}\frac{1}{t} \tag{7.44}$$

$$H = H_0 a^{-\frac{3}{2}(1+w)} \tag{7.45}$$

So for any $w > -1$ the Hubble factor decreases as $\frac{1}{t}$. The conformal Hubble factor, in terms of conformal time, has the same dependence

$$\mathcal{H} = \frac{2}{1+3w}\frac{1}{\tau} \tag{7.46}$$

Interestingly from the Hubble factor (7.44) measured today we can give a quick estimate of the age of the universe at this stage. For the most part, the universe has been in a radiation or matter dominated era. The radiation era only lasted about 10^5 years, a negligible time compared to the expected age of the universe of roughly 10^{10} years. Thus if we take the universe to be mostly matter dominated, $w = 0$, we obtain

$$t_{\text{Uni}} = \frac{2}{3H_0} \tag{7.47}$$

The latest measured Hubble constant is

$$H_0 = 67\frac{km}{s \cdot Mpc} \tag{7.48}$$

which implies the time since Big Bang to be, according to our simplified model,

$$t_{\text{Uni}} = 9.7 \cdot 10^9 \text{year} \tag{7.49}$$

This is certainly a good order of magnitude estimate for little work, but is off the mark by about four billion years. We will understand why when we calculate a more precise estimate in equation (7.76) .

A more interesting thought, is that it is actually the Hubble constant today which marks our point in the evolution of the universe, and not the particular value of the scaling factor. An overall, time independent, re-scaling of the scale factor(for example positing that $a_0 \neq 1$) is simply a re-scaling of the coordinates on the spatial slice and does not represent any physical change. Changing the Hubble constant is equivalent to moving us to a different moment in time. Since we cannot measure the time since the Big Bang directly, it is our measurement of H_0 which tells us *when* we are.

7.5.2 Dark energy dominated universe, $k = 0$, $w = -1$

Let's consider a universe with $k = 0$ and a fluid component with $w = -1$. Equivalently, one can consider a cosmological constant. The Friedmann equation is simply recast as

$$H = H_0 \tag{7.50}$$

which is the immediate result: in a dark energy dominated universe the Hubble factor is a constant. Integrating both sides one gets

$$a(t) = a_i e^{Ht} \tag{7.51}$$

where we dropped the subscript in the Hubble factor. In a dark energy dominated case, there is no time at which $a = 0$. We arbitrarily choose a time as $t = 0$ and a_i is the scale

factor at that moment. Thus, the first thing we note, from an experimental point, is that there is no way to mark time. Indeed the Hubble factor is constant and as such does not mark time as we noticed when solving the case $w \neq 1$. There is no component in the universe whose density is changing, that we could take measurements of. Thus, how could one experimentally determine that the universe is actually expanding?

We can't. To understand this one can plug in the scale factor in the Ricci tensor using (6.33) and (6.34) and find that it manifestly satisfies

$$R_{\mu\nu} = 3H^2 g_{\mu\nu} \tag{7.52}$$

which is the constant curvature condition (6.8) in four-dimensions. The space time is therefore a maximally symmetric four-dimensional spacetime. There is no way to define a special time since the spacetime now is homogeneous in time, just as it is in space. This space is known as the *de-Sitter space*. It can be shown that the de-Sitter space is the unique maximally symmetric four-dimensional spacetime with positive curvature.

Coming back to our universe, it seems that although the universe will eventually be in a dark energy dominated space-time forever, it has not always been like this and therefore it is meaningful to discuss its evolution even in this late stage. The conformal time in particular is of interest, through $d\tau = \frac{dt}{a}$ we obtain

$$\tau = \frac{1}{a_i H}(1 - e^{-Ht}) \tag{7.53}$$

As opposed to the radiation and matter dominated eras, the conformal time decreases as time increases and is exponentially big in the far past. This is a clue to solving the horizon problem through a very early inflationary phase. If, up to $\sim 10^{-32}s$ after the initial singularity, the universe was expanding in this manner, the conformal time in the past may much larger than the size of the observable universe.

7.5.3 Matter dominated, with curvature, $k \neq 0, w = 0, \Lambda = 0$

Our universe, it seems, is flat. However it is very important to understand the case of a non-flat universe, because of the interesting implications on flatness itself. For $k > 0$ there is also the interesting possibility of a *big crunch*, a moment in the future where the scale factor returns to zero. The scale factor increases until some maximum, when $H = 0$ and the decreases. Using the Friedmann equation the maximum scale factor is

$$a_M = \frac{\Omega_0 H_0^2}{k} \tag{7.54}$$

where Ω_0 is the density parameter of matter today. In the following we will use the factor a_M even in the case $k < 0$, taking note that in that case it is negative and does not have a physical significance.

The solution to this problem is easier in conformal time. By noticing that $\frac{d}{dt} = \frac{1}{a}\frac{d}{d\tau}$ it is simple to rewrite the first Friedmann equation as

$$\dot{a}^2 = \Omega_0 H_0^2 a - k a^2 \tag{7.55}$$

where dots indicate derivatives with respect to conformal time. Completing the square $\Omega_0 H_0^2 a - a^2 = -k(a^2 - a_M a + a_M^2/4) + k a_M^2/4$ we get

$$\dot{a}^2 = -k(a - \frac{a_M}{2})^2 + k\frac{a_M^2}{4} \tag{7.56}$$

It's easier now to work with the auxiliary variable

$$\tilde{a} = a - \frac{a_M}{2} \tag{7.57}$$

For $k > 0$ the solution is found by integrating

$$\int \frac{d\tilde{a}}{\sqrt{-\tilde{a}^2 + \frac{a_M^2}{4}}} = \int \sqrt{k} d\tau \tag{7.58}$$

We choose $\tilde{a} = 0$ at the conformal time τ_M as the integration extremes and obtain

$$a = \frac{a_M}{2}(1 + \sin\sqrt{k}(\tau - \tau_M)) \tag{7.59}$$

Since we know that $a = 0$ at $\tau = 0$ we find that

$$\tau_M = \frac{\pi}{2\sqrt{k}} \tag{7.60}$$

so the solution is more explicitly

$$a = \frac{a_M}{2}(1 - \cos\sqrt{k}\tau) \tag{7.61}$$

Integrating with respect to the conformal time, we arrive to an expression for coordinate time

$$t = \frac{a_M}{2}(\tau - \frac{\sin\sqrt{k}\tau}{\sqrt{k}}) \tag{7.62}$$

which could be inverted to find $a(t)$. In particular, the scale factor is maximum at $\tau = \frac{\pi}{\sqrt{k}}$ and is again zero at $\tau = \frac{2\pi}{\sqrt{k}}$. This means that the total lifetime of the universe, from Big Bang to Big Crunch, is finite

$$t_{\text{Uni}} = \frac{\Omega_0 H_0^2}{k^{\frac{3}{2}}}\pi \tag{7.63}$$

The case $k < 0$ is qualitatively different and solution is found in the same way but expressed, unsurprisingly, in terms of hyperbolic functions

$$a = \frac{a_M}{2}(\cosh\sqrt{-k}\tau - 1) \tag{7.64}$$

$$t = \frac{a_M}{2}(\frac{\sinh\sqrt{-k}\tau}{\sqrt{-k}} - \tau) \tag{7.65}$$

As we said, although our universe appears to be flat, it is interesting to understand the non-flat case. Currently, it must be the case that the curvature has not yet had much effect. In fact, we have seen that the curvature "density" $\rho_k \propto a^{-2}$ while matter has a density $\rho_m \propto a^{-3}$. As the scale factor increases the effect on the expansion due to curvature should

become more relevant than that due to matter. This may happen in the future. The real question however is why it has not happened yet. Let's take the case $k > 0$ as an example (the case $k < 0$ is the same). Assuming $\sqrt{k}\tau \ll 1$ then (7.61) can be Taylor expanded and we find

$$a \propto \tau^2 \tag{7.66}$$

which is the result already found for a universe dominated by matter (7.43). The conformal Hubble factor is

$$\mathcal{H} \propto \tau^{-1} \tag{7.67}$$

and combining the definitions of the curvature (7.32) and critical (7.29) density, and recalling that $\mathcal{H} = aH$, we find that the density parameter of the curvature is

$$\Omega_k \propto \frac{1}{\mathcal{H}^2} \propto \tau^2 \propto a \tag{7.68}$$

for small times. As expected, the contribution of the curvature grows with time. The problem may become more evident now. In fact the experimental value of Ω_k today is very close to zero. Considering the experimental error, it is reasonable to take

$$\Omega_k \sim 0.01 \tag{7.69}$$

which means that at the beginning of the universe, the curvature Ω_k must have been extremely close to zero. This is the origin of the flatness problem, which we will discuss more in section 7.7.

7.5.4 Matter and dark energy universe, $k = 0$,$\Lambda \neq 0$,$w = 0$

We now discuss a much more realistic solution of our universe. Experimentally, our universe appear to be flat $k = 0$, have a current density of matter (mostly in the form of dark matter) of about

$$\Omega_{m0} = 0.32 \tag{7.70}$$

a current dark energy density of

$$\Omega_{\Lambda 0} = 0.68 \tag{7.71}$$

The latest measure of the Hubble constant from the PLANCK experiment is

$$H_0 = 67 \frac{km}{s \cdot Mpc} \tag{7.72}$$

The first Friedmann equation (7.25), recasting $\Lambda/3 = \Omega_{0\Lambda} H_0^2$ can be written as

$$\frac{da}{H_0 \sqrt{\Omega_{0\Lambda} a^2 + \Omega_{0m} \frac{1}{a}}} = dt \tag{7.73}$$

Integrating on both sides and choosing $a = 0$ at $t = 0$ we obtain[3]

$$a(t) = (\frac{\Omega_{m0}}{\Omega_{\Lambda 0}})^{\frac{1}{3}} \sinh^{\frac{2}{3}} \left(\frac{3}{2} \sqrt{\Omega_{0\Lambda} H_0^2} t \right) \tag{7.74}$$

For small values of t, $\sinh x \propto x$, and we recover the evolution of the scale factor in the matter dominated universe $a \propto t^{\frac{2}{3}}$. For late times, $t \to \infty$, $\sinh^{\frac{2}{3}} \frac{3}{2} x \propto e^x$ we recover the exponential expansion in a dark energy dominated universe.

The Hubble factor

$$H = \sqrt{\Omega_{0\Lambda} H_0^2} \text{cotanh} \frac{3}{2} \sqrt{\Omega_{0\Lambda} H_0^2} t \tag{7.75}$$

We can invert (7.74) and obtain the time of the universe at a certain scale factor

$$t(a) = \frac{2}{3} \frac{1}{\sqrt{\Omega_{0\Lambda} H_0^2}} \text{arcsinh} \sqrt{\frac{\Omega_{\Lambda 0}}{\Omega_{m0}}} a^{\frac{3}{2}} \tag{7.76}$$

And for $a = a_0 = 1$ we get the age of the universe. Plugging in the experimental values for $\Omega_{m,\Lambda 0}$ and H_0 we obtain

$$t_{\text{Uni}} = 13.8 \cdot 10^9 \text{year} \tag{7.77}$$

which is exactly the estimate given by the latest data. Our universe contained an early era in which radiation was dominant and thus the situation under study here is not correct. However, it turns out that era lasted for only 10^5years and is negligible when estimating the age of the universe. Interestingly, the presence of dark energy has increased the estimated age of the universe by about 4 billion years with respect to our estimate with only matter present (7.49). This can be understood by the fact that in a matter only dominated universe the Hubble factor, which as we pointed out is what really identifies our position along the time evolution, decreases as $1/t$ while in the presence of dark energy the Hubble factor decreases more slowly with time, with an asymptote at $H = \sqrt{\Omega_{0\Lambda}} H_0$. To reach the same value of H_0, starting from some initial H_i , takes more time in the presence of dark energy. Experimentally, astrophysical objects have been found which were formed $t > 12 \cdot 10^9$ years ago, which means the universe must be at least that old, a strong point in favor for an exotic component of the universe.

7.5.5 The Milne universe $\Lambda = 0$, $k \neq 0$, no other matter

The Milne universe is a toy example as there is nothing realistic to it. It contains a curvature, $k \neq 0$ but no other fluid component nor cosmological constant. The Friedmann equation reduces to

$$a'^2 = -k \tag{7.78}$$

so this is a valid solution only for $k < 0$. Thus

$$a(t) = \pm \sqrt{-k} (t - t_0) \tag{7.79}$$

[3]The integral $\int \frac{dx}{\sqrt{x^2 + c^2 x^{-1}}} = \frac{2}{3} \text{arcsinh} \frac{x^{\frac{3}{2}}}{c}$ can be evaluated by using the substitution $x = c^{\frac{2}{3}} \sinh^{\frac{2}{3}} t$

where t_0 is an arbitrary integration constant. The Hubble constant is

$$H = \frac{1}{t - t_0} \tag{7.80}$$

As with the case of an empty, flat, universe with a cosmological constant, we can guess on physical grounds that since there is nothing measurable to mark the expansion of the universe this universe is in some sense not dynamic. With a little more mathematical inclination we may notice that the FRW metric with this scale factor has a coordinate singularity at $t = t_0$. But t_0 is an arbitrary integration constant. Surely this cannot be a singularity of space-time itself. Therefore we immediately check the values of the Riemann tensor and find that it is zero

$$R_{\mu\nu\rho\sigma} = 0 \tag{7.81}$$

So the Milne universe is actually Minkowski space in an unusual coordinate system. One can of course work out what coordinate transformation is needed to return to the canonical flat metric. It turns out the metric we constructed this way covers the interior of a Minkoswki light-cone.

The more interesting point here is that all FRW metrics can be shown to be conformally flat, even in the $k \neq 0$ case. In the case $k \neq 0$ the overall conformal factor will not be simply a function of the time coordinate (in the case $k = 0$ this factor is simply $a(\tau)$). It can also be shown that any conformally flat space-time which solves the Einstein equations in a vacuum is flat. The Milne universe is an example of this.

7.5.6 Dynamical analysis

By taking the two Friedmann equations (7.25), (7.26) in the presence of a single fluid with equation of state $P = w\rho$, we can find that $H' = \frac{dH}{dt} = \frac{a''}{a} - H^2$ can be expressed as[171]

$$\frac{H^{'}}{H^2} = -\frac{\Omega}{2}(1 + 3w) + \Omega_\Lambda - 1 \tag{7.82}$$

or, through $1 + \Omega_k = \Omega + \Omega_\Lambda$, as

$$\frac{H'}{H^2} = -\frac{3}{2}(1 + w)(1 - \Omega_\Lambda) - \frac{3}{2}\Omega_k(\frac{1}{3} + w) \tag{7.83}$$

Also, using the definitions, it is straightforward to show that

$$\Omega_\Lambda^{'} = -2\Omega_\Lambda \frac{H^{'}}{H} \tag{7.84}$$

$$\Omega_k^{'} = -2H\Omega_k - 2\Omega_\Lambda \frac{H'}{H} \tag{7.85}$$

Since both Ω and $\frac{H'}{H^2}$ can be expressed algebraically through Ω_Λ and Ω_k these two equations are almost enough to constrain the dynamical evolution of the universe. There is only a pesky H to consider. To move forward, let's consider the derivatives with respect to $x = \ln \frac{a}{a_0}$. Then $\frac{d}{dt} = \frac{dx}{dt}\frac{d}{dx} = H\frac{d}{dx}$ and so we can analyze the dynamics of the Friedmann

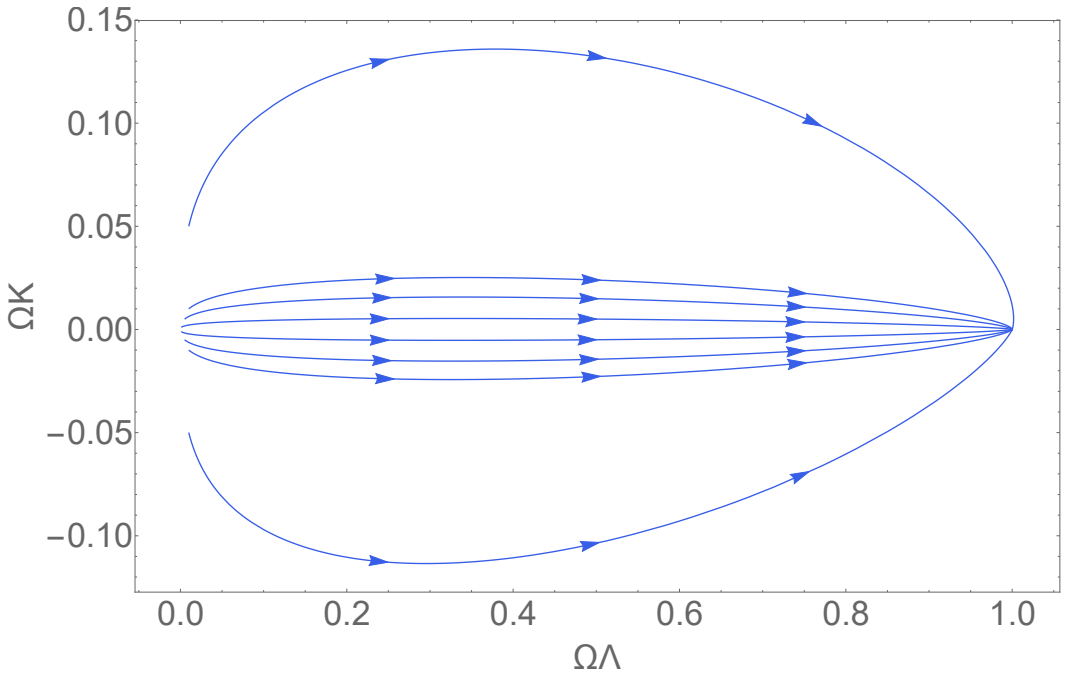

Figure 7.1: Dynamical evolution of the Friedmann equations in a universe with curvature, dark energy and matter. Regardless of the initial conditions, the path through parameter space Ω_Λ, Ω_k evolves towards the fix point at $(1, 0)$, whereas $(0, 0)$, a universe dominated by the matter, is a repeller.

equation through the dependence of the variable x

$$\frac{d}{dx}\begin{pmatrix}\Omega_\Lambda \\ \Omega_k\end{pmatrix} = -2\begin{pmatrix}\frac{H'}{H^2} & 0 \\ 0 & 1+\frac{H'}{H^2}\end{pmatrix}\begin{pmatrix}\Omega_\Lambda \\ \Omega_k\end{pmatrix} \tag{7.86}$$

We are interested in particular in the fixed points of this system. These are points where the tangent vector of the solution curves are zero and could represent points of stable or unstable equilibrium for the system. It turns out there are only three fixed points. We could prove this by explicitly solving a system of equations but we shall simply point out the solutions. An illustration can be found in figure 7.1. The first fixed point is $a)$

$$(\Omega_\Lambda, \Omega_k) = (0, 0) \tag{7.87}$$

a flat universe with no cosmological constant. The second fixed point $b)$ is

$$(\Omega_\Lambda, \Omega_k) = (1, 0) \tag{7.88}$$

where $H^{'} = 0$. This is, in fact, the de-Sitter, or dark energy dominated, universe we have seen before with a constant Hubble factor. We can guess, from our previous analysis, that this is an *attractor,* or point of stable equilibrium for the system. The third fixed point $c)$ is

$$(\Omega_\Lambda, \Omega_k) = (0, -1) \tag{7.89}$$

which is the Milne universe where $H^{'} = -H^2$. We have certainly already seen, when

discussing explicitly a universe with negative curvature and matter, but no cosmological constant, that curvature will dominate. We may guess that the Milne universe is a point of stable equilibrium but in fact we will shortly see that in the presence of dark energy it is a *saddle point;* it is attractive in one direction but repellent in another.

To analyze the dynamics around the fixed points one works at first order in the perturbation around the point

$$\begin{pmatrix}\Omega_\Lambda\\ \Omega_k\end{pmatrix} = \begin{pmatrix}\bar{\Omega}_\Lambda\\ \bar{\Omega}_k\end{pmatrix} + \begin{pmatrix}\omega_\Lambda\\ \omega_k\end{pmatrix} \tag{7.90}$$

where the barred quantities $\bar{\Omega}_x$ are the values at the fixed points and ω_x are the dynamical variables.

Around the first fixed point a), $\frac{H^{'}}{H^2} = -\frac{3}{2}(1+w) + o(\omega)$ and the system is, to order $o(\omega_{\Lambda,k})$

$$\frac{d}{dx}\begin{pmatrix}\omega_\Lambda\\ \omega_k\end{pmatrix} = \begin{pmatrix}3(1+w) & 0\\ 0 & 1+3w\end{pmatrix}\begin{pmatrix}\omega_\Lambda\\ \omega_k\end{pmatrix} \tag{7.91}$$

The fixed point is a stable equilibrium point if both the eigenvalues are negative. For a fluid with with $w \geq 0$, such as matter or radiation the fixed point $(0,0)$ is a repeller. The universe will not spontaneously evolve toward this point, but exponentially move away (at least while the perturbations are small). This is important, since it seems that close to the Big Bang, the universe was in a situation where it was radiation dominated ($w = \frac{1}{3}$) and had negligible curvature and cosmological constant density. It remains a mystery why we seem to be still very close to the repeller, since the curvature is very small today. Again, this is the manifestation of the flatness problem. Of course, one may consider more exotic components of matter which have a negative w. For $w < -\frac{1}{3}$ for example, the point becomes a saddle point and for $w < -1$ it even becomes stable.

Around fixed point b), $\frac{H'}{H^2} = \frac{3}{2}(1+w)\omega_\Lambda - \frac{3}{2}(\frac{1}{3}+w)\omega_k + o(\omega^2)$ and the system is

$$\frac{d}{dx}\begin{pmatrix}\omega_\Lambda\\ \omega_k\end{pmatrix} = \begin{pmatrix}-3(1+w) & 1+3w\\ 0 & -2\end{pmatrix}\begin{pmatrix}\omega_\Lambda\\ \omega_k\end{pmatrix} \tag{7.92}$$

The eigenvalues are $\lambda_1 = -3(1+w)$ and $\lambda_2 = -2$ so for any matter with $w > -1$ $(\Omega_\Lambda, \Omega_k) = (1,0)$ is a stable fixed point. We can generally expect the universe to evolve towards a dark energy dominated phase, regardless of the initial conditions.

Finally consider fixed point c) around this the system is

$$\frac{d}{dx}\begin{pmatrix}\omega_\Lambda\\ \omega_k\end{pmatrix} = \begin{pmatrix}2 & 0\\ 3(1+w) & -1-3w\end{pmatrix}\begin{pmatrix}\omega_\Lambda\\ \omega_k\end{pmatrix} \tag{7.93}$$

The first eigenvalue $\lambda_1 = 2$ is positive and thus this is certainly not a stable equilibrium for the universe. The associated eigenvector is $(1,1)$. The second eigenvalue is $\lambda_2 = -(1+3w)$ which is associated with the eigenvector $(0,1)$. Thus we see that in the absence of dark energy this would be an attractor of the system, as we had found by explicitly solving for a universe with matter and non-zero curvature. In the presence of dark energy it becomes a saddle point.

Parameter	Value
Ω_{r0}	$0.94 \cdot 10^{-4}$
Ω_{b0}	0.022
Ω_{c0}	0.268
$\Omega_{\Lambda 0}$	0.683
H_0	$67.36 \frac{km}{s \cdot Mpc}$
T_{CMB}	$2.72K$

Table 7.1: Measured parameters of our universe in the ΛCDM model [61, 102]. These will be useful to evaluate quantities in the text.

7.6 A brief history of our universe

Let's take a moment to recap our understanding of the discussion so far and how it applies to our universe. The latest data from cosmological observations indicate that the universe is flat, $\Omega_k = 0$, with its content as tabulated in 7.1.

The density parameters refer obviously to the current density of the universe, these have been estimated from cosmological observations and are generally accurate at the percent level. Ω_{r0} is the density parameter of radiation today. This includes photons and the three standard model neutrinos. Although neutrinos have a small mass they are relativistic throughout most of the universe's evolution. The quantity Ω_{b0} refers to the density parameters of *baryons*, which in the terminology of cosmology, generally refers to both baryons (protons, neutrons, etc.) *and charged leptons*(usually only electrons are important). Perhaps stunningly, we see that most of the density of the universe is composed of two quantities which are not present in the standard model. Ω_{c0} is the density of *cold dark matter*. Dark matter is a gravitating component of matter which appears to be non-relativistic for at least most of the time of the universe (hence the name *cold*), which is stable or has a lifetime greater than the age of the universe, and does not interact with regular matter. Its interpretation in terms of a particle physics model is currently unknown. Finally, most of the density of the universe comes from *dark energy* $\Omega_{\Lambda 0}$. Dark energy may be attributed to the presence of a non-zero cosmological constant, although, since we don't have a clear understanding of it yet, we will try to keep the terminology separated. It can be inferred that dark energy acts as a perfect fluid with an equation of state parameter w consistent with -1.

The Hubble constant H_0 determines the expansion rate of the universe and sets the age of it. We remark that, as of this writing, there are inconsistencies in the measurement of H_0 done by different experiments, with some reporting a value as high as $74 \frac{km}{s \cdot Mpc}$[180], while the experimental error is around $\sim 1 \frac{km}{s \cdot Mpc}$. For now, this inconsistency is unimportant. In the literature the parameter h is often used, defined as

$$H_0 = h \cdot 100 \frac{km}{s \cdot Mpc} \tag{7.94}$$

and often the density parameters are reported through the combination $\Omega_0 h^2$, which is proportional to the physical density ρ of a species.

The universe today contains the *cosmic microwave background radiation*, or CMB. This is a black-body spectrum of photons arriving uniformly, within fluctuations of order 10^{-5}, from

all direction of the sky. The temperature T_{CMB} means that this radiation has a spectrum peaked in the microwaves. The CMB is the primary source of knowledge of the early universe.

The presence of a uniform radiation among the sky was historically the "smoking gun" for the existence of the Big Bang. At a finite time in the past the scale factor a must have been zero, and we generally consider the evolution from that moment $t = 0$ until today.

The early universe can be considered a very uniform plasma (at least as uniform as the CMB, and we will see why shortly) containing principally photons, electrons, neutrinos, protons, neutrons, Helium-4 nuclei and dark matter. At least as $t \to 0$ all these components are in thermal and chemical equilibrium, except perhaps dark matter. As we will see in section 8.3, the temperature of this plasma mostly decays as $T \propto a^{-1}$. Since radiation density scales as a^{-4} it is clear that as $a \to 0$ it is the dominant component of the universe. The very early universe is in a radiation dominated era. We can readily calculate the scale factor at which the density of radiation (photons and neutrinos) is equal to the density of matter (baryons and dark matter).

$$\frac{\rho_r}{\rho_m} = \frac{\Omega_{r0} a_{eq}^{-4}}{\Omega_{m0} a_{eq}^{-3}} = 1 \tag{7.95}$$

Is solved by (considering $a_0 = 1$)

$$a_{eq} = 2.9 \cdot 10^{-4} \tag{7.96}$$

This moment is known as *matter-radiation equality.* This happens around

$$t_{eq} \sim 5 \cdot 10^4 \text{year} \tag{7.97}$$

after the Big Bang. Much before this, we say *deep in the radiation era*, the temperature T decreases to a few MeV the neutrinos *decouple* from the plasma. That is, the weak interactions such as $\nu e \leftrightarrow \nu e$, which depend not only on the cross section but on the density and temperature, are not sufficient to keep the neutrinos in thermal equilibrium with the rest of the universe. From then on, the neutrinos evolve separately, contributing only to gravitational interactions. Shortly after at around $T \sim 1MeV$, *Big Bang Nucleosynthesis* takes place. This is the period when protons and neutrons bind to form heavier nuclei, such as Deuterium, Helium and Lithium, and is the only source of nuclei formation in the universe until stellar fusion happens much later. We will discuss some details in section 12.

As the universe cools down to a temperature of a few eV neutral atoms begin to form abundantly. This is crucial. A free photon scatters much more on charged particles than neutral atoms and therefore the universe becomes transparent. The cosmic microwave background is the free streaming radiation from around this time. The universe is in fact so transparent that photons reach us today roughly unimpeded. At those times one can think of the mean free path of the photon being very short. The photons scatters often on the electrons and protons of the plasma until it becomes neutral and suddenly the mean free path becomes cosmologically large. This is why we idealize the photons of the CMB to be from a *last scattering surface* (LSS). From Earth, we are looking back in time and can see clearly in the distance until the universe becomes opaque, much like we can see the outer

surface of a flame. The last scattering of photons happened around

$$t_{LSS} \sim 380.000\text{year} \tag{7.98}$$

after the Big Bang. This period of the universe is also known as *recombination*[167].

We have mentioned that the cosmic microwave background is nearly uniform across all the sky. This implies that the early universe was very uniform itself. There are perturbations across the sky, known as anisotropies, in the CMB of about one part in 10^5[58, 92]. The study of these anisotropies is a main area of research in cosmology. These are due to over and under-densities of the plasma in the early universe. With time, the over-densities can grow to become stars, galaxies and the myriad of structures we observe today. It is also clear that anisotropies must exist in the early universe, for if everything were perfectly in thermal equilibrium, all we would have today would be a CMB with a uniform temperature of T_{CMB} (which is the cosmic temperature today).

Studying structure formation, it has become accepted that dark matter must be the foundation upon which galaxies are formed[129]. Thus dark matter is key to galaxy formation and must, for this purpose, be non-relativistic. It is not known if dark matter has ever been in thermodynamical equilibrium with the rest of the universe. Unless we specify differently, it is usually assumed to be non-relativistic at every era of interest. The production mechanism of dark matter is not known. Although there are many candidates from particle physics, and thus possible early universe production mechanism, none can be confirmed experimentally. Until now it seems that the interactions of dark matter may be purely gravitational.

Dark energy is the other great mystery of modern cosmology. In the simplest case, this is attributed to a cosmological constant. A cosmological constant can arise through particle physics from a vacuum energy. In quantum mechanics a vacuum energy is dropped, since only relative energies are physical. In general relativity, absolute values of energy are physical and the value of the vacuum energy becomes important. Attempts to calculate the cosmological constant from the standard model return values grossly different from what is observed. Something is not right. It is an observable fact however that we have quite recently (in the last ~ 4 billion years) entered an era where the vacuum energy is comparable to the matter density.

This history of the universe, with its components, is known as the ΛCDM model of the universe and is widely successful.

We mention a few other periods of the universe which are likely to have happened in the early universe. At a temperature around $T \sim \Lambda_{QCD} \sim 200MeV$ the universe is expected to have gone through a QCD phase transition[219]. Above $T \sim 200MeV$ the fundamental degrees of freedom are nearly free quarks. A phase transition must have happened, leaving the fundamental degrees of freedom to be mesons and baryons afterwards. The detail of this transition are hard to understand, since they involve QCD in a strong coupling regime. In the most likely scenarios the phase transition is second order. Although the phase transition is complicated, the universe is expected to be in thermal equilibrium both before and after, so most cosmological consequences are "washed out" by the return to equilibrium.

Another phase transition is expected to have happened at the electro-weak scale $T \sim 200GeV$ when the Higgs field relaxes from a symmetric minimum to a non-symmetric one[117]. Particles are expected to go from being massless to massive. It seems that this phase transition is second order as well and that it largely leaves the successive expansion of the universe unaffected.

We should also mention the issue with matter-antimatter asymmetry. In fact, in our thermal history of the universe we have mostly left out anti-matter particles as $e+$ which have every means of being created in a thermal equilibrium. Indeed, by some unknown mechanism there is an excess of matter over anti-matter at least since nucleosynthesis.

7.7 The flatness problem

The observations of the universe imply that the spatial slices of the universe are flat or nearly so. An upper limit for Ω_{k0} is

$$|\Omega_{k0}| \lesssim 0.01 \tag{7.99}$$

This readily implies that the total density of the universe, including dark energy and dark matter, is equal or nearly so to the critical density. This is a highly unnatural situation which warrants an explanation and is known as the *flatness problem.* We have in fact analyzed the dynamics of the Friedmann equations and have found that the situation which $(\Omega_\Lambda, \Omega_k) = (0, 0)$ is a repeller for the system. The early universe was in such a situation and we would expect it to move away from this unstable point in phase space.

For most of its lifetime, the universe has been either in a radiation or matter dominated since *at least* when the scale factor was $a \sim 10^{-9}a_0$, and likely much earlier. At the earliest times the universe must have been radiation dominated. This can be inferred from the present values of Ω_{r0} and Ω_{m0} coupled with the fact that radiation dilutes with an extra factor of a^{-1} with respect to matter. At around $a \sim 10^{-4}a_0$ the matter density becomes comparable with radiation density. As can be seen by (7.32) and (7.29) the curvature density parameter, or deviation from flatness, is proportional to

$$\Omega_k \propto \frac{1}{a^2H^2} \tag{7.100}$$

Combining the scale factor evolutions (7.40) in the matter and radiation dominated case, with the Hubble factor (7.44), we find that during the radiation dominated era $\Omega_k \propto a^2$, whereas in the matter dominated era $\Omega_k \propto a$. This is puzzling. Unlike matter and radiation which we understand have decreased drastically since the early universe, the curvature must have actually increased. Since it is very close to zero today, it must have been extraordinarily so in the early universe. In fact we can simply estimate that around matter-radiation equality it was $\Omega_k \sim 10^{-4}\Omega_{k0}$ and in the earliest phases it must have been $\Omega_k \sim 10^{-14}\Omega_{k0}$. The problem may even be worse if we assume it is correct to extrapolate closer to the Big Bang.

To see it in another light we can rewrite the Friedmann equation as

$$(\rho_c - \rho)a^2 = -\frac{3k}{8\pi G} \tag{7.101}$$

Since the right hand side is a constant, the difference between the total density of the universe and the critical density must scale as a^{-2}. So that indeed, because $\rho \sim \rho_c$ today, within a factor of $\sim 1\%$, they must have been essentially equal as $a \to 0$.

Why was this the case? Is there a problem at all? Indeed , what determines the initial conditions of the universe is completely unknown. It may simply be that $k = 0$ is a principle of nature, just as we assumed homogeneity and isotropy to be. It may be. However, it would be more satisfying if the state of the universe today where to be similar regardless of the initial conditions, since then we mustn't rely on ad-hoc conditions at the Big Bang to explain the here and now. Therefore we ask ourselves if there is, or has been, some mechanism in the universe which *causes* it to be flat, regardless of any specific initial condition we may apply.

Indeed, by looking at a single component universe it is clear that Ω_k will *decrease with time* for $w < -\frac{1}{3}$. In particular, for a dark energy component $w = -1$, Ω_k will exponentially decrease to zero, regardless of any initial value.

This leads us to mention the *theory of inflation* as a possible solution to the flatness, and other problems. Inflation is the idea that in the very first instances after the Big Bang the universe experienced an exponential expansion, a de-Sitter phase, due to some unknown mechanism. After this, it is the regular cosmology proceeds with standard model content.

7.8 Horizons and their problem

Conformal time was introduced in (6.23) as the integral

$$\tau = \int_0^t \frac{dt'}{a(t')} \tag{7.102}$$

and up until now we have only used as a convenient parameter to perform calculations with the metric. It actually has an important physical significance, as it is related to the causal structure of a FRW metric. For a flat, $k = 0$, universe such as ours assume a photon is emitted at $t = 0$ at the co-moving coordinate $r = 0$ and travels in a radial direction. Then, since it travels along a null geodesic,

$$dt = a(t)dr \tag{7.103}$$

Dividing both sides by the scale factor and integrating along the geodesic we obtain

$$\tau = \int_0^r dr' = r \tag{7.104}$$

Of course, by homogeneity the point $r = 0$ where the photon is emitted is arbitrary. What we have found is that the conformal time is the maximum comoving distance a photon may have traveled since the Big Bang, this is known as the *cosmological particle horizon*. In other words, at a time t the past light-cone at $r = 0$ extends to radial comoving coordinates $r = \tau(t)$. So there can only be causal contact between events separated by at most $\tau(t)$. Note that r is not a distance, but simply the coordinate. The proper distance between $r = 0$

and r at a time t is $a(t)r$, but note that this is the instantaneous proper distance, and not the amount of space traversed by the photon. We will return to discussing distances in section 7.9.

For completeness, in the case $k \neq 0$ the relationship between τ and r is found in the same manner and is

$$\tau = \begin{cases} \frac{1}{\sqrt{k}} \arcsin \sqrt{k} r & k > 0 \\ \frac{1}{\sqrt{-k}} \text{arcsinh} \sqrt{-k} r & k < 0 \end{cases} \tag{7.105}$$

Now we must admit we cheated somewhat. In the definition of conformal time we assumed that the integral converges at $t \to 0$. This is indeed the case in a radiation or matter dominated universe, as we showed in (7.42). However, if the integral does not converge at the Big Bang, for example in the case of a dark energy dominated universe, there is no particle horizon; the past light-cone extends to all of space. In a practical situation, whenever one uses the conformal time, if the integral does not converge at $t \to 0$ an initial time t_i is used for its definition, so the quantity is useful, although its physical significance is altered.

A curious situation arises if we consider the behavior of conformal time as $t \to \infty$, far in the future. In our universe, it is apparent that we will eventually end in a dark energy dominated phase where $a \propto e^{Ht}$ at late times. Then it is apparent that as $t \to \infty$, τ does not grow indefinitely. This implies that there are regions of our universe which we will *never* be in causal contact with.

Now we get to the *horizon problem*, which is a very deep issue in our cosmological understanding of the universe. Our first hypothesis of the universe was that it was homogeneous, it is time to test whether this hypothesis is consistent with general relativity, the FRW metric and the evolution we have studied thus far. It turns out it is not.

Observation of the cosmic microwave background has shown that the early universe was homogeneous to a high degree. The conformal time at the surface of last scattering was about

$$\tau_{LSS} \simeq \frac{1}{2\sqrt{2}} (\frac{t_{LSS}}{H_0})^{\frac{1}{2}} \tag{7.106}$$

where we have used (7.42) assuming a radiation dominated, flat, universe until then. This is an over-estimation of the real value since the conformal time grows faster in the radiation dominated universe then the matter dominated one. At the time of last scattering, two points could have a *common causal cause* if they are separated by less than $2\tau_{LSS}$. On the other hand the conformal time difference between last scattering and today is (assuming only matter domination in the meantime)

$$\tau_0 \simeq \frac{1}{3(\frac{3}{2}H_0)^{\frac{2}{3}}} t_0^{\frac{1}{3}} \tag{7.107}$$

where we can easily drop a term due to $t_{LSS}^{1/3}$, since $t_0 \gg t_{LSS}$. This also is an over-estimation since the conformal time grows more slowly in the presence of dark energy. Today, the proper size of a "causal patch" of the CMB we observe is $2a_0\tau_{LSS}$, while the proper distance to it is $a_0\tau_0$. In a flat universe the surface of the last scattering surface today in terms of its radius is the usual one $4\pi(a_0\tau_0)^2$ while we can take the area of a causal patch to be $\sim (2a_0\tau_{LSS})^2$. The ratio of the two quantities is a rough estimate of how many

separate causal patches are observed in the CMB today and is

$$\frac{4\pi(a_0\tau_0)^2}{(2a_0\tau_{LSS})^2} \sim 10^4 \tag{7.108}$$

The very uniform CMB is thus composed of a great deal of areas of the universe which could not have a common causal case! Knowing this, it would be utterly implausible to expect the high degree of uniformity we observe of the early universe. This is known as the *horizon problem*.

As with the flatness problem, the horizon problem can be solved by the theory of inflation. By some mechanism, the first instances of the universe involve an exponential expansion. In fact if there were such an expansion from an early time t_0 to t_1 the change in conformal factor, using $a(t) = a(t_0)e^{H(t-t_0)}$ would be

$$\tau(t_1) - \tau(t_0) = \frac{1}{a(t_0)H}(1 - e^{-H(t_1-t_0)}) \tag{7.109}$$

Taking $\tau(t_0) = 0$ and $t_1 \gg t_0$ we obtain $\tau(t_1) = \frac{1}{a(t_0)H}$. Thus, through an exponential expansion, the conformal time can be made as large as is needed by reducing the scalar factor at the start of it $a(t_0)$. We could conclude that the causal patches at the last scattering surface are much larger than expected. In a theory of inflation, in fact, this exponential expansion is usually thought to begin at around $t_0 \sim 10^{-36}s$ and end around $t_1 \sim 10^{-32}s$ with $a(t_1)/a(t_0) \sim 10^{26}$.

7.9 Distances and redshift

The most reliable observable of an astrophysical object which is at cosmological distances from us is the redshift. As we have shown in (6.64) from distant objects are redshifted as

$$1 + z = \frac{a_0}{a_{Em}} \tag{7.110}$$

where a_{Em} is the scale factor at emission. We shall discuss how z relates to other observables which can be of use in cosmological observations, such as distances. First we shall introduce the *comoving distance*

$$\chi(r) = \int_0^r \frac{dr}{1-kr^2} = \begin{cases} \frac{1}{\sqrt{k}} \arcsin \sqrt{k}r & k > 0 \\ r & k = 0 \\ \frac{1}{\sqrt{-k}} \text{arcsinh} \sqrt{-k}r & k < 0 \end{cases} \tag{7.111}$$

The comoving distance is the radial distance between the coordinates $r = 0$ and r on the spatial slicing. In the FRW spacetime the *proper distance* from $r = 0$ to a coordinate r at time t is given by

$$d_P(r;t) = a(t)\chi(r) \tag{7.112}$$

The proper distance is the instantaneous length of a spacelike, constant time, geodesic. As such, it is not the the world-line of a real particle and its usefulness is limited.

We introduce the *luminosity distance*. Assuming an object with known luminosity $L = \frac{dE}{dt}$, in a flat spacetime we can infer the distance to the object by measuring the energy flux at our location

$$F = \frac{L}{4\pi d_L} \tag{7.113}$$

In a curved space time we must pay more attention to the quantities that appear. Indeed the flux measured at the Earth from an object at fixed co-moving coordinate r will be an energy per unit time and surface

$$dF = \frac{dE_{obs}}{dt_{obs} dS_{obs}} \tag{7.114}$$

Where we have made explicit that energy and time are at the observer's position. Assuming the measurement is made today, so $a = a_0$, the surface element is part of a sphere with proper radius $d_P(r, t_0)$, so $dS_{obs} = a_0^2 \chi^2 d\Omega$, where the elementary solid angle is the same both at emission and observation $d\Omega_{obs} = d\Omega_{em} \equiv d\Omega$. The amount of energy received must be a factor a less than that at emission, since each photon is redshifted. So $dE_{obs} = \frac{a}{a_0} dE_{em}$. On the other hand the rate with which we receive every photon is reduced. In fact if two photons are emitted from the object at times dt_{em} apart, their comoving distance will always be $d\chi = \frac{dt_{em}}{a}$, and the time difference we receive them is $dt_{obs} = a_0 d\chi = \frac{a_0}{a} dt_{em}$. We conclude that

$$F = \frac{dE_{em}}{dt_{em} 4\pi a_0^2 \chi^2(r)} (\frac{a}{a_0})^2 \tag{7.115}$$

Comparing with (7.113) we conclude that the luminosity distance is

$$d_L = a_0 \chi(r)(1 + z) \tag{7.116}$$

These formulas are telling us that cosmological objects appear fainter than they would otherwise, due to the expansion of the universe. Equation (7.116) is not entirely experimentally useful, since the comoving distance r is not easily measurable. We will shortly find an expression solely in terms of z.

Another possible way of measuring distance is the *angular distance*. Suppose that an object at distance coordinate r has a known linear transverse size l, we could measure the distance by the subtended angle as viewed from Earth:

$$\theta = \frac{l}{d_A} \tag{7.117}$$

Of course, we must generalize this to an expanding universe. The comoving transverse size of the object is la_0/a while the distance is $a_0 \chi(r)$ so the subtended angle would be

$$\theta = \frac{l}{a\chi(r)} \tag{7.118}$$

which implies the angular distance is

$$d_A = \frac{a_0 \chi(r)}{1 + z} \tag{7.119}$$

This also implies that distant object appear larger than they would otherwise. Note how angular and luminosity distance have different dependencies on the red-shift

$$\frac{d_L}{d_A} = (1+z)^2 \tag{7.120}$$

This measurement, by itself would be useful test of the cosmological model.

As we have pointed out, it is not easy to measure the comoving coordinate r so we wish to express the angular and luminosity distance purely in terms of z. To do so one would have to find the time for which

$$\int_{t(r)}^{t_0} \frac{dt}{a(t)} = \chi(r) \tag{7.121}$$

and, through the relationship between the scale factor and time, invert to express $r = r(z)$. We shall do this explicitly for small look-back times $(t_0 - t)$ and small z, to second order. We begin with a Taylor expansion of the scale factor around today

$$a(t) = a_0 \left(1 + H_0(t-t_0) - \frac{q_0 H_0^2}{2}(t-t_0)^2\right) + o(t-t_0)^3 \tag{7.122}$$

where we have introduced the deceleration parameter

$$q \equiv -\frac{a''}{aH^2} \tag{7.123}$$

evaluated today. The minus sign is a historical convention. It was expected that the expansion of the universe was decelerating, hence $a'' < 0$, and therefore with the minus sign in the definition, that the deceleration parameter would be positive. Surprisingly, the measurement of q_0 turned out to be negative, which was proof of the accelerated expansion of the universe due to dark energy. In fact, using the second Friedmann equation (7.26) it is immediate to find

$$q_0 = \frac{1}{2}\Omega_{m0} - \Omega_{\Lambda 0} \tag{7.124}$$

Using the expansion (7.122) in (7.121) we find, dropping third order terms,

$$\chi(r) = -\frac{1}{a_0}(t-t_0) + \frac{H_0}{2a_0}(t-t_0)^2 \tag{7.125}$$

To invert this at second order, note that $\chi(r)$ must also be small and that at first order $(t-t_0) = -a_0\chi(r) + o(\chi^2)$ which means that $(t-t_0)^2 = a_0^2\chi^2 + o(\chi^3)$ therefore

$$(t-t_0) = -a_0\chi(r) + \frac{H_0 a_0^2}{2}\chi^2(r) \tag{7.126}$$

Thus, using $z = \frac{a_0}{a} - 1$, we obtain to second order

$$z = -H_0(t-t_0) + H_0^2(t-t_0)^2(\frac{q_0}{2}+1) \tag{7.127}$$

$$z = H_0 a_0 \chi(r) + \frac{H_0^2 a_0^2}{2}\chi^2(r)(1+q_0) \tag{7.128}$$

which can be inverted for small z, χ

$$\chi(r) = \frac{z}{H_0 a_0} - \frac{z^2}{2H_0 a_0}(1 + q_0) \tag{7.129}$$

The luminosity distance may be expressed through the redshift as

$$d_L = D_{H_0}(z + \frac{z^2}{2}(1 - q_0)) \tag{7.130}$$

and the angular distance as

$$d_A = D_{H_0}(z - \frac{z^2}{2}(3 + q_0)) \tag{7.131}$$

having introduced the Hubble distance

$$D_H = \frac{1}{H} \tag{7.132}$$

Note that both measures of distance reduce to the well known Hubble law for small z: $z = H_0 d$.

8 Thermodynamics

We now study the thermodynamics of the fluid which composes the universe. Through most of the early universe, reaction rates between most species of particles are fast enough to keep them in thermal equilibrium. We can consider the early universe to be a homogeneous and isotropic fluid, with every species having a common temperature. There are many situations in which some species reaction rates with all the others falls below the rate of expansion of the universe. It is said that the species *decouples* and will evolve independently from then on, the most obvious case being neutrinos. Thus both the equilibrium thermodynamics and the falling out of equilibrium are important phenomena to understand the evolution of the universe.

8.1 Equilibrium distributions

Assuming reaction rates are fast enough to ensure thermal equilibrium, all species of particles can be thought of as perfect Fermi-Dirac or Bose-Einstein gases with distribution[143][4]

$$f_i(E, T) = g_i \left(e^{\frac{E - \mu_i}{T_i(t)}} \pm 1 \right)^{-1} \tag{8.1}$$

Each species has a degeneracy factor g_i, a chemical potential μ_i and may have its own temperature. If more than one species are in thermal equilibrium together, the temperature must be the same. In principle this distribution may depend on the position $\vec{x}$ and the direction of momentum $\vec{p}$, but the assumptions of homogeneity and isotropy of the universe imply only a dependence on time and on the magnitude of the momentum, or energy.

[4]Note that some source prefer to make the degeneracy factor g_i more explicit. By default, we will always treat it as part of the distribution for any species, thus it will appear implicitly inside any f_i.

Of course, such a thermal distribution for a species is valid only so long as it is in thermal equilibrium, at least with itself, and possibly with the other components of the universe. Generally, the equilibrium is maintained by some reaction rate $\Gamma \sim n\sigma$. We can consider equilibrium to be achieved so long as there are many reactions per Hubble time, that is if $\Gamma > H$. The rates have a dependence on temperature in the cross section, through the average energy of the particles. Usually the rate drops with temperature faster than the Hubble factor. There will be a moment when $\Gamma \simeq H$, and we say the species *decouples* from the rest of the universe. Decoupling is important as it leaves behind a non-negligible amount of thermal relics. The most notable example is neutrinos, which we will analyze explicitly. These drop out of equilibrium at temperatures of around a few MeV and then evolve independently. It may also be the case that dark matter is a thermal relic from the very early universe.

The number density of a species in equilibrium is given by

$$n_i(T) = \int f_i(E(p_i), T) \frac{d_3p}{(2\pi)^3\sqrt{-g}} \tag{8.2}$$

The expression is similar to that of a generic energy-momentum tensor (7.3), where the p_i are the canonically conjugate momenta, and the integration is done with respect to these variables. The factor $\sqrt{-g}$ appears because we are working in a curved spacetime. Regardless of whether the universe is flat or not, we can always put ourselves at the origin $r = 0$ with a roto-translation of coordinates, such that $\sqrt{-g}$ is constant along the coordinate change. Then the energy is $p_0^2 = E^2 = m^2 + a^2|\vec{p}|^2$. We employ the change of variables $q_i = q^i = \frac{1}{a}p_i = ap^i$ and obtain $d_3p/\sqrt{-g} = d_3p/a^3 = d^3qa^3/a^3 = d^3q$. Using the q variables, it is as if we were working in special relativity: the dispersion relation is $E^2 = m^2 + |\vec{q}|^2$, so $d^3q = q^2dqd\Omega_q = qEdEd\Omega$ by use of the usual relation $EdE = qdq$. The q^i have the physical significance of a co-moving momentum. In general, the $\vec{q}$ are the physical momenta and if we use these variables in our distribution, the statistical mechanics works out just as it does in special relativity, and we can mostly forget that we are in a curved space-time. Thus, we get the usual expression for the number density

$$n_i(T_i) = \frac{g_i}{2\pi^2} \int_m^\infty \frac{\sqrt{E^2 - m^2}E}{\exp[(E - \mu_i)/T_i] \pm 1} dE \tag{8.3}$$

Similarly we arrive to the usual expressions for the pressure and density of species i, consistently with the energy-momentum tensor (7.3),

$$\rho_i(T_i) = \frac{g_i}{2\pi^2} \int_m^\infty \frac{\sqrt{E^2 - m^2}E^2}{\exp[(E - \mu_i)/T_i] \pm 1} dE \tag{8.4}$$

$$P_i(T_i) = \frac{g_i}{6\pi^2} \int_m^\infty \frac{(E^2 - m^2)^{\frac{3}{2}}}{\exp[(E - \mu_i)/T_i] \pm 1} dE \tag{8.5}$$

These integrals could be evaluated numerically, but their expressions in certain limits are usually enough for any practical calculation. We summarize them in table 8.1.

Limit	B/F	n	ρ	P
$T \gg m, \mu$	F	$g\frac{3\zeta(3)}{4\pi^2}T^3$	$g\frac{7}{8}\frac{\pi^2}{30}T^4$	$\frac{\rho}{3}$
$T \gg m, \mu$	B	$g\frac{\zeta(3)}{\pi^2}T^3$	$g\frac{\pi^2}{30}T^4$	$\frac{\rho}{3}$
$\mu \gg T \gg m$	F	$g\frac{1}{6\pi^2}\mu^3$	$g\frac{1}{8\pi^2}\mu^3$	$\frac{\rho}{3}$
$T \gg m$ and $\mu < -T$	B,F	$\frac{g}{\pi^2}e^{\frac{\mu}{T}}T^3$	$\frac{3g}{\pi^2}e^{\frac{\mu}{T}}T^4$	$\frac{\rho}{3}$
$T \ll m$	B,F	$g(\frac{mT}{2\pi})^{\frac{3}{2}}e^{\frac{(\mu-m)}{T}}$	$(m+\frac{3T}{2})n$	nT

Table 8.1: Summary of limits of thermodynamical quantities for fermions (F) and bosons (B). $\zeta(z)$ is the Riemann Zeta function.

In the limit $T \gg m, \mu$ we can drop the chemical potential in the exponential and the mass, both in the integrand and in the limit of integration which we set to $E = 0$. With a suitable change of variables the integrals can be put in the form[4]

$$\int_0^\infty \frac{x^s}{e^x \pm 1} dx = \mp\Gamma(s+1)L_{i(s+1)}(\pm 1) \tag{8.6}$$

where $\Gamma(z)$ is the gamma function and $L_{i(s+1)}(z)$ is the polylogarithm function. The polylogarithm function has the square relationship property

$$L_{i(s)}(-z) + L_{i(s)}(z) = 2^{1-s}L_{i(s)}(z^2) \tag{8.7}$$

and evaluated at $z = 1$ is equal to the Riemann zeta function $L_{i(s)}(1) = \zeta(s)$. The particular values of the Riemann zeta function needed to complete the expressions are Apery's constant[3] $\zeta(3) \simeq 1.20205..$ and $\zeta(4) = \frac{\pi^2}{90}$. The results in table 8.1 follow.

Clearly this is the ultra-relativistic limit where the particles can be considered massless. Photons and neutrinos are in this situation, as well as electrons, at sufficiently early times. The first notable result is that the energy density of radiation is proportional to the fourth power of the temperature $\rho \propto T^4$. Combining this with the fact that $\rho \propto a^{-4}$ we can deduce that $T \propto a^{-1}$. This is not entirely correct because heat exchanges between species may alter the temperature dependence[5], as we will see.

The second limit to look at is the situation in which there is a very large chemical potential, compared to the temperature, which is itself much larger than the mass, thus we are still in an ultra-relativistic case. This can only happen for fermions. Now we are in the well known situation where every state is occupied up to $E = \mu$. In fact, we can approximate the Fermi-Dirac distribution as $\theta(\mu - E)$, and the integrals are easily evaluated.

In the converse case, when the chemical potential is negative $\mu < -T$, but the gas is still relativistic $T \gg m$, we can approximate the denominator as $\frac{1}{e^{\frac{E-\mu}{T}} \pm 1} \simeq e^{\mu/T}e^{-E/T}$ and reduce ourselves to integrals of the form

$$\int_0^\infty E^s e^{-\frac{E}{T}} dE \tag{8.8}$$

which are trivial. We note that in the standard ΛCDM cosmology, the chemical potentials

[5]Recall that the dependence $\rho \propto a^{-4}$ was derived from the continuity equation (7.21) for a *single* perfect fluid. When more than one fluid exists, one must consider energy exchanges amongst them.

are mostly zero, except for a very small value $\mu/T \simeq 10^{-9}$ for matter to dominate over anti-matter. We will return to this.

The final limit is the non-relativistic one, and perhaps the most important after the ultra-relativistic limit. Here $T \ll m$ and the integrals can be solved by making a non-relativistic substitution $E = m + \frac{mv^2}{2}$. The integral for the number density can be approximated as

$$n \simeq \frac{g_i}{2\pi^2} e^{-\frac{m-\mu}{T}} m^3 \int_0^\infty v^2 e^{-\frac{mv^2}{2T}} dv = g(\frac{mT}{2\pi})^{\frac{3}{2}} e^{\frac{(\mu-m)}{T}} \tag{8.9}$$

so we have reduced the Fermi-Dirac and Bose-Einstein distributions to the Maxwell distribution. Indeed the non-relativistic case is just a perfect gas, and we find usual expression for internal energy of a perfect gas as well.

When taking the total density of the universe it is obvious that relativistic species dominate, the contribution of non-relativistic species is exponentially suppressed. Therefore the total density of the universe can be expressed by neglecting the non-relativistic components as

$$\rho = \frac{\pi^2}{30} \tilde{g}(T) T_\gamma^4 \tag{8.10}$$

where we take the temperature of the photon as a reference temperature and $\tilde{g}$ represents the number of degrees of freedom of the relativistic species defined as

$$\tilde{g}(T) = \sum_{\text{bosons}} g_i (\frac{T_i}{T_\gamma})^4 + \frac{7}{8} \sum_{\text{fermions}} g_i (\frac{T_i}{T_\gamma})^4 \tag{8.11}$$

8.2 Chemical potential and particle-antiparticle asymmetry

We have tacitly supposed that only matter exists in the early universe. This indeed seems to be the case from experiment. The thermodynamical formulas are symmetric with respect to matter and antimatter and reactions such as $e^+ + e^- \leftrightarrow \gamma + \gamma$ can be shown to be efficient in the early universe, in the sense that the rate Γ is larger than the Hubble factor. The difference in densities between matter and antimatter is due to the presence of a chemical potential $\mu_{e^-} = -\mu_{e^+}$ which favors electrons over anti-electrons. The same can be said for baryons.

The chemical potential is the energy required to add a particle to a system. The defining thermodynamic relation is, in fact, for a single fluid

$$\mu \equiv \frac{\partial U}{\partial N}|_{S,V} \tag{8.12}$$

where U is the internal energy, N is the particle number and the derivative is taken on the manifold of constant entropy S and volume V. We can associate a chemical potential μ_i to every species. If a reaction

$$A + B \leftrightarrow C + D \tag{8.13}$$

is in equilibrium, namely that it happens efficiently in both directions, then the change in

internal energy of the system during each reaction is

$$\pm\, dU = \mu_A dN_A + \mu_B dN_B - \mu_C dN_C - \mu_D dN_D \tag{8.14}$$

since $dN_{A,B,C,D} = 1$ and energy is conserved, this implies

$$\mu_A + \mu_B = \mu_C + \mu_D \tag{8.15}$$

which holds *only if the reaction is efficient.*

What chemical potentials do the various species have? First consider the photon. Since there exist many efficient reactions which do not conserve photon number, for example Compton scattering with Bremsstrahlung $e^- + \gamma \leftrightarrow e^- + 2\gamma$, it follows that

$$\mu_\gamma = 0 \tag{8.16}$$

Matter and their respective anti-matter particles must have opposite chemical potentials. Although we won't prove it, at high temperature the reaction $e^+ + e^- \leftrightarrow \gamma + \gamma$ is efficient which implies

$$\mu_{e^-} = -\mu_{e^+} \tag{8.17}$$

In principle they could be zero, but they are not. It is not at present clear why the universe has preferred matter to anti-matter $\mu_{e^-} > 0$.

Let's take the number density (8.3) difference $n_{e^-} - n_{e^+}$. In the ultra-relativistic case we drop the mass and we Taylor expand the integrand with respect to μ/T around zero. Up to third order, we obtain integrals of the form

$$\int x^2 e^{nx} \frac{1}{(e^x+1)^m} dx \tag{8.18}$$

with $n < m$. Noticing that $e^x/(1+e^x)^m = -\frac{1}{m-1}\frac{d}{dx}\frac{1}{(1+e^x)^{m-1}}$ through successive integration by parts these integrals can be put in the form (8.6). The calculation is not particularly enlightening, and it involves the value of Riemann Zeta function $\zeta(2) = \frac{\pi^2}{6}$, so we simply state the result

$$n_{e^-} - n_{e^+} = \frac{gT^3}{6}\left(\frac{\mu_{e^-}}{T} + \frac{1}{\pi^2}(\frac{\mu_{e^-}}{T})^3\right) \tag{8.19}$$

In the non-relativistic limit it is easily seen that the asymmetry becomes exponentially suppressed. Using the result in table 8.1

$$n_{e^-} - n_{e^+} = 2g\left(\frac{mT}{2\pi}\right)^{\frac{3}{2}} e^{-\frac{m}{T}} \sinh\frac{\mu_{e^-}}{T} \tag{8.20}$$

so the asymmetry is exponentially suppressed as the electrons annihilate and very little excess remains. Assuming electrical neutrality in the universe, one can use the equality $n_p = n_{e^-} - n_{e^+}$, since the number of anti-protons is largely suppressed by the $\sim 1GeV$ mass. The ratio of protons to photons is well constrained by nucleosynthesis measurements[99] and

$$\frac{n_p}{n_\gamma} \sim 5 \cdot 10^{-10} \tag{8.21}$$

so that at nucleosynthesis, $T \sim 1MeV$, the ratio is

$$\frac{\mu_e}{T} \sim 5 \cdot 10^{-10} \tag{8.22}$$

The chemical potentials can generally be dropped in practical calculations. We note, however, that the chemical potential of neutrinos is not well constrained by experiment and may in principle be quite large (although one would have to explain why).

8.3 Entropy

Entropy turns out to be a key quantity in the evolution of the universe. Consider the fundamental thermodynamic relation for the content of the universe in thermal equilibrium

$$dU = -PdV + TdS + \mu dN \tag{8.23}$$

where the values have the regular thermodynamical meaning. Since there are many species, it would be more correct to write the chemical potential term as $\sum_i \mu_i dN_i$. Keeping track of this will not be very illuminating at the moment. We are considering the main component which is in thermodynamic equilibrium, such as photons and electrons, and neutrinos at early times. The universe may also contain other species, such as dark matter, which are not in thermodynamic equilibrium. The volume under consideration can be any proper-volume $V = a^3 V_0$, with V_0 being the comoving volume. We will assume that the expansion of the universe is slow enough so the evolution can be considered quasi-static. To be general, we may consider that the energy and number of the particles may change both due to internal thermodynamics and due to external action. An external action may be an exchange of energy with a species, such as dark matter, which is not in equilibrium. For example, although dark matter is not in equilibrium, it may decay, in some models, to standard model particles heating up the fluid. Such a process is of course irreversible.

Writing $U = \rho V$, $N = \rho V$ and $S = sV$ where s is the entropy density, we obtain

$$d(sa^3) = \frac{1}{T} d(\rho a^3) + \frac{P}{T} da^3 - \frac{\mu}{T} d(na^3) \tag{8.24}$$

We now want to apply the continuity equation (7.21) in its differential form $d(\rho a^3) = -Pda^3$. This imposes the Einstein equations on the thermodynamics, since the continuity equation comes from the Bianchi identities. However, this equation assumes there is no interaction of the fluid with other elements. If there were an energy transfer from some component of the universe not in equilibrium the equation would look like

$$d(\rho a^3) = -Pda^3 + \delta(qa^3) \tag{8.25}$$

where q is a density of heat. In any specific calculation we would have to determine the form of the function $q(a)$, which may depend on the physics involved. Using this modified continuity equation we obtain

$$d(sa^3) = \frac{\partial(qa^3)}{T} - \frac{\mu}{T} d(na^3) \tag{8.26}$$

Thus the entropy can change either by heat injection from another component of matter or by a change in the number of particles. In standard cosmology, there is no component which can heat up the universe, therefore the first terms is usually disregarded.

We'd like to find an explicit expression for the entropy density in terms of the distributions, or other known quantities. One way would be to begin with the definition of entropy in terms of microstates. This is the most physically correct way, but very lengthy. We will adopt a shorter route by taking the fundamental relation

$$dS = \frac{\rho + P}{T} dV + \frac{V}{T} \frac{\partial \rho}{\partial T} dT - \frac{\mu}{T} dN \tag{8.27}$$

The independent state variables in the above are V, T and N. The density, pressure and chemical potential are intensive quantities and cannot depend explicitly on the extensive variables V and N. By checking the integrability conditions $\frac{\partial^2 S}{\partial V \partial T} = \frac{\partial^2 S}{\partial T \partial V}$ and $\frac{\partial^2 S}{\partial T \partial N} = \frac{\partial^2 S}{\partial N \partial T}$ we obtain the two useful relations

$$\frac{dP}{dT} = \frac{\rho + P}{T} \tag{8.28}$$

$$\frac{d\mu}{dT} = \frac{\mu}{T} \tag{8.29}$$

We define the entropy density as

$$s = \frac{\rho + P - \mu n}{T} \tag{8.30}$$

Using the found relations for dP/dT and $d\mu/dT$ we find that

$$\frac{ds}{dT} = -\frac{\mu}{T} \frac{\partial n}{\partial T} \tag{8.31}$$

The entropy so defined satisfies the same conservation equation for the thermodynamical entropy (8.26); they are the same up to a constant. From now on (8.30) will be our entropy.

The main result is that we have a well defined expression based on fundamental quantities and a conservation equation. In usual cosmological settings, we have seen that $\frac{\mu}{T}$ is very small. In addition, for non-relativistic species there is no change in particle number (as the thermal energies would be too small to create a new particle). This implies that for most of the history of the universe

$$d(sa^3) = 0 \tag{8.32}$$

the total entropy is constant. Alternatively

$$s \propto a^{-3} \tag{8.33}$$

This is the functional formula through which we can relate the change to thermodynamical quantities with the expansion of the universe. For an ultra-relativistic species, $T \gg m, \mu$,

we obtain (see table 8.1).

$$s = \frac{2g\pi^2}{45}T^3 \begin{cases} 1 & \text{bosons} \\ \frac{7}{8} & \text{fermions} \end{cases} \tag{8.34}$$

For a non-relativistic species

$$s \simeq gm(\frac{mT}{2\pi^2})^{\frac{3}{2}}e^{\frac{\mu-m}{T}} \tag{8.35}$$

The entropy contribution of non-relativistic species is very small. Because of this, their contribution is usually dropped in the computation of the total entropy

$$s = \frac{2\pi^2}{45}T_\gamma^3\left(\sum_{\text{bosons}} g_i(\frac{T_i}{T_\gamma})^3 + \frac{7}{8}\sum_{\text{fermions}} g_i(\frac{T_i}{T_\gamma})^3\right) \tag{8.36}$$

The sum extends only to relativistic species. We have pulled out the temperature of the photons T_γ which we will use as reference temperature. In fact, decoupled species may have a different temperature. When we refer to the temperature of the universe, we will take this to mean the temperature of photons. Now we introduce the total number of relativistic degrees of freedom of the universe as

$$g^*(T) = \sum_{\text{bosons}} g_i(\frac{T_i}{T})^3 + \frac{7}{8}\sum_{\text{fermions}} g_i(\frac{T_i}{T})^3 \tag{8.37}$$

$$s = \frac{2\pi^2}{45}g^*(T_\gamma)T_\gamma^3 \tag{8.38}$$

Which one must take care not to confuse with the similar quantity (8.11), in which the ratio of temperatures contributes with the fourth power. In particular, $\tilde{g}$ and g^* differ after neutrinos decouple, since the temperature of the the neutrinos $T_\nu \neq T_\gamma$, as we shall see shortly.

In this manner, the constant entropy condition becomes

$$(g^*)^{\frac{1}{3}}T \propto a^{-1} \tag{8.39}$$

The value of g^* is constant whenever no species is transitioning from being relativistic to non-relativistic. Of course, among the transition the above equation is not exact, but it is much earlier and after the transition. The temperature generally decreases as a^{-1}, except when a species of particles is becoming non-relativistic. An example are the electrons which become non-relativistic as $T \sim 0.5MeV$. When this happens, the energy density of the electrons decreases drastically and this energy heats up the rest of the plasma. In fact, since $dS = \partial Q/T$ the reduction in entropy of the electrons is related to a heat loss towards the other components of the fluid.

8.4 Decoupled species

A species may only be in thermal equilibrium with the main fluid of the universe, namely the photons, if its reaction rate Γ is larger than the Hubble factor $\Gamma \gtrsim H$. When the reaction rate slows, a full treatment, involving the Boltzmann equation is required. However, with

some general arguments one can draw many correct conclusions without a detailed treatment. We will suppose in this paragraph, that a species i is in thermal equilibrium until a decoupling time t_D with temperature T_D and has afterwards a negligible interaction. At decoupling its phase space distribution will be a Bose-Einstein or Fermi-Dirac

$$f(E,t) = [\exp \frac{E-\mu}{T_D} \pm 1]^{-1} \tag{8.40}$$

Afterwards there are no interactions that keep the distribution at the equilibrium one. Instead the particles are freely propagating along the geodesics. A particle with energy E_D at decoupling has an energy $E = E(E_D, t)$ at any later time and so the distribution becomes

$$f(E,t) = [\exp \frac{E_D(E)-\mu}{T_D} \pm 1]^{-1} \tag{8.41}$$

since free motion does not change the phase space volume element. If a species is ultra-relativistic when it decouples, then the energy of a single particle is simply redshifted away as $E \propto a^{-1}$. Neglecting the chemical potential, the distribution becomes

$$f(E,t) = [\exp \frac{E}{T(t)} \pm 1]^{-1} \tag{8.42}$$

where the temperature is $T(t) = \frac{a_D T_D}{a}$, a_D being the scale factor at decoupling. A relativistic decoupled species maintains its Bose-Einstein or Fermi-Dirac distribution with a temperature that decreases as a^{-1}. This may be the case of neutrinos, which decouple at a temperatures of a few MeVs. It must be noted, and this may be relevant for neutrinos, that as T becomes comparable to the small mass m_ν it is no longer true that $E \propto a^{-1}$ but the distribution stays in the form (8.41). The more complicated dependence on energy implies that the shape of the distribution is no longer a Fermi-Dirac or Bose-Einstein.

For non-relativistic species $E \simeq m + \frac{p^2}{2m}$, one can easily show that the geodesic equation, keeping terms up to order p^1, implies $p \propto a^{-2}$. If a species decouples non-relativistically with $\mu = 0$ it is easy to show that its distribution remains Maxwell-Boltzmann with a temperature $T = (\frac{a_D}{a})^2 T_D$.

Based on the thermodynamics we developed thus far, let's see what we can say about neutrinos present in the universe. There are three known families of neutrinos, ν_e, ν_μ and ν_τ. They can be considered ultra-relativistic until at least very recently and possibly even today, depending on their mass. They may be Dirac or Majorana but this cannot be important in the massless limit. In fact for $m = 0$ a Dirac particle is equivalent to two Majorana particles, but only the left chirality neutrinos interact with the standard model and can be in equilibrium. So if neutrinos are Dirac, there are two interacting degrees of freedom, one for the neutrino and one for the anti-neutrino. If they are Majorana there are again two degrees of freedom.

Neutrinos decouple at a temperature of about $\sim 1MeV$. We will work in approximation where they decouple instantaneously. Once they decouple at T_D , with scale factor a_D, their temperature will decrease as $T_\nu = \frac{a_D}{a} T_D$. Some time later at around $m_e = 0.511MeV$ the electrons will start to become non-relativistic. As this happens, the entropy of the neutrinos is conserved separately, since they have decoupled and the temperature T_ν is not

altered. The total entropy of electrons, anti-electrons and photons is also conserved. Before electrons become non-relativistic, the number of interacting degrees of freedom is

$$g^*_{\text{before}} = g^*(T \gg m_e) = 2 + \frac{7}{8} \cdot 2 \cdot 2 \tag{8.43}$$

where we counted two degrees of freedom for the photon, a boson, and two each for the electron and anti-electron, fermions. Well after the electrons became non-relativistic $g^*_{\text{after}} = 2$. By conservation of entropy the temperature of the photon fluid is

$$T_\gamma = \left(\frac{g^*_{\text{before}}}{g^*_{\text{after}}} \right)^{\frac{1}{3}} T_D \frac{a_D}{a} \tag{8.44}$$

where we used (8.39) $g^* T^3 a^3 = \text{const}$. This now allows us to establish the relationship between the temperature of the neutrinos and of the photons.

$$T_\nu = \left(\frac{4}{11} \right)^{\frac{1}{3}} T_\gamma \tag{8.45}$$

Today the temperature of the photons is, of course, that of the CMB which is measured to be $T_\gamma = 2.7K = 2.3 \cdot 10^{-4} eV$. So long as the mass of the neutrino is not larger than this temperature, we can hold this formula to be true. We conclude that there exists a Cosmic Neutrino Background (CνB). Unfortunately the energy of these neutrinos is too small to be realistically detected in the near future. Using the formulas for the density of a relativistic gas in table 8.1 one arrives at the density parameters for neutrinos

$$\Omega_\nu = \frac{7}{8} (\frac{4}{11})^{\frac{4}{3}} N_\nu \Omega_\gamma \tag{8.46}$$

N_ν being the number of neutrino families. In the standard model $N_\nu = 3$. When one repeats the calculation more accurately, including the finite time neutrinos take to decouple, it turns out that the above formula is correct, provided one replaces $N_\nu \to N_{eff}$, the *effective number of neutrinos*. The predicted value is then $N_{eff} = 3.046$. This corresponds to a slightly higher value of the neutrino density than we have estimated with an instant decoupling. The difference is mainly due to neutrinos not having completely decoupled by the time the electrons start to become non-relativistic, so that the electrons will actually heat up the neutrino fluid slightly.

Part III

Perturbed universe

We have examined the space-time structure and the thermodynamics of a universe which is isotropic and homogeneous. This has followed from the assumption of a cosmological principle. As we had noted, the principle implicitly needed some distance scale above which it would work. The universe at smaller scales is quite evidently not homogeneous. Clusters, galaxies and stars exist and these could not have formed from a perfectly homogeneous and isotropic initial state. In fact, with the non-relativistic limit of our thermodynamic formulas, the universe today would simply be a fluid of photons with $T_{CMB} = 2.7K$, all other matter having annihilated away. Our understanding of the universe must therefore involve an understanding of the in-homogeneities and the processes that happen out of equilibrium. This is the cornerstone of modern cosmology.

9 The Boltzmann Equation

The starting point for any calculation involving non-equilibrium thermodynamics is the Boltzmann equation. Let's take a phase space distribution $f(x^\mu, p_\nu)$. Liouville's theorem states that *the distribution must be constant over every trajectory in phase space*. In ordinary Hamiltonian mechanics this is a trajectory through time, in general relativity the "time" coordinate x^0 can considered to be a coordinate in phase space and we parametrize a trajectory through some affine parameter s which we can take to be the proper time, except for trajectories of massless particles. Along a trajectory we may write

$$\frac{df}{ds} = \dot{x}^\mu \frac{\partial f}{\partial x^\mu} + \dot{p}_\mu \frac{\partial f}{\partial p_\mu} = 0 \tag{9.1}$$

which is *Liouville's equation*. The dot indicates a derivative with respect to s and $p^\mu = \dot{x}^\mu$. $\dot{p}_\mu$ can be found by the geodesic equation $\dot{p}^\mu = -\Gamma^\mu_{\alpha\beta} p^\alpha p^\beta$. What we are describing with these equations is an ensemble which does not interact with itself, or anything else, and simply follows the geodesics of space-time. Next, we include instantaneous collisions of the particles amongst themselves and with a different species, which will have a different distribution g. To do this, we supplement the Liouville equation with a collision term[143]

$$\frac{df}{ds} = p^\mu \frac{\partial f}{\partial x^\mu} - \Gamma^\mu_{\alpha\beta} p^\alpha p^\beta \frac{\partial f}{\partial p^\mu} = (\frac{\partial f}{\partial s})_C \tag{9.2}$$

The collision terms represent the change per unit s of the distribution in the phase space element $d^4x d_4p$. If the fluid has many species, every species has its own distribution, equation and collision term. The collision term, having many terms , can be tricky to write out correctly, but is by no means mysterious.

Let's connect the collision term with a matrix element $\mathcal{M}_{i\to j}$ calculated in quantum field theory, where i and f indicate the initial and final states respectively. To this end, we use a not explicitly covariant formalism and we give preferential treatment to time $t = x^0$. Given

that the collisions are microscopic, we expect the usual special relativity of quantum field theory to suffice in their calculation. We calculate the collision factor with respect to time, $(\frac{\partial f}{\partial t})_C$. Let's analyze a process $a_1 + a_2 + ... + a_n \rightarrow c + b_1 + b_2 + ... + b_m$ where c is the species who's current collision term we are studying. We will denote f_i the distributions of the incoming particles a_i , $\vec{q}_i$ their momenta, which we will integrate over, and E_{q_i} their energies. For outgoing particles, we will denote g_i the distribution for the species b_i, $\vec{p}_i$ their momenta , E_{p_i} their energies, and simply f for the distribution of c particles, with momenta $\vec{k}$ and energy E. The energies are all fixed by the on-shell relation $E^2 = p^2 + m^2$. Assume the scattering takes place at $\vec{x}$ and a time t. We need to know the number of particles in the elementary volume of phase space around $\vec{x}$ and $\vec{q}_i$. For n particles, this would be given by the n particle distribution $f_n(\vec{x}_1, \vec{x}_2, ..., \vec{q}_1, \vec{q}_2)$ which could account for correlations between particles. If we adopt the *molecular chaos hypothesis*, also known as *Stosszahlansatz*[28], we assume there are no correlations between particles long before or after the collision. The n particle distribution may be written simply as the product of the one-particle distributions. Hence, in the elementary volume of phase space the number of particles of each incoming species is

$$dN_{a_1} = f_i(\vec{x}, \vec{p}, t) \frac{d^3x d^3p}{(2\pi)^3} \tag{9.3}$$

From quantum field theory, the differential probability of the process taking place during a time T , in a volume V and with momenta $\vec{p}_i$ in the final state is[189]

$$\begin{aligned} dP &= (2\pi)^4 \delta^4(\sum_{i=1}^{n} q_i^\mu - \sum_{f=1}^{m} p_f^\mu - k^\mu) |\mathcal{M}_{i\rightarrow f}|^2 \times \\ &\frac{T}{V^{n-1}} \frac{1}{2E} \prod_{i=1}^{n} \frac{1}{2E_{q_i}} \prod_{f=1}^{m} \frac{1}{2E_{p_f}} \times \\ &\frac{d^3k}{(2\pi)^3} \prod_{f=1}^{m} \frac{d^3p_f}{(2\pi)^3} \end{aligned}$$

The time and volume in the field theory formula have the physical meaning of the time over which the scattering takes place and the size of the volume it takes place in. Assuming that the interaction has a small enough typical length and time, none other than the instantaneous collision approximation, we can identify V with the volume d^3x and T with the time dt. This is the usual idea of taking an element which is macroscopically small but contains many microscopic particles. The probability dP refers to only one 1 incoming particle per species a_i. Therefore the total number of interactions happening in the volume $V = d^3x$ during a time $T = dt$ is

$$d\tilde{N}_{int} = dP \prod_{i=1}^{n} dN_{a_i} \tag{9.4}$$

which we won't write in full. We note that there are exactly n terms $V = d^3x$ in the numerator, while $n-1$ terms V in the denominator. Thus, the number of interactions dN_{int} is proportional to $d^3x dt$, as it should be.

In addition, we must consider that the final states may already be occupied. If the species

is a fermion, this means we cannot put a particle in that particular final state. In the case of a boson, a transition to the final state is enhanced. Therefore we multiply by $(1 \pm g_i(\vec{x}, \vec{p}_i))$ for each particle in the final state, including particle c, where the$-$ is for fermions and the $+$ for bosons. We marginalize on the momentum all the final states with exception of particle c, by integrating the d^3p_i. Finally, we sum over, integrating, all initial momenta q_i since we are interested in all possible collisions which leave us with a new c particle with momenta $\vec{k}$.

$$dN_{int}(\vec{k}) = \int d\tilde{N}_{int}(\vec{q}_i, \vec{p}_i, \vec{k}) \tag{9.5}$$

The quantity dN_{int} now has the significance of the number of collisions which produce an extra particle of momenta $\vec{k}$ at position $\vec{x}$ during the time dt . We may identify

$$\frac{dN_{int}}{dt} = (\frac{\partial f}{\partial t})_C \frac{d^3x d^3k}{(2\pi)^3} \tag{9.6}$$

the change in the distribution is equal to the number of collisions normalized by the phase space volume. Explicitly, the collision term for the process $a_1 + a_2 + ... + a_n \to c + b_1 + ... + b_m$ is given by

$$\begin{aligned}
(\frac{\partial f}{\partial t})_C \quad = \quad & \prod_{i=1}^{n} \int \frac{d^3q_i}{(2\pi)^3} \prod_{f=1}^{m} \int \frac{d^3p_f}{(2\pi)^3} \times \\
& \prod_{i=1}^{n} f_i(q_i) \left(\prod_{f=1}^{m} 1 \pm g_f(p_f) \right) (1 \pm f(k)) \times \\
& (2\pi)^4 \delta^4 (\sum_{i=1}^{n} q_i^\mu - k^\mu - \sum_{f=1}^{m} p_f^\mu) |\mathcal{M}_{i\to f}|^2 \times \\
& \frac{1}{2E} \prod_{i=1}^{n} \frac{1}{2E_{q_i}} \prod_{f=1}^{m} \frac{1}{2E_{p_f}}
\end{aligned}$$

This expression is nearly symmetric if we change the final and initial state. If the process above exists, so does the inverse $c + b_1 + ... + b_m \to a_1 + ... + a_n$. For the opposite process one exchanges the initial particle distributions with the final ones. The matrix element may differ in principle, although for a CP (charge, parity) invariant, or equivalently time reversal invariant, process $\mathcal{M}_{i\to f} = \mathcal{M}_{f\to i}$. We will not study any CP violating process and take this to be true henceforth. If the particle c is in the final state in the first process, the inverse process will reduce its density in phase space, so we add an overall $-$ sign. We give the collision term for the two way process $a_1 + a_2 + ... + a_n \leftrightarrow c + b_1 + ... + b_m$

$$
\begin{aligned}
(\frac{\partial f}{\partial t})_C \quad = \quad & \prod_{i=1}^{n} \int \frac{d^3 q_i}{(2\pi)^3} \prod_{f=1}^{m} \int \frac{d^3 p_f}{(2\pi)^3} \times \\
& \left[\prod_{i=1}^{n} f_i(q_i) \left(\prod_{f=1}^{m} 1 \pm g_f(p_f) \right) (1 \pm f(k)) - f(k) \prod_{f=1}^{m} g_f(\vec{p}_f) \prod_{i=1}^{n} (1 \pm f_i(\vec{q}_i)) \right] \times \\
& (2\pi)^4 \delta^4 (\sum_{i=1}^{n} q_i^\mu - k^\mu - \sum_{f=1}^{m} p_f^\mu) |\mathcal{M}_{i \to f}|^2 \times \\
& \frac{1}{2E} \prod_{i=1}^{n} \frac{1}{2E_{q_i}} \prod_{f=1}^{m} \frac{1}{2E_{p_f}}
\end{aligned}
\tag{9.7}
$$

Obviously, we must sum over all relevant processes that involve the particle c.

Given the the collision term with respect to time it is immediate to write the Liouville equation with time, instead of s

$$
\frac{df}{ds} = \frac{df}{dt} \frac{1}{E} = \frac{1}{E} (\frac{\partial f}{\partial t})_C \tag{9.8}
$$

so in practice we will use coordinate time when dealing with the Boltzmann equation.

The Boltzmann equation is in principle a partial differential equation of f and can be very hard to solve in the most general case. We will assume that the interactions of a species with itself is fast enough to keep the form of f to functionally be either a Bose-Einstein or Fermi-Dirac distribution, but with a temperature, and possibly chemical potential, which depend on position, momentum and time. With some additional assumptions of symmetry and small deviation from an equilibrium, we will be able to make progress by reducing the equations to a series a ordinary differential equations on T and μ.

Finally we note that one can consider f a function of any generalized coordinates (q^i, k_i) related to (x^i, p_j) by some transformation, even if such a transformation is not canonical. In that case the right hand side of the Boltzmann equation may be $\frac{df}{ds} = \dot{q}^i \frac{\partial f}{\partial q^i} + \dot{k}_i \frac{\partial f}{\partial k_i} + \frac{\partial f}{\partial s}$. The $\dot{q}^i$ and $\dot{k}_i$ will only be given by Hamilton's equations if the transformation is canonical.

10 Relics and decay out of equilibrium

10.1 Cold relics

To illustrate the power of the Boltzmann equation, we study a simple example in which a very massive particle X, which interacts little with the other components of the universe and is stable, falls out of equilibrium[80, 143]. Such a situation may indeed be the case of cold massive dark matter. In fact, if a very massive particle were to stay in equilibrium until today, its equilibrium density would be exponentially suppressed and could not account for any large measured quantity. In order for it to be a *relic* from the early universe today, it must have fallen out of equilibrium when its density was not yet suppressed. It is said to be cold, in the sense that it is non-relativistic today.

Let's examine the left hand side of the Boltzmann equation (9.2) in a situation where the

distribution f is isotropic and homogeneous. This implies that f only depends on the energy $E = q^2 + m^2$, where $q^2 = g_{ij}p^i p^j$[6], and on time so

$$\frac{df}{dt} = \frac{\partial f}{\partial t} - H\frac{q^2}{E}\frac{\partial f}{\partial E} \tag{10.1}$$

Let's integrate on the 3-momenta $\frac{d^3q}{(2\pi)^3}$. The term $\int \frac{d^3q}{(2\pi)^3}\frac{\partial f}{\partial t} = \frac{d}{dt}\int \frac{d^3q}{(2\pi)^3} f = n'$, and the second one is equal to $3Hn$ through an integration by parts in E. n is the number density by definition, so we don't need to know the peculiarities of f at this stage. The left hand side of the Boltzmann equation, or the integrated Liouville term for a homogeneous and isotropic ensemble is,

$$a^{-3}\frac{d}{dt}(a^3 n) \tag{10.2}$$

If there is no collision term, we simply obtained a continuity equation which conserves the number of particles per comoving volume.

We now suppose the X particles can annihilate into light standard model particles β according to a reaction

$$X + X \leftrightarrow \beta + \beta \tag{10.3}$$

These β particles may be photons or electrons. We will assume that the interaction rate of Xs is much smaller than a standard model interaction rate, so that the β are always considered to be in equilibrium with the rest of the universe, through regular interactions. We take the collision term, integrated over $\frac{d^3q}{(2\pi)^3}$, the momenta of one incoming X. We will denote p^μ the momenta of the other X and k^μ, k'^μ the momenta of the two β. Then the right hand side of the Boltzmann equation is

$$\int \frac{d^3p d^3q d^3k d^3k'}{(2\pi)^{12} 16 E_p E_q E_k E_{k'}} (2\pi)^4 \delta^4(p^\mu + q^\mu - k^\mu - k'^\mu) |\mathcal{M}_{XX\to\beta\beta}|^2 \times \Big[f_\beta(k) f_\beta(k') - f_X(p) f_X(q)\Big]$$

where we will be neglecting the enhancement/suppression factors $(1 \pm f)$. Other than that, this equation is exact. We shall now make a few reasonable approximations to put this in a more tractable form. We shall assume that the chemical potential for the β particles is negligible and that $f_\beta(k) \simeq e^{-E/T}$. On the other hand we will suppose that $f_X(p) \simeq e^{\mu/T} e^{-E/T}$ where the chemical potential μ is not the equilibrium one and is an independent variable we could solve for. We shall see we won't need to solve for μ exactly.

Next, we note that due to the Dirac Delta in the integral, $E_k + E_{k'} = E_p + E_q$, so the term $f_\beta(k) f_\beta(k') = e^{-(E_p+E_q)/T} = f_X^{eq}(p) f_X^{eq}(q)$ where $f_X^{eq}(p)$ is the distribution the Xs would have *if they were in thermodynamic equilibrium*, with $\mu = 0$. The astute reader would have already made the connection that the collision factor is closely related to a thermally

[6] As we had noted when discussing equation (8.2), we may use the physical momenta $q^i = ap^i$ instead of the canonical ones and the thermodynamics looks like the usual special relativistic one. Thus this q^i is the same kind of momentum which appears in the collision term.

averaged cross section of the process. Indeed

$$n_{eq}^2 < \sigma v >_{eq} = \int \frac{d^3p d^3q d^3k d^3k'}{(2\pi)^{12} 16 E_p E_q E_k E_{k'}} (2\pi)^4 \delta^4(p^\mu + q^\mu - k^\mu - k'^\mu) |\mathcal{M}_{XX\to\beta\beta}|^2 f_X^{eq}(p) f_X^{eq}(q) \tag{10.4}$$

is the thermally averaged cross section times velocity over the equilibrium distribution. The right hand side can now be put in a simple for with this relation and noting that $n = e^{\mu/T} n_{eq}$. The Boltzmann equation for an annihilating particle can be written as

$$\frac{dn}{dt} + 3Hn = - < \sigma v > (n^2 - n_{eq}^2) \tag{10.5}$$

Now we simply must solve for the number density. It is not as easy as it seems, since there is a lot of physics, and therefore time dependence, in $< \sigma v >$. We now introduce a very useful quantity, the *abundance of a species* i

$$Y_i = \frac{n_i}{s} \tag{10.6}$$

since both n_i and s decrease as a^{-3} for non-relativistic, or non-interacting, particles, the abundance is the number of particles per comoving volume, up to a multiplicative constant. It is easy to note that

$$s \frac{d}{dt} Y = \frac{dn}{dt} + 3Hn \tag{10.7}$$

Furthermore, since the dependence of thermodynamic quantities is directly on temperature, rather than time, it's useful to work with a variable which represents that. So we define

$$x = \frac{m}{T} \tag{10.8}$$

where m is some arbitrary energy scale useful for the problem. We shall choose m to be the mass of the X. Let's assume we are deep in the radiation era, since a relic must have fallen out of equilibrium very early. Then $\frac{dx}{dt} = Hx$, since $T \propto a^{-1}$, so long as there is no change in the number of relativistic degrees of freedom g^*. In terms of Y and x equation (10.5) can be written as

$$\dot{Y}_X = - < \sigma v > \frac{s}{xH} (Y_X^2 - Y_{eq}^2(x)) \tag{10.9}$$

where the dot indicates a derivative with respect to x . Alternatively, through the difference from equilibrium $\Delta = Y_X - Y_{eq}(x)$

$$\dot{\Delta} = -\dot{Y}_{eq} - < \sigma v > \frac{s}{xH} \Delta(\Delta + 2Y_{eq}) \tag{10.10}$$

Taking the entropy density (8.38), the Hubble factor expressed via the Friedmann equation and the total energy density

$$H^2 = \frac{8\pi G}{3} \frac{\pi^2}{30} \tilde{g} T^4 \tag{10.11}$$

we define

$$\lambda = < \sigma v > \sqrt{\frac{\pi}{45G}} \frac{g^*}{\tilde{g}^{\frac{1}{2}}} m \tag{10.12}$$

and write the equation again as

$$\dot{\Delta} = -\lambda(x)\frac{1}{x^2}\Delta(\Delta + 2Y_{eq}(x)) \tag{10.13}$$

In principle, supplemented with an explicit form for $< \sigma v >$ we could solve this equation. $< \sigma v >$ depends on the temperature through the equilibrium distribution over which it is defined. In likely situations its dependence on x is of the form $a + bx^{-1}$. All the ingredients now exist to solve this equation numerically.

A few conclusions can be observed immediately. When the abundance is at equilibrium, $\Delta = 0$, there is no instantaneous change in the relative abundance $\dot{\Delta}$. However, $Y_{eq}(x)$ will depend on x (via T) and the actual abundance will only be able to track the equilibrium value so long as λ/x^2 is large enough. In the limit of large x, when $T \ll m$, the term in $Y_{eq}(x)$ can be neglected and the equation reduces to

$$\dot{\Delta} = -\lambda(x)\frac{\Delta^2}{x^2} \tag{10.14}$$

which can be integrated analytically, and $\Delta \simeq Y_X$ in this limit. This formula can help in a numerical integration.

Qualitatively we understand that the smaller $< \sigma v >$is at $T \sim m$ the larger the relic density will be. In fact, a smaller cross section means the actual abundance will abandon the equilibrium one sooner. Since the equilibrium abundance is decreasing, we will be left with a larger final density. This can be confirmed by numerical and analytical estimates so that $Y_0 \propto (< \sigma v >_{T=m})^{-1}$. Therefore, species which have very small cross sections may have large relic densities today. This is a powerful idea for what dark matter could be, in fact one class of models for dark matter is known as Weakly Interacting Massive Particles (WIMP) which have the characteristics in mass and cross section we have described above.

10.2 Hot Relics

We can discuss what happens when a stable species X exits equilibrium while it is still ultra-relativistic[143]. In that case it is called a *hot relic*. A real example in cosmology is the neutrino, which we discussed in section 8.4. Assuming an annihilation channel to lighter standard model particles exists, just like in the case of the cold relic we discussed, one can arrive to the equation (10.9) in the same manner. In the case of neutrinos, where ν annihilates with the anti-neutrino $\bar{\nu}$ the modifications are straightforward. In the ultra-relativistic case, the equilibrium abundance is actually constant. In fact n and s both scale with T^3, assuming there is no change in relativistic degrees of freedom in the mean time. This implies that the details of the freeze out for a hot relic are largely unimportant. The abundance will already be the equilibrium abundance long before freeze out, and that is constant, so long after decoupling the abundance is still the equilibrium one. As long as the mass of X is not larger than the temperature of the photons today, the same arguments as for the neutrino in section 8.4 apply.

If the mass of X is larger than the temperature of the universe today, the distribution is no longer Bose-Einstein or Fermi-Dirac. In that case the relic density today is $\rho_X = m_X s_0 Y_{X0}$

where s_0 is the entropy density today and Y_0 the relic abundance, which we can take to be equal to $Y_X(T_f)$ where T_f is the freeze-out temperature. The temperature of photons today is $T_{CMB} = 2.7K = 2.3 \cdot 10^{-4} eV$, so for any particle more massive than that we can calculate the current density parameter as

$$\Omega_X = m_X s_0 Y(T_f) \frac{1}{\rho_c} \tag{10.15}$$

Plugging everything in we find

$$\Omega_X h^2 = \frac{1}{53eV} m_X \epsilon_{FB} \frac{g_0^* g_X}{g^*(T_F)} \tag{10.16}$$

To easily compute the numerical factor one can refer to appendix A. ϵ_{FB} is a factor equal to 1 if X is a boson, and $3/4$ if X is a fermion. The relativistic degrees of freedom today are $g_0^* = 2 + 3 \cdot 2\frac{4}{11}\frac{7}{8} = 3.91$, counting the neutrinos with their reduced temperature $\frac{T_\nu^3}{T_\gamma^3} = \frac{4}{11}$. Note that even if the neutrinos have become non-relativistic, since their entropy is conserved separately it is still equal to its value at decoupling, scaled by a^{-3}. We can immediately set a bound on the mass of these particles with the lax assumption that $\Omega_X h^2 < 1$ (recall that $h \simeq 0.7$). Then a bound on the mass is

$$m_X < 13.5eV \cdot \frac{g^*(T_f)}{g_x \epsilon_{FB}} \tag{10.17}$$

As noted, the only dependence on the cross section with the standard model particles is in the freeze-out temperature, and the dependence on the temperature is very weak, since $g^*(T)$ is a constant almost always. For a neutrino with a small mass (which we shall assume for simplicity is the same between all flavors), and is larger than $\sim T_{CMB}$, the freeze-out happens when $g^* = \underbrace{2}_{\gamma} + \underbrace{2 \cdot 2 \cdot \frac{7}{8}}_{e^-,e^+} + \underbrace{3 \cdot 2 \cdot \frac{7}{8}}_{\nu} = 10.75$ and

$$\Omega_\nu h^2 = \frac{m_\nu}{91.5eV} \tag{10.18}$$

The present density of the neutrino background can only be measured indirectly. We can assume a bound $\Omega_\nu < 0.1$ and using $h = 0.67$ one obtains that $m_\nu \lesssim 5eV$. Thus cosmology gives us a useful upper limit on the mass of the neutrino.

The present relic density (A.2) is inversely proportional to $g^*(T_f)$. This means that hot relics which abandon equilibrium sooner have a smaller relic density (as opposed to cold relics which have a larger density if they decouple sooner). When all the degrees of freedom of the standard model are relativistic at $T \gtrsim 200GeV$ and even the top quark is ultra-relativistic, the value $g^* \sim 107$.

10.3 Out of equilibrium decay

In the standard cosmological picture, the total entropy of the primordial universe is constant, the evolution is adiabatic. We have proved that the entropy density $s \propto a^{-3}$. Indeed, this fact is often taken for granted and we find it easier to deal with abundances $Y = n/s$

which is constant for a species whose total particle number is conserved. We had pointed out that a non-equilibrium species which exchanges energy (heat) with the main primordial fluid of photons and electrons can cause the entropy to increase, $dS = \frac{\partial Q}{T}$. If this were to happen it would change how we relate primordial abundances, and densities, with the same today. In the ΛCDM model this does not happen. It is believed to happen following the hypothesized inflation, and before the "standard" evolution of the universe, a process known as *reheating*. In that case, the fields which have caused inflation decay into standard model particles, heating up the plasma.

Let's consider a massive *unstable* particle X which has exited equilibrium very early in the universe when it was non-relativistic. We can consider it a cold relic even at temperatures $T > 10MeV$ before the neutrinos have decoupling. We could calculate its abundance Y_X as explained in section (10.1), we shall just assume it has some non-negligible value. If X is unstable it may decay to standard model particles with a lifetime τ_X. The number density n_X then satisfies the equation

$$\frac{dn_X}{dt} + 3Hn_X = -\frac{1}{\tau_X} n_X \tag{10.19}$$

This equation can of course be derived from the Boltzmann equation, considering some process $X \rightarrow l + l$ where l are some standard model particles. We assume the reverse process, where standard model particles interact to produce an X is thermally suppressed. If the X is non-relativistic, the energy density is $\rho_X = m_X n_X$ and the solution to the above equation is

$$\rho_X = \rho_{Xi} (\frac{a_i}{a})^3 e^{-\frac{(t-t_d)}{\tau_X}} \tag{10.20}$$

where we have chosen a time $t = t_d$ when the density of X is ρ_{Xi} and the scale factor is a_i. Conversely, the density for radiation is given by the continuity equation with the injection of heat. By conservation of energy it must follow

$$\frac{d\rho_r}{dt} + 4H\rho_r = \frac{1}{\tau_X} \rho_X \tag{10.21}$$

Note that this formula is only valid as long as the degrees of freedom $\tilde{g}$ is constant, since then we can use $P_r = \rho_r/3$. If $\tilde{g}$ is changing, the system of equations must be more complex, accounting for all the different species. We will avoid this complication. A solution to the equation for ρ_r can be found and is

$$\rho_r = \rho_{ri} (\frac{a_i}{a})^4 + \frac{\rho_{Xi}}{\tau_X} \int dt (\frac{a_i}{a})^3 e^{-\frac{(t-t_d)}{\tau_X}} \tag{10.22}$$

Augmented with the Friedmann equation $H^2 = \frac{8\pi G}{3}(\rho_r + \rho_X)$ we know have three equation describing the evolution of the scale factor during the decay, in three variables a, ρ_r, ρ_X. Thus we could solve them numerically. Once they are solved, we can use $\rho_r = \frac{\pi^2}{30} \tilde{g} T_\gamma^4$ to invert and find the temperature of the universe T_γ as a function of time. Let's focus on the entropy. Assuming the process is quasi-static, the decay time τ_X is longer than the thermalization time of the fluid, so we can assume the photon-electron-neutrino fluid to be

in equilibrium, its infinitesimal change in entropy in a comoving volume is

$$dS = \frac{\partial Q}{T_\gamma} = -\frac{1}{T_\gamma}\frac{d(a^3\rho_X)}{dt}dt = \frac{\rho_{Xi}}{T_\gamma}a_i^3\frac{1}{\tau_X}e^{-\frac{(t-t_d)}{\tau_X}}dt \tag{10.23}$$

So that the change in density across the process is

$$\Delta S = \frac{\rho_{Xi}a_i^3}{\tau_X}\int_{t_d}^{\infty} e^{-\frac{(t-t_d)}{\tau_X}}\frac{dt}{T_\gamma(t)} \tag{10.24}$$

Obviously the precise value of the upper limit of integration becomes irrelevant after several lifetimes τ_X. What this formula clearly shows, is that there is a change in entropy of the universe. The thermodynamic evolution is no longer dictated by the simple $s \propto a^{-3}$. Since $s \propto T^3$, the change in entropy means the universe temperature is increased by the decay.

The change in ΔS must be calculated numerically. To make an estimate we can use the approximation of "instant decay". In this approximation we suppose there is no decay until a time t_d and then, instantly, all the X particles decay into standard model ones, which thermalize immediately. By conservation of energy the energy density of the Xs must entirely be accounted for in an immediate heating of the photons to a temperature T_{RH} (since it is instantaneous, we can forget that the universe is expanding and that the density dilutes)

$$\rho_{Xi} + \frac{\pi^2}{30}\tilde{g}T_D^4 = \frac{\pi^2}{30}\tilde{g}T_{RH}^4 \tag{10.25}$$

As the process is immediate, there is no change of degrees of freedom $\tilde{g}$. The value of the temperature $T_{RH} > T_D$ can be computed. From this we can get the final entropy density s_f and its ratio to s_i.

$$\frac{s_f}{s_i} = (\frac{30\rho_{Xi}}{\pi^2\tilde{g}T_D^4} + 1)^{\frac{3}{4}} \tag{10.26}$$

The initial density of the X can be expressed through its initial abundance $\rho_{Xi} = m_X s_i Y_{Xi} = \frac{2\pi^2}{45}m_X g^* T_D^3 Y_{Xi}$ so that

$$\frac{s_f}{s_i} = (1 + \frac{4}{3}\frac{g^*}{\tilde{g}}\frac{m_X}{T_D}Y_{Xi})^{\frac{3}{4}} \tag{10.27}$$

We note that the entropy production may not be an irrelevant feature. In fact we had assumed from the outset that $m_X \gg T_D$, that the particle was non-relativistic. If the abundance Y_{Xi} is not too small the ratio may be much larger than 1. This may be the case for reheating. Suppose there is a reheating after inflation when the fields or particles which drove inflation decay into standard model particles. Assuming standard model particles were not able to maintain equilibrium as the universe expanded exponentially, their temperature could be exponentially suppressed and we may indeed be in a situation where m_X is much larger than T_D. Thus, by decay the universe heats up to a temperature where the standard cosmology can begin.

11 Recombination, decoupling and the cosmic microwave background

The most important observable to understand the early universe is the cosmic microwave background (CMB). This is the relic photon density which has since decoupled from the rest of the matter of the universe, whose initial over-densities went on to form the large scale structures we see today. We will now study the era of the universe known as *decoupling*. This is the moment when the the photon Compton scattering rate on electrons and protons becomes longer than a Hubble time and its free streaming length becomes comparable to the size of the observable universe. While the primordial universe is opaque to photons, the universe after decoupling is very much transparent, so much so that we observe the photons from that era today, relatively unperturbed. Assuming the universe is isotropic, and neglecting the in-homogeneities of the fluid, we can say that we observe CMB photons today arriving from a spherical surface centered on us, known as the *last scattering surface (LSS)*.

Decoupling is connected to *recombination*, the process that brings free protons and electrons to form neutral hydrogen atoms. Neutral hydrogen atoms may only form in non-negligible quantities when the temperature becomes comparable with the binding energy $E_I = 13.6eV$. At higher temperatures any bound state will be quickly ionized by a thermal photon. Indeed, since there are many more photons than protons and electrons, decoupling happens at a temperature about an order of magnitude less at $T = 0.25eV$, corresponding to a, well measured, redshift z at decoupling $z_d = 1090$.

Around temperatures $T \sim 10eV$ the reaction

$$e^- + p \leftrightarrow H + \gamma \tag{11.1}$$

becomes relevant. There are also reactions which produce neutral helium atoms from Helium nuclei which we neglect, but are needed for a more precise treatment. We define the fraction of free electrons n_e to total baryons as

$$X_e \equiv \frac{n_e}{n_p + n_H} = \frac{n_e}{n_b} \tag{11.2}$$

We assume electrical neutrality of the universe, which is confirmed by astrophysical observations, so we can always take $n_p = n_e$. The total baryon density is $n_b = n_p + n_H$ (neglecting Helium density). As we noted several times, decoupling of a species happens when the scattering rate is roughly equal to the Hubble factor. We willestimate when this happens by looking at the scattering rate of a photon on electrons

$$n_e \sigma_T = X_e n_b \sigma_T \tag{11.3}$$

where $\sigma_T = 0.665 \cdot 10^{-24} cm^2 = 1.70 \cdot 10^{-15} eV^{-2}$ is the Thompson cross section, the non-relativistic limit of Compton scattering $e^- + \gamma \to e^- + \gamma$. There is also scattering of photon on protons. In the non-relativistic limit the cross section for this process is suppressed by $\frac{m_e}{m_p}$ with respect to Compton scattering and so we neglect it.

Assuming baryon number conservation, the baryon number can be related to the baryon density today with $n_b = \frac{3}{8\pi G}\frac{\Omega_{b0}H_0^2}{m_p a^3}$, where m_p is the mass of the proton and the density parameter Ω_{b0} is reported in table 7.1[7]. Using the Friedmann equation, (7.25), recalling that dark energy is very subdominant in the primordial universe, $\frac{H}{H_0} = \frac{\Omega_{m0}}{a^{3/2}}\sqrt{1+\frac{a_{eq}}{a}}$, given the matter-radiation equivalence scale factor $a_{eq} = \frac{\Omega_{r0}}{\Omega_{m0}}$. Putting everything together, and exchanging the scale factors for redshift factors we get the ratio of scattering rate to Hubble factor

$$\frac{n_e\sigma_T}{H} = \frac{8\pi G}{3}\frac{m_p}{\sigma_T}X_e\frac{\Omega_{b0}H_0^2}{\sqrt{\Omega_{m0}H_0^2}}\frac{(1+\frac{1+z_d}{1+z_{eq}})^{\frac{1}{2}}}{(1+z_d)^{\frac{3}{2}}} \tag{11.4}$$

Plugging in all the numbers, we can find that this ratio is equal to 1 when the fraction of free electrons $X_e \simeq 0.01$, or when around 99% of proton-electron pairs have formed a neutral hydrogen atom. If we hadn't measured the details of the cosmic microwave background, we would like to find a functional dependence $X_e(z)$ so that the above equation could be used to determine and predict z_d. Of course, such a dependence can be found by examining the Boltzmann equation.

Before looking at the Boltzmann equation in full, we make an estimate of X_e on the temperature (therefore on z, through $T(z)$) assuming the number densities of electrons and hydrogen atoms are given by their equilibrium density, so that $n_e = n_e^{eq}$, $n_H = n_H^{eq}$. By using the non-relativistic limits for number densities, with zero chemical potential, shown in table 8.1, we obtain the *Saha equation*

$$\frac{X_e^2}{1-X_e} = \frac{1}{n_b}\left(\frac{m_e m_p T}{2\pi m_H}\right)^{\frac{3}{2}} e^{-\frac{E_I}{T}} \tag{11.5}$$

where in the prefactor we can simplify $m_p \simeq m_H$ and $m_e + m_p - m_H = E_I$ is the binding energy of the hydrogen. The Saha equation turns out to be a good estimate of the temperatures required to have $X_e \simeq 10^{-2}$. It works well so long as $T \gtrsim 0.2eV$, when the reaction rates are sufficient to maintain the equilibrium density. Below that, the densities decrease much more slowly than exponentially, as an equilibrium distribution would presume.

We wish to find a Boltzmann equation to describe the number density n_e of free electrons. For all intents and purposes, hydrogen can be thought as a new species of particles, with the reaction $e^- + p \leftrightarrow H + \gamma$ "annihilating" an electron and a proton to create a hydrogen atom and a photon. Then, the first steps are the same as those we had taken when discussing a cold relic in section 10.1, with the difference that the distribution functions now refer to different particles. We assume that each species thermalizes on its own faster than the rate of hydrogen forming reactions, so that we assume a common temperature but chemical potentials μ_i which are not the equilibrium ones. These contain all the unknown we have to solve for. Then we integrate the Boltzmann equation (9.2) over the momenta $\int \frac{d^3p}{(2\pi)^3}$ of the electron. The Liouville term of the Boltzmann equation is

$$a^{-3}\frac{d}{dt}(n_e a^3) \tag{11.6}$$

[7] Ω_b also count the density of charged leptons, since cosmologists refer to "baryons" as actual baryons and electrons. However since $m_e \ll m_p$ and assuming charge neutrality, the contribution to the density is almost entirely due to the baryons.

The integrated collision term is

$$\int \frac{d^3p_e d^3p_p d^3k_H d^3k_\gamma}{(2\pi)^{12} 16 E_e E_p E_H E_\gamma}(2\pi)^4\delta^4(p_e^\mu + p_p^\mu - k_H^\mu - k_\gamma^\mu)|\mathcal{M}_{e^-p\to H\gamma}|^2 \times \\ [f_H(\vec{k}_H)f_\gamma(\vec{k}_\gamma) - f_{e^-}(\vec{p}_e)f_p(\vec{p}_p)]$$

Typically, the energies $E-\mu$ are larger than T and so we have neglected the Bose-Einstein or Fermi-Dirac suppression and enhancement factors. Furthermore, this allows us to approximate every distribution with $f_i(p) \simeq g_i e^{\mu_i/T}e^{-E/T}$, since the exponentials become large compared to the ± 1 in the denominators. With these simplifications, we have that

$$f_H(\vec{k}_H)f_\gamma(\vec{k}_\gamma) = e^{(\mu_H+\mu_T)/T}e^{-(E_H+E_\gamma)/T} = e^{(\mu_H+\mu_\gamma)}f_{e^-}^{eq}(\vec{p}_e)f_p^{eq}(\vec{p}_p) \tag{11.7}$$

where the equilibrium distributions are those at zero chemical potential, ie. $f_{e^-}^{eq} = f_{e^-}(\mu_e = 0) = g_i e^{-E/T}$. Now, since the number density of a species is

$$n_i = g_i e^{\frac{\mu_i}{T}}\int \frac{d^3p}{(2\pi)^3}e^{-E/T} = g_i e^{\frac{\mu_i}{T}} n_i^{eq} \tag{11.8}$$

we can introduce the thermally averaged cross section times velocity as

$$n_p^{eq}n_e^{eq} < \sigma v > \quad = \\ \int \frac{d^3p_e d^3p_p d^3k_H d^3k_\gamma}{(2\pi)^{12} 16 E_e E_p E_H E_\gamma}(2\pi)^4\delta^4(p_e^\mu + p_p^\mu - k_H^\mu - k_\gamma^\mu) \times \\ |\mathcal{M}_{e^-p\to H\gamma}|^2 f_{e^-}^{eq}(\vec{p}_e)f_p^{eq}(\vec{p}_p)$$

and the Boltzmann equation can be reduced to the simple form

$$a^{-3}\frac{d(n_e a^3)}{dt} = n_e^{eq}n_p^{eq} < \sigma v > \left(\frac{n_H n_\gamma}{n_H^{(eq)}n_\gamma^{(eq)}} - \frac{n_e n_p}{n_e^{eq}n_p^{eq}}\right) \tag{11.9}$$

We can assume that the actual photon density is equal to its equilibrium density. This can be put down to the fact that photons have a much higher thermalization rate and their number is not conserved, but also to experiment. In fact the spectrum of the cosmic microwave background is well measured to have no chemical potential. The ratio of equilibrium densities is just the Saha equation, in a different form

$$\frac{n_e^{eq}n_p^{eq}}{n_H^{eq}} = \left(\frac{m_e m_p T}{2\pi m_H}\right)^{\frac{3}{2}} e^{-\frac{E_I}{T}} \tag{11.10}$$

Noting that $n_p = n_e$ (but $n_p^{eq} \neq n_e^{eq}$!), due to charge neutrality, and $n_e = X_e n_b$, $n_H = (1-X_e)n_b$

$$a^{-3}\frac{d(n_e a^3)}{dt} = n_b < \sigma v > \left((1-X_e)(\frac{m_e T}{2\pi})^{\frac{3}{2}}e^{-\frac{E_I}{T}} - X_e^2 n_b\right) \tag{11.11}$$

We divide both sides by $n_b a^3$. Since the total baryon number per comoving volume is constant we can pull this factor into the time derivative on the left hand side to obtain the

Boltzmann equation in the standard form

$$\frac{dX_e}{dt} = [\beta(1 - X_e) - X_e^2 n_b \alpha^{(2)}] \tag{11.12}$$

Where the quantities

$$\beta \equiv < \sigma v > \left(\frac{m_e T}{2\pi}\right)^{\frac{3}{2}} e^{-\frac{E_I}{T}} \tag{11.13}$$

$$\alpha^{(2)} \equiv < \sigma v > \tag{11.14}$$

Unfortunately, although correct in form, the above factors are not the correct ones when calculating recombination numerically[167]. We have neglected the existence and importance of the excited states of hydrogen. This turns out to be important in determining recombination accurately at $\sim 10\%$. We would need to consider the different populations of excited states and their interactions with photons in a Boltzmann equation to get a better estimate. There are several interesting phenomena at play.

Hydrogen produced directly in its ground state contributes extremely little to recombination. This is the reason the term $\alpha^{(2)}$ has its suffix and, in a numerical integration, we should use the cross section for electron capture by a proton in the $n = 2$ excited state. The reason for this is that any hydrogen produced directly in the ground state will also produce a photon with energy E_I. This photon has a resonant cross section with neutral atoms in the ground state and therefore will immediately re-ionize a second neutral atom. The net production of neutral hydrogen this way turns out to be very close to zero.

If an H is produced in a more excited state however, it will then decay to the ground state, either directly or through a successive series of decays. Every decay has a photon which has the energy of the energy difference of the initial and final bound state, therefore it could, and does, also resonantly excite other hydrogen atoms which then have a good chance of being ionized through thermal photons. We understand that the most important state is the $n = 2$, and this idea is at the basis of the *effective three-level atom* calculation. Producing a hydrogen in the $n = 2$ state which then decays to $n = 1$ is the dominant recombinaton mechanism. The fine structure is important. In fact the $n = 2$ state could be $2s$ or $2p$.

Decays from $2s$ to $1s$ in hydrogen are zero at leading order due to conservation of energy and angular momentum. The decay can happen with emission of two photons and takes a longer time, which means in the meantime that no other hydrogen are ionized by the emission. In the mean time the universe is getting less dense. This is a competitive way to produce neutral hydrogen in the ground state. The alternative is the decay from $2p$ to $1s$ which will almost always ionize or excite another hydrogen atom, except for the relatively rare cases where the energy of the photon gets redshifted sufficiently in order to not have a resonant energy. This is not the most probable thing for every given photon, but it is systematic and it turns out it produces a H_{1s} atom with about the same rate as those produced from the $2s \to 1s$ decay.

To account for this, the simplest way, introduced by Peebles[167], is to multiply all of the right hand side of equation (11.12) by a factor C_r, known as the reduction coefficient. The reduction coefficient is the probability that an $n = 2$ atom decays to the lowest state by one of the two ways above before being photo-ionized.

12 Nucleosynthesis

Big Bang Nucleosynthesis (BBN) is the main mechanism to form nuclei from protons and neutrons in the universe[71]. The only other known mechanism is stellar nucleosynthesis, which happened much later in the history of the universe[43]. Big Bang nucleosynthesis happens in the first minutes after the Big Bang when the temperature is falling down to $T \sim 1MeV$. Large are needed temperatures to overcome the Coulomb barrier.[8]

At these temperatures, the dominant particles are photons and neutrinos, which decouple at a temperature of a few MeV. The actual decoupling temperature is important, as we shall see. The universe also contains protons, neutrons and heavier baryons. Since the masses of the baryons are at least $m_p = 0.938GeV$, we would expect all baryons to be exponentially depleted. However, as we had remarked in section 8.2, there exists a matter-antimatter asymmetry in the universe and a non-zero chemical potential for protons and neutrons which allows a non-negligible density. This small but non-zero density is responsible for all the matter we have today. Indeed, measurement of nuclear abundances in the universe give the best measurement for the chemical potential of matter. In a sense, baryons can be though of as a relic; there can be no more number changing reactions and the abundance has frozen to some non-equilibrium value given by the freeze-out. The mystery is why this non-equilibrium value has preferred protons over anti-protons.

It is common to parametrize the matter-antimatter asymmetry through the ratio of baryons to photons

$$\eta \equiv \frac{n_b}{n_\gamma} \tag{12.1}$$

This value is very small $\eta \sim 10^{-10} \div 10^{-9}$. We shall see its main determination comes from BBN. Since baryon abundance is conserved, so long as evolution of the universe is isentropic, $\eta = \frac{sY_b}{n_\gamma}$.

Beginning from very high temperatures, before neutrinos have decoupled, we want to understand what abundances of the various heavy elements are produced. As the temperature decreases, thermal equilibrium will prefer lighter baryons, the proton being the lightest. However nuclear reactions scale roughly as the density squared and will not maintain equilibrium for long. The problem is not very different from how we treated recombination. In fact, formulas will be similar.

At very high temperatures, $T \sim 10MeV$, weak interactions maintain the thermodynamical equilibrium between protons, neutrons, neutrinos, electrons, anti-electrons and photons. The relevant weak reactions are $\nu_e + n \leftrightarrow p + e$, $e^+ + n \leftrightarrow p + \bar{\nu}_e$ and $n \leftrightarrow p + e + \bar{\nu}_e$. These are important since they are the only nuclear isospin changing interactions and keep the ratio of protons and neutrons to its equilibrium value

$$\frac{n_n^{eq}}{n_p^{eq}} = (\frac{m_n}{m_p})^{\frac{3}{2}} e^{-\frac{Q}{T}} \tag{12.2}$$

where $Q = m_n - m_p = 1.293MeV$. In deriving this we have neglected the chemical potential for the electrons $\mu_e \ll Q$ and that of the neutrinos μ_ν. There is very little experi-

[8] Actually, quantum tunneling under the barrier is quite important due to the Gamow factor[108]. The probability of two nucleons to tunnel across their Coulomb barrier is $P_g(E) \simeq \exp -\sqrt{\frac{E_g}{E}}$, where $E_g = \pi^2\alpha^2 m_p$.

mental evidence on the value of the chemical potential of the neutrinos. We would assume $\mu_\nu \simeq \mu_e$, but this is an assumption which should be challenged. The prefactor above can be simplified, but the mass difference is important in the exponential. At very high temperatures the density of protons and neutrons is the same. The neutrinos freeze out at a temperature $T \sim 0.8MeV$, when $e^{-Q/T} \sim 1/5$. After this, we will have to consider nuclear reactions which produce heavier elements. Before we do, we note that if there were no nuclear reactions the neutrons would simply decay over time. Luckily, the neutrons lifetime is $\sim 10min$, longer than the time it will take to form heavier nuclei, but the depletion in neutrons over time will affect the final abundances.

Let's consider equilibrium densities for a general nucleus X_Z^A, A being the mass number and Z the proton number. The nucleus binding energy is given by

$$B_A = Zm_p + (A-Z)m_n - m_A \tag{12.3}$$

As long as the reaction rate is greater than the expansion of the universe, and these nuclei are in equilibrium, the chemical potentials must satisfy (8.15), with a reaction of the form $Zp + (A-Z)n \leftrightarrow X_Z^A$ (this is true even if the reaction proceeds in stages)

$$\mu_A = Z\mu_p + (A-Z)\mu_n \tag{12.4}$$

The density of this species is then the usual non-relativistic one

$$n_A = g_A \left(\frac{m_A T}{2\pi}\right)^{\frac{3}{2}} e^{-\frac{m_A}{T}} e^{Z\frac{\mu_p}{T}} e^{(A-Z)\frac{\mu_n}{T}} \tag{12.5}$$

which for $A = 1, Z = 1$ is the density of the proton and for $A = 1, Z = 0$ is the density of the neutron(with $g_{p,n} = 2$). Thus we can re-express the above in terms of the proton and neutron density (using $m_n = m_p$ in the prefactor)

$$n_A = \frac{g_A}{2^A} \left(\frac{(2\pi)^{A-1} m_A}{m_p^A T^{A-1}}\right)^{\frac{3}{2}} n_p^Z n_n^{A-Z} e^{\frac{B_A}{T}} \tag{12.6}$$

We define the *mass fraction*

$$X_A \equiv \frac{A n_A}{n_b} \tag{12.7}$$

where n_b is the total number of baryons

$$n_b = n_n + n_p + \sum A n_A \tag{12.8}$$

With some straightforward algebra, using η and the full expression for the number density of photons n_γ as in table 8.1, the *equilibrium mass fraction is*

$$\begin{aligned} X_A &= g_A A^{\frac{5}{2}} [\zeta(3) \frac{2^{(3A-5)/(2A-2)}}{\sqrt{\pi}}]^{A-1} \times \\ &\left(\frac{T}{m_p}\right)^{\frac{3}{2}(A-1)} \eta^{A-1} X_p^Z X_n^{A-Z} e^{\frac{B_A}{T}} \end{aligned} \tag{12.9}$$

A lengthy expression, but not mysterious. We immediately note that for $A > 1$ the mass fraction is suppressed by powers of η, which we have already remarked is very small. The exponential is the dominating factor and $B_A \sim A \cdot MeV$, so for large temperatures $e^{B_A/T} \sim 1$ and the mass fractions are suppressed. This is the first important result: *at the start of BBN the only baryons are protons and neutrons*. Physically, what is happening is that the heavier nuclei, such as deuterium (D) can be produced by nuclear reactions but are quickly photo-dissociated and large abundances do not form. Therefore, nucleosynthesis can only begin in earnest when $e^{B_A/T} \sim \eta^{-1}$. By plugging in all the numbers one finds that the mass fraction X_D of deuterium becomes of order unity when $T \sim 0.07MeV$.

We note that deuterium must be produced before any heavier elements, such as helium is produced. The probability to produce ${}^3H$ or ${}^3He$ directly from three nucleons is very small, so these heavier elements can only be produced after Deuterium. Deuterium fraction becomes relevant only when the temperature is becoming smaller than that needed for nuclear reactions. This is known as the *Deuterium bottleneck*.

The fact that nucleosynthesis begins at $T \sim T_{nuc} = 0.07MeV$ means that we don't have to deal with the electrons becoming non-relativistic and heating up in the meantime, this has already happened at $T \sim 0.5MeV$. So when nucleosynthesis begins, there are no weak interactions and temperature will simply decrease as $T \propto a^{-1}$. However, some time has passed since the proton and neutron ratio fell out of equilibrium: a fraction neutrons have decayed. We can simply account for this by finding the time that elapsed between the freeze-out of weak interactions and damping the neutron density by e^{-t/τ_n}. A proper freeze-out calculation, with equations similar to (11.9) for the weak reactions, to find the freeze-out abundance and then applying the decay factor gives

$$\frac{n_n}{n_p}(T_{nuc}) \simeq 0.133 \tag{12.10}$$

which corresponds to a mass fraction

$$X_n(T_{nuc}) \simeq 0.11 \tag{12.11}$$

At this temperature deuterium can be formed through the nuclear reaction $n + p \leftrightarrow D + \gamma$ and it can be shown that this reaction is efficient until well after the end of nucleosynthesis. Very quickly, a large amount of deuterium is produced in accordance with its equilibrium mass fraction (12.9). However, as soon deuterium is produced, the production of Helium through $D + D \rightarrow n +^3 He$, $D +^3 He \rightarrow p +^4 He$ begins efficiently as well. Although the Helium fraction has an extra factor of η to deal with, it is exponential in the binding energy which is $B_4 \simeq 28MeV$, much larger than that of deuterium. Thus, the exponential factor for the Helium-4 dominates and it quickly becomes the most abundant nucleus, after the bare proton. *The production of Helium-4 continues undisturbed until the universe runs out of neutrons*. Since this happens quickly, we can ignore neutron decays. The final abundance of ${}^4He$ is half the abundance of neutrons at T_{nuc}, since 2 neutrons are required per ${}^4He$. Thus $n_{^4He} = 2n_n(T_{nuc}) \times (\frac{a_{nuc}}{a})^3$. In the mass fraction, using conservation of baryon number,

the scale factor dependence drops out

$$X_{^4He}(T \to \infty) = \frac{4 \cdot \frac{1}{2} n_n(T_{nuc})}{n_n(T_{nuc}) + n_p(T_{nuc})} \simeq 0.23 \tag{12.12}$$

while the mass fraction of protons after all neutrons are exhausted is

$$X_p(T \to \infty) \simeq 0.77 \tag{12.13}$$

The main prediction of nucleosynthesis is that Hydrogen and Helium-4 are the dominant elements in the universe. In fact, once the Helium is produced, the penetration of the Coulomb barrier becomes smaller and nuclear rates proceed more slowly. Some Lithium is produced, either via the reaction $^4He +^3 H \to^7 Li + \gamma$ or $^4He +^3 He \to^7 Be + \gamma$ followed by a β-decay of the beryllium to lithium via an electron capture. These rates are not very efficient, since they involve the densities of ^3H and ^3He which are small. In addition, there exist no stable nuclei with mass $A = 5, 8$ which is another bottleneck for nuclear fusion. There is no reaction which can combine a Helium-4 with a proton to produce something heavier and stable. In stars, heavier elements, such as carbon $^{12}_6C$ which has a much larger binding energy and whose equilibrium fraction would be larger (just as Helium-4's was compared to Deuterium), are only produced through the *triple-alpha process*, which involves three ^4He nuclei and is indeed rare. In fact, even in the large densities of stars it takes thousands of years to produce a reasonable amount of ^{12}C, compared to *minutes* of BBN.

We have seen that Helium-4 is most abundant nucleus synthesized and its abundance depends mainly on the neutron lifetime, whose measurement has a large uncertainty. It depends weakly(~ logarithmically) on the matter asymmetry η since a slightly larger value of η will permit earlier production of Deuterium and hence BBN begins with a slightly larger amount of neutrons. With numerical integrations of the Boltzmann equations[35, 67, 173], one can show that the final mass fraction of ^2H, ^3H and ^3He is $\sim 10^{-5}$, while the mass fraction of lithium is predicted to be $\sim 10^{-10}$.

12.1 Primordial light elements observations and conclusions

Observation of the abundances of primordial element are greatly complicated by the fact that stellar activity has processed nuclei since BBN. Stars, for example, also produce Helium-4 throughout the lifetime of the universe and separating what has been produced in stellar fusion from the primordial abundance is the main experimental complication. For all the light element abundances, systematic errors are the dominant errors. To measure primordial abundances one must look for astrophysical sites with low metallicity. We expect heavier elements, and metals in particular, to not be produced at all in BBN (except perhaps $^{12}_6C$ in extremely small and yet unmeasured quantities), but be copiously produced in stellar fusion. Sites with low metallicity can be supposed to have a fractional composition similar to that of the early universe. Composition is, of course, determined by measuring the emission spectrum.

Deuterium, because of its relatively light binding energy, is entirely destroyed by stellar reactions[89]. Observed deuterium can be supposed to be from the BBN and any measurement then presents a lower limit on the abundance after BBN. Measurement of deuterium

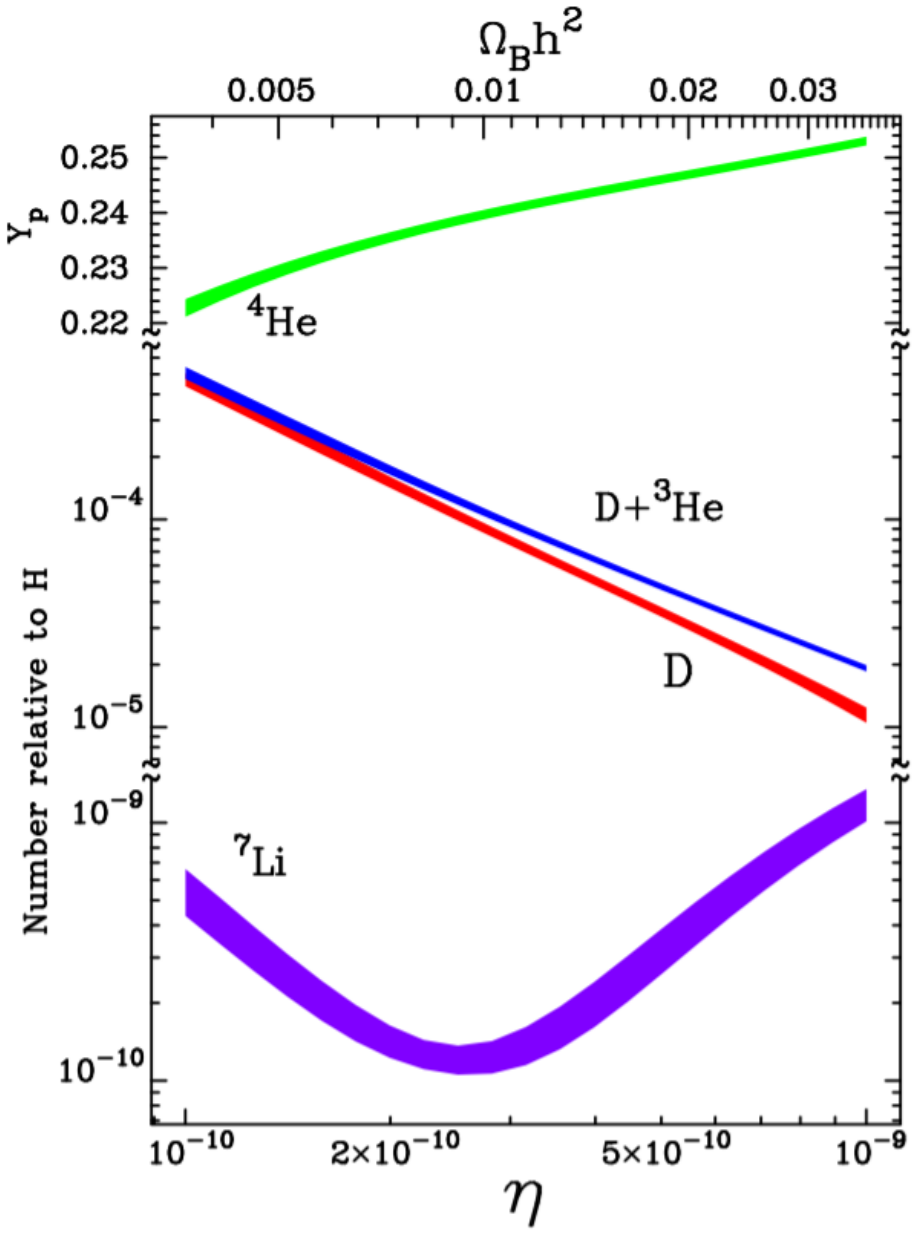

Figure 12.1: Predictions of Big Bang Nucleosynthesis primordial abundances shown with 95% C.L. bands. Figure from ref. [35].

abundance can be done by observing distant matter which has seen little or no stellar activity[172]. Alternatively, one can measure the ratio of deuterium to Helium in the solar system. This gives a lower limit to the local deuterium abundance before solar activity. Its easy to see (12.9)[9] that $X_D^2/X_{^4He} \propto \eta^{-1}$, which implies that a larger value of η will result in a smaller deuterium fraction. Measurements of deuterium put an upper limit to the value of η.

The primordial value of Helium-4 can be measured by observing emission lines in HII regions (galactic and extra-galactic)[132, 133, 162]. Helium-4 is also produced by stellar activity. To discern what fraction is primordial, one can compare the measured Helium-4 fraction with the metallicity of the region and extrapolate to zero metallicity. Since the observed system are complex, the main source of error is definitely a systematical one. Two measurements of the primordial Helium-4 fraction are[99]

$$X_{^4He} = 0.2449 \pm 0.0040 \tag{12.14}$$

$$X_{^4He} = 0.2551 \pm 0.0022 \tag{12.15}$$

Since they are discordant, this indicates there is some unaccounted for systematic effect.

By fitting the results of a numerical integration as a function of η , the measurements of Helium-4 and deuterium are in agreement with the theory and each other, providing a value of η

$$5.8 \cdot 10^{-10} < \eta < 6.6 \cdot 10^{-10} \tag{12.16}$$

at 95% confidence level. Thus BBN provides an important constraint on the matter-antimatter asymmetry. η can be related to the baryon content of the universe (through T_{CMB}) giving a value

$$0.021 < \Omega_b h^2 < 0.024 \tag{12.17}$$

See appendix A for the definition of h.

Finally we should mention Lithium primordial abundance. Measurements of Lithium compared to metallicity can be done in a similar fashion as with Helium-4. However the stellar and nuclear physics is more complicated, possibly leading to unknown systematics. In fact, the measurements of Lithium are very much in tension with the well measured values of deuterium and Helium-4. It would require a different value of η. This is known as the Lithium problem[101] and its resolution could come from better understanding the systematics involved in measurement. The solution may also arise from new physics which alters the primordial production.

13 Evolution of primordial in-homogeneities and anisotropies

In studying Big Bang Nucleosynthesis and recombination, we assumed that the universe was homogeneous. Although the density of the species deviates from the equilibrium val-

[9] Although the neutron fraction is not that of equilibrium, the reactions keeping D in equilibrium are efficient and the ratio between nuclei is that of equilibrium.

ues, we always assumed this was the same throughout the universe. Actually, the density of the universe was not perfectly homogeneous in early times. This is evidenced by the fact that the cosmic microwave background has temperature fluctuations across the sky of order $\delta T/T \sim 10^{-5}$. These are known as *Cosmic Microwave Background anisotropies*. The smallness of the anisotropies implies that we can use perturbation theory, and assume the density and local speed fluctuations were small compared to the homogeneous, zero-order, density, at least for early times. Eventually, the over-dense regions of the universe condense to form the largest structures we see today, while the under-dense regions become voids. Understanding the evolution of these fluctuations is the cornerstone to modern cosmology and we shall study them in detail.

13.1 Perturbed space-time

We assume that space-time is at lowest order given by the flat ($k = 0$) FRW metric, and the time evolution follows the Friedmann equation. Then we study small perturbations around this zero-order metric. Working with conformal time, the metric can be perturbed generically[171]

$$ds^2 = a^2(\tau)\left[-(1+2A)d\tau^2 + 2B_i dx^i d\tau + [(1+2D)\delta_{ij} + \chi_{ij}]dx^i dx^j\right] \tag{13.1}$$

Where χ_{ij} is taken to be traceless, as D is the spatial-trace perturbation. We shall generically write the metric tensor as

$$g_{\mu\nu} = \bar{g}_{\mu\nu} + \delta g_{\mu\nu} \tag{13.2}$$

where $\bar{g}_{\mu\nu}$ is the unperturbed metric and $\delta g_{\mu\nu}$ is the perturbation. The perturbations, which are encoded in the functions A, B_i, D and χ_{ij}, are functions of space-time. We are now abandoning the cosmological principle, and so we will not assume isotropy and homogeneity except in a statistical sense, later on. In the meanwhile, nothing special should be assumed about these functions, except that they are small. The quantities carry only spatial indexes an we'd interpret them as three-vectors, or spatial tensors. In order to construct quantities that are invariant under three-dimensional rotations we define the raising and lowering through the spatial metric $a^2[(1+2D)\delta_{ij} + \chi_{ij}]$. Since these quantities are small, at first order raising and lowering is trivial: $B_i = B^i$ and similarly for other quantities.

The three-vector B_i can be decomposed into the sum of a gradient and a divergenceless term

$$B_i = B_{,i} + \bar{B}_i \tag{13.3}$$

where B is a function of space and time, and $\bar{B}_i$ a vector function which has zero divergence

$$\partial_i \bar{B}^i = 0 \tag{13.4}$$

As we said, $\bar{B}^i = \frac{1}{a^2}\bar{B}_i$, and the presence of the τ-dependent scale factor does not mess with the spatial derivatives. In the same manner, the traceless and symmetric tensor χ_{ij} can be decomposed into

$$\chi_{ij} = (\partial_i\partial_j - \frac{1}{3}\delta_{ij})C + \frac{1}{2}(\bar{C}_{i,j} + \bar{C}_{j,i}) + \tilde{C}_{ij} \tag{13.5}$$

with $\bar{C}^i_{,i} = 0$ and $\tilde{C}^{ij}_{,j} = 0$. With these decompositions, we have separated the metric perturbations into four scalar functions, A, B, C, D, two divergenceless vector functions $\bar{B}_i$,$\bar{C}_i$, and one symmetric, traceless, and divergence free tensor $\tilde{C}_{ij}$. This is known as the *scalar-vector-tensor decomposition* (SVT). Although we will not prove it in the general case, it turns out that the perturbations to the metric can always be separated into terms that transform as a scalar, vector or tensor under spatial rotations. The same can be done to the stress-energy tensor so that, at lowest order in perturbation theory, the three types of perturbations are decoupled and can be studied separately.

Each function $\bar{B}_i$, $\bar{C}_i$, expresses two degrees of freedom. The tensor function $\tilde{C}_{ij}$ expresses $6 - 1 - 3 = 2$ degrees of freedom, since $\tilde{C}^{ij}_{,j} = 0$ expresses three independent conditions and the tracelessness another. The perturbations seem to have ten degrees of freedom. This turns out to be too much.

We have forgotten about gauge invariance of the theory. This is more than a technical quibble. A choice of the functions above has implicitly chosen a coordinate system to work in. For example, even in a perfect FRW metric, we could choose conformal time coordinate which is no longer the proper time of a comoving observer, smudging the time coordinate in different manners and at different points. Then the metric may look somewhat like the perturbed one above. In order to understand the physical degrees of freedom, we need to see how the generic perturbed metric (13.1) transforms under coordinate changes.

To this end, lets take a coordinate transformation between x^μ and y^μ

$$y^\mu = x^\mu + \xi^\mu(x^\rho) \tag{13.6}$$

Where the functions ξ^α are of the same order as the perturbations. The ξ^μ are functions and not vectors on the tangent space, so one can simply define $\xi_\mu = \xi^\mu$ in the following. More specifically we can write the spatial vector part as the sum of a gradient and a divergenceless vector

$$\xi_i = S_{,i} + \bar{S}_i \tag{13.7}$$

Let's see how the metric changes when changing coordinates. The metric in the y^μ coordinates is $g^{'}_{\mu\nu}$ and has the same form as (13.1) with the new functions denoted with a prime. We will calculate these function at the *same space-time coordinate* which, in general is not the same physical coordinate. This allows us to interpret the x^μ coordinate as the background coordinate on the unperturbed metric. Working at first order we have

$$g_{\mu\nu}(x^\rho) = \frac{\partial y^\alpha}{\partial x^\mu}\frac{\partial y^\beta}{\partial x^\nu} g^{'}_{\alpha\beta}(y^\gamma) = \frac{\partial y^\alpha}{\partial x^\mu}\frac{\partial y^\beta}{\partial x^\nu}(g^{'}_{\alpha\beta}(x^\rho) + g_{\alpha\beta,\gamma}\xi^\gamma) \tag{13.8}$$

Keeping at most first order terms

$$g_{\mu\nu}(x^\rho) = g^{'}_{\mu\nu}(x^\rho) + g^{'}_{\mu\nu,\alpha}\xi^\alpha + \xi^\alpha_{,\mu} g^{'}_{\alpha\nu} + \xi^\alpha_{,\nu} g^{'}_{\alpha\mu} \tag{13.9}$$

The time-time component is the simplest. For the metric terms multiplying ξ^α, we can use the FRW metric which is diagonal and only time dependent. Explicitly we get

$$-a^2(1+2A) = -a^2[1 + 2A^{'} + 2\mathcal{H}\xi^0 + 2\dot{\xi}^0] \tag{13.10}$$

where we remind that a dot indicates derivatives with respect to conformal time and $\mathcal{H} = \frac{\dot{a}}{a}$ is the conformal Hubble factor. Thus

$$A = A^{'} + \mathcal{H}\xi^0 + \dot{\xi}^0 \tag{13.11}$$

The time space component gives (recall that $\xi^i = \xi_i$)

$$B_i = B_i^{'} + \xi_{i,0} - \xi_{0,i} \tag{13.12}$$

Separating out the gradient and divergence free vector we obtain

$$B = B^{'} + \dot{S} - \xi^0 \tag{13.13}$$

$$\bar{B}_i = \bar{B}_i^{'} + \bar{S}_{i,0} \tag{13.14}$$

The space-space component gives

$$2D\delta_{ij} + \chi_{ij} = 2D^{'} + \chi_{ij}^{'} + 2\mathcal{H}\delta_{ij}\xi^0 + (\xi_{j,i} + \xi_{i,j}) \tag{13.15}$$

Taking the trace on $i - j$, the term $\xi^i_{,i} = \nabla^2 S$, and we obtain

$$D = D^{'} + \mathcal{H}\xi^0 + \frac{1}{3}\nabla^2 S \tag{13.16}$$

Plugging back in its obvious by comparison that

$$C = C^{'} + 2S \tag{13.17}$$

$$\bar{C}_i = \bar{C}_i^{'} + \bar{S}_i \tag{13.18}$$

$$\tilde{C}_{ij} = \tilde{C}_{ij}^{'} \tag{13.19}$$

Interestingly enough, the tensor perturbation $\tilde{C}_{ij}$ does not change under small transformations. In fact, this term can be shown to express *gravitational waves* which are a physical degree of freedom. We could have guessed there would be no change before going through the calculation. In fact the small coordinate change involves only two scalar functions, ξ^0 and S, and one vector function $\bar{S}^i$. Due to the SVT decomposition, the scalars will enter into the transformations of the scalar quantities of the metric, and the vector in the vector quantities. Nothing in the coordinate change transforms as a tensor and can affect $\tilde{C}_{ij}$. These decouplings will keep cropping up and are always a manifestation of the SVT theorem.

Although when we proceed to solve the Boltzmann equations we will choose a gauge to work in, we need to know which quantities are gauge invariant, and thus can be taken have some physical meaning. There were four scalar degrees of freedom, but the coordinate transformation provides us with two scalar functions, so there are only two physical scalar modes. By the same reasoning there is only one divergence free vector mode, in addition

to two degrees of freedom for the tensor mode. The total is six physical degrees of freedom (as expected: $10 - 4 = 6$). The gauge invariant quantities are[13]

$$\lneqq_A = A + \mathcal{H}(B - \frac{\dot{C}}{2}) + (\dot{B} - \frac{\ddot{C}}{2}) \tag{13.20}$$

$$\Phi_H = D - \frac{1}{2}(\frac{\nabla^2}{3}C + \mathcal{H}\dot{C}) + \mathcal{H}B \tag{13.21}$$

$$\bar{\Phi}_i = \dot{\bar{C}}_i - \bar{B}_i \tag{13.22}$$

$$\tilde{C}_{ij} \tag{13.23}$$

13.2 Metric of scalar modes and common gauges

We now will look at two specific gauges for scalar modes of the metric perturbation (13.1). Due to the SVT decomposition, scalar modes can be treated separately from the vector and tensor modes[80]. Experimentally, only scalar modes have been detected so far[61]. In fact, it seems that the whole universe can be described by these scalar modes alone, at least at early times when the fluctuations don't grow outside the perturbative regime. Scalar modes are the most important for cosmology.

The first gauge we discuss is the *conformal-Newtonian gauge*[155]

$$ds^2 = a^2(\tau)\left[-(1+2\psi)d\tau^2 + (1-2\phi)dx^i dx_i\right] \tag{13.24}$$

The convenience of this metric is that it is diagonal and, in the non-relativistic limit, -2ψ is the Newton gravitational potential and the interpretation is easy. In the same limit the equations of motion will set $\phi = \psi$. One should take care of the signs used to define the potentials ψ and ϕ , since different sources use different conventions. We use the conventions in ref. [155].

In terms of the notation of the previous section $A = \psi$ and $D = -\phi$, the other functions being zero, so that the gauge invariant quantities are $\Phi_A = \psi$ and $\Phi_H = -\phi$. This confirms that the gauge choice is appropriate.

All the perturbations we encounter depend on space. It will be more useful to work with the spatial Fourier transforms, for example

$$\psi(\vec{x}, \tau) = \int d^3k e^{i\vec{k}\cdot\vec{x}}\tilde{\psi}(\vec{k}, \tau) \tag{13.25}$$

Henceforth, we will usually drop the tilde in the Fourier transform indicating $\tilde{\psi}(\vec{k}, \tau) \equiv \psi(\vec{k}, \tau)$. It will be clear by context if we are working with the transforms or not (usually we will be, except for the first present definitions). Indeed, when working with Fourier transforms, spatial derivative will be replaced by $\partial_i\psi \to ik_i\psi$. Physically, $\vec{k}$ lives on the three-dimensional spatial slicing, and has the interpretation of a comoving wave vector.

We define the raising and lowering in a trivial manner $k^i = k_i$[10]. Its "wavelength" $\lambda = \frac{2\pi}{k}$ can be thought of as a comoving wavelength.

The other common gauge, used in calculations and Boltzmann codes is the *synchronous gauge*

$$ds^2 = a^2(\tau)\left[-d\tau^2 + (\delta_{ij} + h_{ij})dx^i dx^j\right] \tag{13.26}$$

The perturbation h_{ij} lives on the spatial-slicing, so again we may raise and lower its indexes simply $h^i_j = h_{ij}$. It is obviously symmetric and we separate out its trace as

$$h \equiv h^i_i \tag{13.27}$$

With this separation we obtain a spatial part of the metric of a form like (13.1) with $2D = \frac{h}{3}$ and

$$\chi_{ij} = h_{ij} - h\delta_{ij}/3 \tag{13.28}$$

As we had seen the traceless symmetric tensor can be written in terms of a scalar, a vector and a tensor. At this moment we are only interested in the scalar perturbation

$$\chi_{ij} = (\partial_i\partial_j - \frac{1}{3}\delta_{ij}\nabla^2)\mu \tag{13.29}$$

where μ is the scalar degree of freedom, along with h. With the notation of the previous section, we identify $C = \mu$. The scalar gauge invariant quantities are

$$\Phi_A = -\frac{1}{2}(\dot{\mu}\mathcal{H} + \ddot{\mu}) \tag{13.30}$$

$$\Phi_H = \frac{h}{6} - \frac{\nabla^2}{6}\mu - \frac{\mathcal{H}}{2}\dot{\mu} \tag{13.31}$$

The convenience of this gauge is not fully evident at this stage. When we perturbed the matter density of the universe, we will be perturbing the energy-momentum tensor of a perfect fluid (7.17), which will acquire not only density fluctuations but peculiar velocities through a perturbed u^μcomoving vector. By choosing the synchronous gauge, it will happen that we can set the velocity perturbations to zero for a pressureless matter component, such as dark matter. This simplifies the Boltzmann equations to be solved. Indeed, the above form of the metric is not sufficient to fix the synchronous gauge condition, since one can make a small coordinate transformation which keeps the same form but mixes μ and h. We make the most common choice, to set the peculiar velocity of dark matter to zero in the synchronous gauge.

Another convenience of this gauge is that it is very simple to add vector and tensor perturbations if we wanted to. In fact, we would simply choose a more general form for χ_{ij} as in (13.5).

In Fourier space, which we use most often, the physical degrees of freedom we use are

[10]Equations in Fourier space are no longer explicitly covariant. Just by taking the Fourier transform we have split space and time. k_i is not a tensor since the derivative ∂_i isn't either.

defined by the relation

$$h_{ij}(\vec{x},\tau) \equiv \int d^3k e^{i\vec{k}\cdot\vec{x}} \left[\hat{k}_i\hat{k}_j h(\vec{k},\tau) + (\hat{k}_i\hat{k}_j - \frac{1}{3}\delta_{ij})6\eta(\vec{k},\tau)\right] \tag{13.32}$$

With $\hat{k}_i = k_i/k$. The η and h will be the degrees of freedom that appear in the synchronous gauge equations. While $h(\vec{k},\tau)$ is the Fourier transform of $h(\vec{x},\tau)$, η is not the Fourier transform of μ. The relationship can be found by taking the Fourier transform of $h\delta_{ij}/3 + (\partial_i\partial_j - \delta_{ij}\nabla^2/3)\mu$

$$\int d^3k e^{i\vec{k}\cdot\vec{x}}\tilde{h}(\vec{k},\tau)\frac{\delta_{ij}}{3} - k^2(\partial_i\partial_j - \frac{\delta_{ij}}{3}\nabla^2)\tilde{\mu}(\vec{k},\tau) \tag{13.33}$$

We want this to be equal to the Fourier transform for h_{ij} , so we can drop the $d^3k e^{i\vec{k}\cdot\vec{x}}$ integral, by the Fourier inversion theorem, and find that

$$\tilde{\mu} = -\frac{1}{k^2}(h + 6\eta) \tag{13.34}$$

Which implies that the Fourier transform of χ_{ij} is

$$\tilde{\chi}_{ij} = (\hat{k}_i\hat{k}_j - \delta_{ij}/3)(h + 6\eta) \tag{13.35}$$

Let's find how to relate the quantities in the two gauges. By using the gauge invariant quantities(13.30) and (13.31) we obtain, in real space,

$$\psi(\vec{x},\tau) = -\frac{1}{2}(\dot{\mu}\mathcal{H} + \ddot{\mu}) \tag{13.36}$$

$$\phi(\vec{x},\tau) = -\frac{h}{6} + \frac{1}{2}(\frac{1}{3}\nabla^2\mu + \mathcal{H}\dot{\mu}) \tag{13.37}$$

Passing to Fourier space,

$$\psi(\vec{k},\tau) = \frac{1}{2k^2}(\ddot{h} + 6\ddot{\eta} + \mathcal{H}(\dot{h} + 6\dot{\eta})) \tag{13.38}$$

$$\phi(\vec{k},\tau) = \eta + \frac{\mathcal{H}}{2k^2}(\dot{h} + 6\dot{\eta}) \tag{13.39}$$

It will also be useful to determine what coordinate change is needed to switch between a gauge and the other. Suppose we take a coordinate change as in (13.6) $y^\mu = x^\mu + \xi^\mu$. Here y^μ are supposed to be the coordinates in the synchronous gauge and x^μ in the conformal-Newtonian one. We can take the spatial term to be a gradient $\xi_i = S_{,i}$, in the same way as in equation (13.7), but with only the scalar term. The vector term isn't needed since we are dealing with scalar perturbations, or alternatively, if we added it, it's trivial to show it must be zero. By using the relation between the quantity C in the two gauges we obtain, from (13.17) using the unprimed quantities as those in the synchronous gauge,

$$S = -\frac{\mu}{2} \tag{13.40}$$

Using (13.13)

$$\dot{S} = \xi^0 \tag{13.41}$$

In Fourier space, the coordinate transformation is identified by the function

$$S(\vec{k}, \tau) = \frac{1}{2k^2}(h + 6\eta) \tag{13.42}$$

It can be of use to give the Christoffel symbols explicitly. Their calculation is trivial, but lengthy. The zero-order terms are given in .(6.41). Here we give the first order terms $\delta\Gamma^{\mu}_{\rho\sigma}$, the total symbol being $\Gamma^{\mu}_{\rho\sigma} =^{(0)} \Gamma^{\mu}_{\rho\sigma} + \delta\Gamma^{\mu}_{\rho\sigma}$. For the conformal-Newtonian gauge, in Fourier space,

$$\text{(Newtonian)}\delta\Gamma^{\tau}_{\tau\tau}(\vec{k}, \tau) = \dot{\psi} \tag{13.43}$$

$$\text{(Newtonian)}\delta\Gamma^{\tau}_{\tau i} = ik_i\psi \tag{13.44}$$

$$\text{(Newtonian)}\delta\Gamma^{\tau}_{ij} = -\delta_{ij}\left(\dot{\phi} + 2\mathcal{H}(\phi + \psi)\right) \tag{13.45}$$

$$\text{(Newtonian)}\delta\Gamma^{i}_{\tau\tau} = ik^i\psi \tag{13.46}$$

$$\text{(Newtonian)}\delta\Gamma^{i}_{\tau j} = -\delta^i_j\dot{\phi} \tag{13.47}$$

$$\text{(Newtonian)}\delta\Gamma^{i}_{jk} = i\phi(\delta_{jk}k^i - \delta^i_j k_k - \delta^i_k k_j) \tag{13.48}$$

For the synchronous gauge, in terms of the Fourier transform $\chi_{ij} = (\hat{k}_i\hat{k}_j - \delta_{ij}/3)(h + 6\eta)$

$$\text{(Synchronous)}\delta\Gamma^{\tau}_{ij} = \chi_{ij}\mathcal{H} + \frac{\dot{\chi}_{ij}}{2} + \delta_{ij}(\mathcal{H}\frac{h}{3} + \frac{\dot{h}}{6}) \tag{13.49}$$

$$\text{(Synchronous)}\delta\Gamma^{i}_{\tau j} = \frac{\dot{\chi}_{ij}}{2} + \frac{\delta_{ij}}{6}\dot{h} \tag{13.50}$$

$$\text{(Synchronous)}\delta\Gamma^{i}_{jk} = \frac{i}{2}(k_k\chi_{ij} + k_j\chi_{ik} - k_i\chi_{jk}) + \frac{ih}{6}(\delta^i_j k_k + \delta^i_k k_j - \delta_{jk}k_i) \tag{13.51}$$

The other symbols being zero. Most importantly we give the first-order components of the Einstein tensor. In the conformal-Newtonian gauge

$$\text{(Newtonian)}\delta G^{\tau}_{\tau} = \frac{6\mathcal{H}}{a^2}(\mathcal{H}\psi + \dot{\phi}) + \frac{2k^2}{a^2}\phi \tag{13.52}$$

$$\text{(Newtonian)}\delta G^{\tau}_{i} = -\delta G^{i}_{\tau} = -\frac{2ik_i}{a^2}(\mathcal{H}\psi + \dot{\phi}) \tag{13.53}$$

$$\text{(Newtonian)}\delta G^{i}_{j} = \frac{k^i k_j}{a^2}(\psi - \phi) + \delta^i_j\left[\frac{2}{a^2}\ddot{\phi} + 4\frac{\ddot{a}}{a^3}\psi + \frac{2\mathcal{H}}{a^2}(\dot{\psi} + 2\dot{\phi}) - 2\frac{\mathcal{H}^2}{a^2}\psi - \frac{k}{a^2}(\psi - \phi)\right] \tag{13.54}$$

Of which the trace and longitudinal traceless space parts are

$$(\text{Newtonian})\frac{1}{3}\delta G^i_{\ i} = \frac{2}{a^2}\ddot{\phi} + \frac{4\ddot{a}}{a^3}\psi + \frac{2\mathcal{H}}{a^2}(\dot{\psi} + 2\dot{\phi}) - \frac{4\mathcal{H}^2}{a^2}\psi - \frac{2k^2}{3a^2}(\psi - \phi) \tag{13.55}$$

$$(\text{Newtonian})(\hat{k}^j\hat{k}_i - \frac{\delta^j_i}{3})\delta G^i_{\ j} = \frac{2k^2}{3a^2}(\psi - \phi) \tag{13.56}$$

Whereas in the synchronous gauge

$$(\text{Synchronous})\delta G^\tau_{\ \tau} = 2\frac{k^2}{a^2}\eta - \frac{\mathcal{H}}{a^2}\dot{h} \tag{13.57}$$

$$(\text{Synchronous})\delta G^\tau_{\ i} = -\frac{2ik_i}{a^2}\eta \tag{13.58}$$

$$(\text{Synchronous})\delta G^i_{\ j} = (k^ik_j - \frac{\delta^i_j}{3}k^2)\left[\frac{\ddot{h} + 6\ddot{\eta}}{2a^2k^2} + \frac{\mathcal{H}}{a^2k^2}(\dot{h} + 6\dot{\eta}) - 3\frac{\eta}{a^2}\right] + 2k^ik_j\eta - \delta^i_j(\frac{\ddot{h}}{3a^2} + \frac{2\mathcal{H}}{3a^2}h) \tag{13.59}$$

The trace and longitudinal traceless space parts can be read off

$$(\text{Synchronous})\delta G^i_{\ i} = \frac{2k^2}{a^2}\eta - \frac{\ddot{h}}{a^2} - \frac{2\mathcal{H}}{a^2}\dot{h} \tag{13.60}$$

$$\begin{aligned} &(\text{Synchronous}) \\ &(\hat{k}^i\hat{k}_j - \frac{\delta^i_j}{3})\delta G^j_{\ i} \quad = \quad \frac{1}{3a^2}(\ddot{h} + 6\ddot{\eta}) + \frac{2}{3a^2}\mathcal{H}(\dot{h} + 6\dot{\eta}) - \frac{2}{3}\frac{k^2}{a^2}\eta \end{aligned} \tag{13.61}$$

The spatial part of the Einstein tensor is separated this way because it will allow us to decouple scalar, vector and tensor perturbations.

13.3 Tensor modes

Tensor modes in the primordial universe have not been observed yet, however they are predicted to arise in non-standard cosmological scenarios. The most notable case is inflation, which predicts a small background of tensor perturbations.

As we had shown in (13.19), the tensor perturbation can be described through a traceless and divergenceless tensor h_{ij}, which does not change under small coordinate changes. The metric of a tensor perturbation only is[80]

$$ds^2 = a^2(\tau)\left(-d\tau^2 + (\delta_{ij} + h_{ij})dx^i dx^j\right) \tag{13.62}$$

which is the same notation we used when defining the synchronous gauge (13.26), but h_{ij} is a different object. This notation, which is the standard one, may appear to be somewhat confusing, however this confusion will be surpassed momentarily as we introduce the degrees of freedom we are working with.

In Fourier space the divergenceless property $h^{ij}_{,j} = 0$ becomes

$$k^i h_{ij} = 0 \tag{13.63}$$

The tensor perturbation is transverse, orthogonal to its wave-vector. This turns out to have a deep meaning. Tensor perturbations are none-other than gravitational waves, as anyone familiar with them might have guessed. In the usual treatment of gravitational waves, one perturbs around the Minkowski space-time $\eta_{\mu\nu}$ with a small perturbation $\tilde{h}_{\mu\nu}$. We are perturbing around a FRW metric, which is conformally flat

We now write the metric for a tensor perturbation explicitly, taking the wave vector $k^i = k\delta^i_z$ as pointing in the z direction. As long as we are studying only one single Fourier component, a *gravitational plane wave*, we may always choose the spatial coordinates such that $\vec{k} \parallel \hat{z}$. We will have to be careful when summing over many waves propagating in different directions. Then $k^i h_{ij} = 0 \implies h_{iz} = 0$, so, recalling the traceless condition, we choose h_{ij} to be of the form

$$h_{ij} = \begin{pmatrix} h_+ & h_\times & 0 \\ h_\times & -h_+ & 0 \\ 0 & 0 & 0 \end{pmatrix} \tag{13.64}$$

where $h_{+,\times}$ are functions of conformal time τ and the magnitude k.

The Einstein tensor can be computed and we give in the two combinations

$$\delta G^x_x - \delta G^y_y = \frac{\ddot{h}_+}{a^2} + 2\frac{\mathcal{H}}{a^2}\dot{h}_+ + \frac{k^2 h_+}{a^2} \tag{13.65}$$

$$\delta G^x_y + \delta G^y_x = \frac{\ddot{h}_\times}{a^2} + 2\frac{\mathcal{H}}{a^2}\dot{h}_\times + \frac{k^2 h_\times}{a^2} \tag{13.66}$$

One can introduce the variable $\tilde{h}_{+,\times} = \frac{h}{a}$ and the right hand sides above become

$$\frac{1}{a^3}\left[\dddot{\tilde{h}} + (k^2 - \frac{\ddot{a}}{a})\tilde{h}\right] \tag{13.67}$$

This formula makes clear the oscillatory nature of the metric, which acts as a harmonic oscillator with a time dependent mass. The right hand side of the Einstein equations would then represent forcing terms on this oscillator.

13.4 Matter Perturbations

The right hand side of the Einstein equations need the matter perturbations. We assume the universe can be described as a nearly-perfect fluid. Thus we take the expression for the energy-momentum tensor of a perfect fluid

$${}^{(0)}T^\mu_{\ \nu} = P g^\mu_\nu + (P + \rho) u^\mu u_\nu \tag{13.68}$$

where the four-velocity of a comoving observer in conformal time coordinates is $u^\mu = (\frac{1}{a}, 0, 0, 0)$. We will perturb the pressure, density and local velocity. So $\rho \to \rho + \delta\rho$,

$P \to P + \delta P$ and $u^\mu \to \mu^\mu + a v^\mu$, where we have extracted a term a in the definition of the four-velocity perturbation for future convenience. In addition, we can posit that the perturbed fluid have an anisotropic pressure term $\Sigma_{\mu\nu}$ which is symmetric, traceless and satisfies $u^\mu \Sigma_{\mu\nu} = 0$. This implies $\Sigma_{\mu\nu}$ is a spatial tensor, and we will often "forget" its time indexes, indicating it simply as Σ_{ij}. In the same sense of the metric perturbations previously discussed, all these are small quantities.

Using a generic perturbed metric (13.1), we take care to note that $u_\mu = (-a(1+2A), aB_i)$ and we separate the zero-order part from the perturbation as $u_\mu = \bar{u}_\mu + \delta u_\mu$. We wish that the total velocity perturbation $\mu^\mu + a v^\mu$ have a norm of -1 as is required of a world-line. Then

$$(u^\mu + a v^\mu)(u_\mu + a v_\mu) = -1 \tag{13.69}$$

implies

$$a v_\mu u^\mu = A \tag{13.70}$$

where we have used the fact that the index on v_μ can be raised or lowered with the unperturbed metric. It follows that $v_0 = A$ and $v^0 = -\frac{A}{a^2}$. Now we are ready to write the first order term of the energy-momentum tensor. Note that $g^\mu_\nu = \delta^\mu_\nu$, even when perturbed.

$$\delta T^\mu_\nu = \delta P g^\mu_\nu + (\delta P + \delta\rho) u^\mu \bar{\mu}_\nu + (P+\rho)(a v^\mu \bar{u}_\nu + a u^\mu v_\nu + u^\mu \delta u_\nu) + \Sigma^\mu_\nu \tag{13.71}$$

Explicitly, we obtain

$$\delta T^0_0 = -\delta\rho \tag{13.72}$$

$$\delta T^0_i = (\rho + P)(v_i + B_i) \tag{13.73}$$

$$\delta T^i_0 = (\rho + P)(-v_i) \tag{13.74}$$

$$\delta T^i_j = \delta P \delta^i_j + \Sigma^i_j \tag{13.75}$$

In both the conformal-Newtonian and synchronous gauges the term B_i of the metric is zero. Together with the Einstein tensors that we obtained in the various gauges, these quantities will give us the Einstein equations, which will determine the evolution of space-time. That will not be all of the story. In fact, we will have to somehow relate the perturbed matter quantities to the underlying distribution in phase space, whose evolution is governed by the Boltzmann equations. For future reference, it is useful to define the quantities, in Fourier space

$$(\rho + P)\theta \equiv i k^i \delta T^0_i \tag{13.76}$$

$$(\rho + P)\sigma \equiv -(\hat{k}^i \hat{k}_j - \frac{\delta^i_j}{3})\Sigma^j_i \tag{13.77}$$

In both the conformal-Newtonian and synchronous gauges $\theta = i k^i v_i$. In addition we also

introduce the fractional density change

$$\delta \equiv \frac{\delta\rho}{\rho} \tag{13.78}$$

Let's take a moment to note that the velocity field may be separated into a gradient and a divergenceless vector as

$$v_i = v_{,i} + \bar{v}_i \tag{13.79}$$

This is actually a decomposition into a scalar mode $v_{,i}$ and a vector mode. The vector term $\bar{v}_i$ is related to the vorticity of matter. In most cosmological models the vector modes are very small and we will neglect them. For scalar modes *the velocity field is irrotational.*

The energy-momentum tensor must satisfy a continuity equation, generated by the fact that it is divergenceless. At zero order this is the continuity equation for a perfect fluid (7.21) : $\dot{\rho} = -3\mathcal{H}(\rho + P)$.

$$T^{\mu}_{\nu;\mu} = T^{\mu}_{\nu,\mu} + \Gamma^{\mu}_{\alpha\mu}T^{\alpha}_{\nu} - \Gamma^{\alpha}_{\nu\mu}T^{\mu}_{\alpha} = 0 \tag{13.80}$$

Explicitly, in the *conformal-Newtonian* gauge we obtain, in terms of δ, θ and σ.

$$\dot{\delta} = -3\mathcal{H}(\frac{\delta P}{\rho} - \frac{P\cdot\delta}{\rho}) - (1+\frac{P}{\rho})(\theta - 3\dot{\phi}) \tag{13.81}$$

$$\dot{\theta} = -\theta(\mathcal{H} + \frac{\dot{P}}{\rho+P}) + k^2\psi + k^2\frac{\delta P}{\rho+P} - \sigma k^2 \tag{13.82}$$

And in the *synchronous gauge*

$$\dot{\delta} = -3\mathcal{H}(\frac{\delta P}{\rho} - \frac{P\cdot\delta}{\rho}) - (1+\frac{P}{\rho})(\theta + \frac{\dot{h}}{2}) \tag{13.83}$$

$$\dot{\theta} = -\theta(\mathcal{H} + \frac{\dot{P}}{\rho+P}) + k^2\frac{\delta P}{\rho+P} - \sigma k^2 \tag{13.84}$$

These equations are valid when the energy-momentum tensor describes all the components of the universe. They are not true for any single component, unless it is non-interacting like dark matter. For interacting components the exchange in momentum and energy must be added. If we are describing a single component, these equations can be supplemented by an equation of state describing $P = P(\rho)$, for example the usual $P = w\rho$. For the variations δP , it is often the case that one introduces the adiabatic sound speed defined through $c_s^2 = \frac{dP}{d\rho} = w + \rho\frac{dw}{d\rho}$, taking into account the possibility that w is not a constant. We note that we consider the interactions which transfer energy between components to be first-order so that the zero-order continuity equation is valid for each matter component separately.

Let's see how to relate the matter perturbation in the two gauges, evaluating the perturbations at the same coordinate point (which is not in general the same physical event). Repeating the same argument as we did when transforming the metric (13.8) we obtain

with a coordinate change (13.6)

$$T^{\mu}_{S\nu} = T^{\mu}_{N\nu} + \xi^{\mu}_{,\alpha} T^{\alpha}_{N\nu} - \xi^{\alpha}_{,\nu} T^{\mu}_{N\alpha} - \xi^{a} T^{\mu}_{\nu,\alpha} \tag{13.85}$$

where S indicates the tensor is the synchronous gauge and N in the conformal Newtonian. Recall that we make a coordinate change which affects the scalar modes only so $\xi^0 = \dot{S}$, $\xi^i = S_{,i} \rightarrow ik_i S$ and S is given by (13.42). For the specific quantities we have

$$\delta_S = \delta_N - \frac{\dot{\rho}}{\rho}\dot{S} \tag{13.86}$$

$$\theta_S = \theta_N - \dot{S}k^2 \tag{13.87}$$

$$\delta P_S = \delta P_N - \dot{S}\dot{P} \tag{13.88}$$

$$\sigma_S = \sigma_N \tag{13.89}$$

Finally, let's return to the issue we had raised following our definition of the synchronous gauge metric (13.26). We had pointed out that there is a class of coordinate changes under which the metric takes the same form. The generic perturbation quantities A, B defined in (13.1) change via equations (13.11) and (13.11), so it is clear that in order to keep $A = B = 0$ before and after a transformation of the scalar modes, we must have

$$\dot{\xi}^0 = -\mathcal{H}\xi^0 \implies \xi^0 = fa^{-1} \tag{13.90}$$

for some function of space $f = f(\vec{x})$ and

$$\dot{S} = \xi^0 \tag{13.91}$$

Now, assume that the dark matter has a peculiar velocity field v^i in some frame. By equation (13.85) in Fourier space we obtain the velocity field $v_i^{'}$ in the new coordinate system

$$v_i^{'} = v_i + ik_i f a^{-1} \tag{13.92}$$

And taking the contraction with ik^i

$$\theta^{'} = \theta - k^2 f a^{-1} \tag{13.93}$$

If the time dependence for θ were as $a^{-1}(\tau)$ then there is enough freedom in f , which is a function of space alone, to set $\theta^{'} = 0$. Regardless of the speed of dark matter, we can always set it to zero in the synchronous gauge. In fact, *setting $\theta_{DM} = 0$ is required to fix the synchronous gauge completely.* Can this be done? Yes, by analyzing the equation for θ in the synchronous gauge, since dark matter is pressureless $\delta P = \dot{P} = \sigma = 0$ and the equation for θ becomes

$$\dot{\theta} = -\mathcal{H}\theta \tag{13.94}$$

which gives precisely the required time dependence.

With this in mind we can give our first result, which is the perturbation equations for dark matter in the synchronous and conformal-Newtonian gauge. These are simply the continuity equations for a pressureless and non-interacting non-relativistic fluid, with the added bonus of one equation being fixed in the synchronous gauge[155].

$$\text{(Synchronous)}\dot{\delta}_{cdm} = -\frac{\dot{h}}{2} \tag{13.95}$$

$$\text{(Synchronous)}\theta_{cdm} = 0 \tag{13.96}$$

$$\text{(Newtonian)}\dot{\delta}_{cdm} = -\theta_{cdm} + 3\dot{\phi} \tag{13.97}$$

$$\text{(Newtonian)}\dot{\theta}_{cdm} = -\mathcal{H}\theta_{cdm} + k^2\psi \tag{13.98}$$

13.5 The energy-momentum tensor from the perturbed distribution

So far, we have found the perturbed Einstein equations by giving the perturbed Einstein tensors and the perturbed energy-momentum tensors. To continue our quest to write a complete set of differential equations, we need to connect the energy momentum tensor to an underlying distribution of states. This distribution will evolve according to the Boltzmann equations. Recall, as in (7.3), that the energy momentum tensor is given by

$$T^{\mu}_{\ \nu} = \int \frac{d_3p}{(2\pi)^3} \frac{1}{\sqrt{-g}} \frac{p^{\mu}p_{\nu}}{p^0} f(\tau, x^i, p_j) \tag{13.99}$$

where $p_\mu = g_{\mu\nu}\frac{dx^\nu}{ds}$ are the canonically conjugate momenta to x^μ and $d_3p = dp_1dp_2dp_3$. When we perturb the metric, these x^μ and p_μ remain canonically conjugate, but they are cumbersome to work with. So we will use different variables in a process not dissimilar to the discussion following (8.2), when dealing with the unperturbed distribution. We will now definitively break any explicit covariance. This shouldn't surprise, as even specifying a distribution $f(x^i, p_i)$ means selecting a frame of reference.

We are still working in perturbation theory, thus we shall assume that f is a distribution close to being either Bose-Einstein or Fermi-Dirac. Denoting $f_0(E, T(\tau))$ as the background, unperturbed distribution, with temperature T, we write

$$f(\tau, \vec{x}, \vec{p}) = f_0(E, T)\,(1 + \Psi(\tau, \vec{x}, \vec{p})) \tag{13.100}$$

where Ψ encodes the perturbation in the distribution. Not only can the distribution now depend on position, but it is no longer assumed isotropic. It depends explicitly on the direction of $\vec{p}$ and not just on the magnitude. Ψ is assumed to be small in the same sense as the other perturbation variables.

In principle this is all is needed to "connect" the energy momentum tensor to the distribution, but we will change variables to make the dependence similar to the usual special relativistic case. In fact note that $\sqrt{-g}$ now contains explicitly the metric perturbations, which we would like to hide inside a redefinition of variables. Let's work in the conformal-

Newtonian gauge first. We define the variables

$$q^i = q_i \equiv \frac{(1+\phi)}{a} p_i \tag{13.101}$$

which implies, working at first order, $(1+\phi)^{-1} = 1-\phi$,

$$p_i = a(1-\phi)q_i \tag{13.102}$$

$$p^i = \frac{1}{a}(1+\phi)q_i \tag{13.103}$$

With this definition, for a particle of mass m, which may also be zero, $g_{\mu\nu}p^\mu p^\nu = -m^2$ is explicitly

$$-a^2(1+2\psi){p^0}^2 + q^2 = -m^2 \tag{13.104}$$

where $q^2 = q_i q^i$. Defining the energy $E = \sqrt{q^2+m^2}$, we obtain

$$p^0 = \frac{E}{a}(1-\psi) \tag{13.105}$$

and $p_0 = -aE(1+\psi)$. Therefore, the momenta q satisfies the dispersion relation of a particle with mass m and energy E. There is no problem exchanging the variables p_μ with $\vec{q}$ and E in the distribution f, but we must take care in changing the measure of integration. In fact $\frac{d^3q}{(2\pi)^3} f(\vec{x},\vec{q})$ is *not* the number density. The measure $d_3p = a^3(1-3\phi)d^3q$, while the determinant $\sqrt{-g} = a^4(1+\psi-3\phi)$. Putting everything together

$$\frac{d_3p}{\sqrt{-g}}\frac{1}{p^0} = d^3q\frac{1}{E} \tag{13.106}$$

With this, the energy-momentum tensor (13.99) is becoming remarkably similar to the expressions for density and pressure that we are usually accustomed to, written in terms of $\vec{q}$. The terms $p^\mu p_\nu$ are left to discuss. When both indexes are time, the metric perturbations cancel out, since $p^0 p_0 = -E^2(1+\psi)(1-\psi) = -E^2$ and we recover the usual expression for the density of a distribution (with an added $-$sign, as is in the definition of T^μ_ν). The same happens when both indexes are space, and we obtain the same expression for the pressure that we are used to. When one index is time and one is space the cancellation of metric perturbations doesn't occur. However, in this case we note that the integral over $f_0 q^i$ is zero, since it is odd in q^i. Therefore, only the perturbation to the distribution Ψf_0 can contribute to the integral, the corrections due to the metric are second order and can be dropped. To conclude, we separate the magnitude of $\vec{q}$ from its direction.

$$\vec{q^i} = qn^i \tag{13.107}$$

where $\delta_{ij}n^i n^j = 1$. In the rest of our discussion of the perturbed Boltzmann equations, we will use q, n^i as our momentum variables. In these terms

$$T^0_0 = -(\rho+\delta\rho) = -\int \frac{q^2 dq d\Omega_q}{(2\pi)^3} E f_0(E,T)(1+\Psi(\vec{x},\vec{q},\tau)) \tag{13.108}$$

$$T^0_i = -T^i_0 = (\rho + P)v_i = \int \frac{q^2 dq d\Omega_q}{(2\pi)^3} q_i f_0(E,T)\Psi(\vec{x},\vec{q},\tau) \tag{13.109}$$

$$T^i_j = (P+\delta P)\delta^i_j + \Sigma^i_j = \int \frac{q^2 dq d\Omega_q}{(2\pi)^3} \frac{q^i q^j}{E} f_0(E,T)(1+\Psi(\vec{x},\vec{q},\tau)) \tag{13.110}$$

The matter perturbation variables δ (eq. (13.78)), θ(eq. (13.76)) and σ (eq. (13.77)) are explicitly

$$\delta = \frac{\int q^2 dq d\Omega_q f_0 \Psi}{\int q^2 dq d\Omega f_0} \tag{13.111}$$

$$(\rho + P)\theta = \int \frac{q^2 dq d\Omega_q}{(2\pi)^3} i k_i q^i f_0 \Psi \tag{13.112}$$

$$(\rho + P)\sigma = -\int \frac{q^2 dq d\Omega}{(2\pi)^3} (\hat{k}^i \hat{k}^j - \frac{\delta^{ij}}{3}) \frac{q_i q_j}{E} f_0 \Psi \tag{13.113}$$

For non-relativistic species, where $P \simeq 0$, we understand this to be related to the small value of q^i/E which is the velocity of the particle. Thus, neglecting any term $o(\frac{q^2}{E^2})$ is the same as taking the non-relativistic limit. This means that the higher moments of non-relativistic distributions, such as the anisotropic stress σ , can be neglected for non-relativistic species. All the perturbation for a non-relativistic species is approximately given by the values of δ and θ, the local over-density and velocity field. The two non-relativistic species under consideration are dark matter and baryons. For dark matter, all the perturbation is taken to be encoded in θ_{DM} and δ_{DM}. Since it does not interact, its Boltzmann equations will reduce to the continuity equations (13.81) and (13.82) (or (13.83) and (13.84) in the synchronous gauge). For the baryons, the δP terms can sometimes become important for high values of k, since $\frac{\delta P}{\delta \rho} = w$ appears in the equation and is small but non-zero. The σ term for baryons can be safely neglected.

Had we worked in the synchronous gauge the final expressions for T^μ_ν , δ, θ and σ are the same, as long as everything is calculated in the synchronous gauge. The derivation is the same when we define the variables

$$q_i = a(\delta_{ij} + \frac{1}{2}h_{ij})p^j \tag{13.114}$$

where h_{ij} is the total metric perturbation in the synchronous gauge. In this case $P^0 = E/a$.

13.6 Einstein equations and scalar-tensor decomposition

Putting together the results from the previous sections, we can now fully write the Einstein equations. We begin with scalar modes and will work in Fourier space. There are two independent degrees of freedom in the metric, ψ and ϕ in the conformal-Newtonian gauge, h and η in the synchronous gauge, therefore only two independent Einstein equations are needed. We choose the time-time and longitudinal traceless space-space equations, but in many applications other equations can be used.

$$\delta G^\tau_\tau = 8\pi G \delta T^\tau_\tau \tag{13.115}$$

$$(\hat{k}^i\hat{k}_j - \frac{\delta^i_j}{3})(\delta G^j_i - 8\pi G\delta T^j_i) = 0 \tag{13.116}$$

This particular combination of the spatial components is chosen because it projects on the scalar modes. Even if they were present, the tensor and vector modes would not contribute in this equation. We will come back to this point shortly.

In the conformal-Newtonian gauge, combining equations (13.52), (13.56) with the definitions (13.76) and (13.77) we obtain

$$\text{(Newtonian)}k^2\phi + 3\mathcal{H}(\dot{\phi} + \mathcal{H}\psi) = -4\pi G a^2 \sum_i \rho_i\delta_i \tag{13.117}$$

$$\text{(Newtonian)}k^2(\phi - \psi) = 12\pi G a^2 \sum_i (\rho_i + P_i)\sigma_i \tag{13.118}$$

For the synchronous gauge we combine equations (13.57), (13.61),

$$\text{(Synchronous)}k^2\eta - \frac{\mathcal{H}}{2}\dot{h} = -4\pi G a^2 \sum_i \rho_i\delta_i \tag{13.119}$$

$$\text{(Synchronous)}\ddot{h} + 6\ddot{\eta} + 2(\dot{h} + 6\dot{\eta})\mathcal{H} - 2k^2\eta = -24\pi G a^2 \sum_i (\rho_i + P_i)\sigma_i \tag{13.120}$$

where we have allowed the energy momentum tensor to be the sum of several matter components. In most calculations the independent components are considered to be photons, baryons, neutrinos and dark matter. They all contribute to the metric contributions via the Einstein equations. By "baryons" in this setting, it is customary to consider electrons and protons together, since their scattering rate amongst each other is very large and they can be treated as a single fluid.

They will be of use, so we also write the trace space-space and time-space Einstein equations. These are not independent of the above due to the Bianchi identities, but we will not prove it explicitly.

$$\text{(Newtonian)}\ddot{\phi} + \mathcal{H}(\dot{\psi} + 2\dot{\phi}) + \psi(2\frac{\ddot{a}}{a} - \mathcal{H}^2) + \frac{k^2}{3}(\phi - \psi) = 4\pi G a^2 \sum_i \delta P_i \tag{13.121}$$

$$\text{(Synchronous)}\ddot{h} + 2\mathcal{H}\dot{h} - 2k^2\eta = 24\pi G a^2 \sum_i \delta P_i \tag{13.122}$$

$$\text{(Newtonian)}k^2(\dot{\phi} + \mathcal{H}\psi) = 4\pi G a^2 \sum_i (\rho_i + P_i)\theta_i \tag{13.123}$$

$$\text{(Synchronous)}k^2\dot{\eta} = 4\pi G a^2 \sum_i (\rho_i + P_i)\theta_i \tag{13.124}$$

For a tensor perturbation in Fourier space with a wave-vector $\hat{k} \parallel \hat{z}$, we can use the $^1_1{-}^2_2$ and $^1_2{+}^2_1$ Einstein equations. Taking (13.65) and (13.66)

$$\text{(Tensor)}\frac{\ddot{h}_+}{a^2} + 2\frac{\mathcal{H}}{a^2}\dot{h}_+ + \frac{k^2 h_+}{a^2} = 8\pi G(\Sigma_{11} - \Sigma_{22}) \tag{13.125}$$

$$\text{(Tensor)}\frac{\ddot{h}_\times}{a^2} + 2\frac{\mathcal{H}}{a^2}\dot{h}_\times + \frac{k^2 h_\times}{a^2} = 8\pi G(\Sigma_{12} + \Sigma_{21}) \tag{13.126}$$

In most models of the universe, the anisotropic pressure is mainly due to neutrinos and is relatively small during most of the evolution. We will see this better when studying the Boltzmann equations. Now we simply point out that when σ and Σ_{ij} are negligible, the tensor perturbation have no source and represent a freely propagating gravitational wave in an expanding universe. For scalar modes in the conformal-Newtonian gauge it turns out that $\psi \simeq \phi$, a useful approximation which is utilized often.

Now we wish to highlight a further aspect of the scalar-tensor decomposition (we are ignoring vectors, but the same ideas would work out the same). The idea of the decomposition is that scalar perturbations and tensor perturbations *in the metric and matter* completely decouple. The equations of motion can be studied independently.The first question to ask is, if we had considered both tensor and scalar perturbations together, would scalar terms appear in the Einstein equations we have written for tensors, and vice versa? Let's look at the left hand sides first. Consider the space-space Einstein tensor in either the conformal-Newtonian (13.54) or synchronous (13.59) gauge. For a wave-vector $\hat{k} \parallel \hat{z}$ it is obvious that $\delta G_2^1 = \delta G_1^2 = 0$ and $\delta G_1^1 - \delta G_2^2 = 0$. We conclude that a scalar mode would not contribute to the left hand side of the tensor mode equations. Vice versa, tensor modes only contribute the $x-x$, $y-y$, $y-x$ and $x-y$ components of the Einstein tensor. In particular, $\delta G_1^1 = -\delta G_2^2$, so the contributions cancel out when we project on the longitudinal traceless components. Thus, we have proven the left hand side of the Einstein equations have decoupled tensor and scalar modes.

Now the right hand sides. From the above equations, it would seem that since the anisotropic stress Σ_{ij} contributes to both the scalar perturbations, through $\sigma = -(\hat{k}^i\hat{k}^j - \frac{\delta^{ij}}{3})\Sigma_{ij}$, and the tensor perturbations. Where is the decoupling? We must dig a little deeper. In fact, we must make a separation of scalar and tensor modes in the matter perturbations as well. We have not done this yet. The decoupling is done at the level of the perturbation to the distribution of states $f_0\Psi(\vec{x}, \vec{q}, \tau)$. Lets consider the perturbations to the space-space energy momentum tensor, from (13.110),

$$\delta T^i_{\ j} = \int \frac{q^2 dq d\cos\theta_q d\phi_q}{(2\pi)^3}\frac{q^2}{E} n^i n^j f_0 \Psi(q, \cos\theta_q, \phi_q) \tag{13.127}$$

where we have made explicit the angular dependence of the n^i. We define θ_q as the angle between $\hat{n}$ and the z axis and ϕ_q the azimuthal angle. Now consider the difference $\delta T_1^1 - \delta T_2^2$, the ϕ_q dependence in the integral can be summed up as

$$\int d\phi_q \cos 2\phi_q \Psi(\phi_q) \tag{13.128}$$

The ϕ_q dependence in the distribution can be examined by taking the Fourier series in ϕ_q, since Ψ must be a periodic function on ϕ_q. Then expanding

$$\Psi(q, \cos\theta_q, \phi_q) = \Psi_{c0} + \sum_n (\Psi_{sn}\sin n\phi_q + \Psi_{cn}\cos n\phi_q) \tag{13.129}$$

where Ψ_{cn}, Ψ_{sn} are functions of q and $\cos\theta_q$. In the ϕ_q integral, only the term proportional to $\cos 2\phi_q$ is picked out from the sum. This means that the only contribution to the h_+ tensor perturbation comes from a particular term of the distribution which has an azimuthal dependence around the wave-vector $\hat{k}$ (recall, the z axis was arbitrarily chosen parallel to $\hat{k}$). Repeating the same argument with the $\delta T^1_2 + \delta T^2_1$ term, we find that the part of the distribution of states that contributes to matter the $h_\times$ equation has a dependence as $\sin 2\phi_q$.

On the other hand, consider the contribution to the scalar mode equations σ, which is the longitudinal traceless projection of the energy-momentum tensor. With $k^i = \delta^i_3$

$$-(\rho+P)\sigma = (\hat{k}^i\hat{k}^j - \frac{\delta^{ij}}{3})\Sigma_{ij} = \int \frac{q^2 dq d\cos\theta_q d\phi_q}{(2\pi)^3}\frac{q^2}{E}(\cos^2\theta_q - \frac{1}{3})f_0\Psi(q,\cos\theta_q,\phi_q) \quad (13.130)$$

Now it is clear that the contribution to scalar modes arises from a perturbation to the distribution which has *no azimuthal dependence.*

We have completed our proof on the Einstein equations. The matter perturbations which source scalar modes do not source tensor modes and vice versa. We will have to see that this separation happens in the Boltzmann equations as well, to complete our decoupling theorem.

13.7 Vector modes

So far we have foregone discussing vector modes. As we shall see, in absence of a source from matter, they decay very quickly. In standard cosmology, the source term is very small, mainly due to neutrinos. For this reason we will not discuss them much in this text. They should be mentioned for completeness.

As we have seen, tensor modes also are sourced very little in standard cosmology, however their free equation of motion has an oscillatory characteristic which may allow them to survive. They may be cosmologically relevant if they were produced by some mechanism in the earliest moments of the universe. On the other hand, vector perturbations decay away.

As with tensors, it is convenient to describe vector modes in Fourier space. In a generic perturbed metric (13.1), they are described by two divergence-free fields $\bar{B}_i$ and $\bar{C}_i$. In Fourier space this implies $k_i\bar{B}^i = k_i\bar{C}^i = 0$. We choose the z axis to lie parallel to $\hat{k}$, so that $\bar{B}_3 = \bar{C}_3 = 0$. Using their properties under a small gauge transformations (13.18) and (13.14), one of the two quantities can be set to zero. We shall choose $\bar{C}_i = 0$ as our gauge condition. Then the metric is[171]

$$ds^2 = a^2(\tau)\left(-d\tau^2 + 2\bar{B}_i dx^i d\tau + \delta_{ij}dx^i dx^j\right) \quad (13.131)$$

We now introduce the quantity $\beta_i = i\epsilon_{ijk}k_j\bar{B}_k$, the curl in Fourier space. The reason to define this quantity is to project out an actual vector perturbation, in case there were a scalar mode. In the generic case, the time-space component of a metric may contain the scalar mode, as in equation (13.3). This is an aspect of the scalar-vector-tensor decomposition we want to point out. For $k^i = \delta^i_3$ we obtain $\beta_1 = -i\bar{B}_2$, $\beta_2 = i\bar{B}_1$. The $\frac{1}{3}$ Einstein equation is

then

$$\dot{\beta}_2 + 2\mathcal{H}\beta_2 = -16\pi G\Sigma_{13} \tag{13.132}$$

and similarly for β_1.

By repeating the arguments relating to scalar tensor decomposition in 13.6 one can show that the contributions to Σ_{13} are due to terms which have a $\cos\phi_q$ or $\sin\phi_q$ dependence on the azimuth in the distribution. Again the SVT decomposition theorem at play.

In most cosmological models Σ_{13} is very small and we may set it to zero. This implies that the metric perturbations for vector modes decay as

$$\bar{B}_i \propto a^{-2} \tag{13.133}$$

Vector modes are related to vorticity of the matter perturbations. The velocity field that appears in the perturbation of the energy-momentum tensor can be decomposed as $v_i = v_{,i} + \bar{v}_i$, as we had explained in section 13.4 when describing scalar modes. The $\bar{v}_i$ is the vector mode perturbation. To project onto it we take the curl

$$\omega_i = i\epsilon_{ijk}k_j v_k \tag{13.134}$$

Now we can derive the continuity equation for vector modes $T^{\mu}_{\nu;\nu} = 0$. The $\nu = \tau$ does not give any information. The spatial term gives

$$\dot{v}_k(\rho + P) + v_k(\dot{P} + \mathcal{H}(\rho + P)) + \bar{B}_k\mathcal{H}(\rho + P) + ik_l\Sigma_{kl} + ik_k\delta\rho = 0 \tag{13.135}$$

We project out any possible scalar mode in the spirit of the decomposition theorem. In fact, the density perturbation is typical of a scalar mode. We multiply and contract with $i\epsilon_{ijk}k_j$

$$\dot{\omega}_i = -\omega_i(\mathcal{H} + \frac{\dot{P}}{\rho + P}) - \beta_i\mathcal{H} - k_ik_l\epsilon_{ijk}\Sigma_{kl} \tag{13.136}$$

It is straightforward to see that the last term projects out the anisotropic stress terms Σ_{13} and Σ_{23} which, we have pointed out, contribute to vector modes. If these are negligible, we find again a damping equation for the vorticity ω_i. Since β_i is small in this limit, and $\frac{\dot{P}}{\rho+P}$ can be usually thought of as a constant, the vorticity decays approximately with the inverse of the scale factor. Thus, we expect the cosmological fluid to have a negligible vorticity.

Of course, the fact that the right anisotropic stress terms are small should not be taken for granted, but be observed from the Boltzmann equation and the matter interactions which we will describe next.

13.8 Collisionless Boltzmann equation

Let's evaluate the Liouville term of the Boltzmann equation (9.2) to first order in perturbation variables[155]. In the absence of a collision term, the Liovuille terms is all there is and we will have the *collisionless Boltzmann equation*. Since the Dark Matter in the universe does not interact, it satisfies the collisionless equation. For non-relativistic dark matter we had already seen that reference to the underlying distribution are not needed, since the en-

tirety of the perturbations are encoded in the density and velocity fields, whose evolution is given by a continuity equation. For other components it is not as simple. Neutrinos are also collisionless, at eras of usual interest, however we can not neglect the higher moments of the distribution.

We are studying the evolution of a distribution

$$f(\tau, \vec{x}, q, \hat{n}) = f_0(E, \tau)(1 + \Psi(\tau, \vec{x}, q, \hat{n})) \tag{13.137}$$

in terms of the momentum variables q and $\hat{n}$. The Liouville term is the total derivative of f along the world-line of the particle. We are not using canonical variables, but the form is the same

$$\frac{df}{d\tau} = \frac{\partial f}{\partial \tau} + \frac{\partial f}{\partial x^i}\frac{dx^i}{d\tau} + \frac{\partial f}{\partial q}\frac{dq}{d\tau} + \frac{\partial f}{\partial n^i}\frac{dn^i}{d\tau} \tag{13.138}$$

$$\frac{df}{d\tau} = \frac{\partial f}{\partial \tau} + ik_i f\frac{dx^i}{d\tau} + \frac{\partial f}{\partial q}\frac{dq}{d\tau} + \frac{\partial f}{\partial n^i}\frac{dn^i}{d\tau} \tag{13.139}$$

The derivatives of the coordinates with respect to τ are total derivatives, since τ is present in lieu of the affine parameter of the world-line. In the second line we have passed to Fourier space, replacing the spatial derivatives with ik_i and dropping the traditional tilde which indicates Fourier transforms. Working in Fourier space is immensely useful. For small perturbations, modes corresponding to different values of $\vec{k}$ don't mix. This is the well known fact that a linear partial differential equation can be transformed in Fourier space into an infinite number of *decoupled* ordinary differential equations. In cosmology, photon modes remain small and therefore decoupled even today. Matter perturbations grow enough in late time that a non-linear treatment is needed. However, in the early universe they are still small enough to be reduced in this manner. Although the number of equations is formally infinite, in practice the calculation will be done numerically on a grid of vectors $\vec{k}$.

Let's assume only scalar perturbations to the metric, to begin. The equations of motion must be calculated from the geodesic equation. Indeed that will be the first thing we do. We will be working at first order in the perturbation variables. Because f_0 is homogeneous, the derivative $\frac{\partial f}{\partial x^i}$ is first order and we only need the unperturbed value of $dx^i/d\tau$. By definition

$$\frac{dx^i}{d\tau} = \frac{dx^i}{ds}\frac{1}{P^0} = \frac{qn^i}{E} \tag{13.140}$$

where s is the proper time (or an affine parameter, in the case of a massless particle). We have used the zero order fact $p^0 = E/a$ and $p^i = qn^i/a$ regardless of the gauge we use.

Next we examine the term $\frac{\partial f}{\partial n^i}\frac{dn^i}{d\tau}$. The derivative of the distribution with respect to the direction of momentum is first order, since the unperturbed distribution is isotropic. On the other hand, due to homogeneity and isotropy of the unperturbed universe, $\frac{dn^i}{d\tau}$ must also be first order, since the spatial direction of the world-line cannot change. This term is at least second order and we may drop it.

Finally we consider the term containing the derivative with respect to q. This is a first order term, its structure is more elaborate and it depends on the gauge. Mainly, we must find $\frac{dq}{d\tau}$

from the geodesic equations. It is easiest to study the geodesic equation for p^0

$$\frac{dp^0}{ds} = -\Gamma^0_{\mu\nu} p^\mu p^\nu \tag{13.141}$$

In performing the calculation, we must keep track of both the zero-order and first-order term. By using the expression for the Christoffel symbols in the conformal-Newtonian gauge (equation (13.43) and successive) as well as the definitions for $p^i = \frac{1+\phi}{a} q_i$ (equation (13.101)), $p^0 = \frac{E}{a}(1-\psi)$, we obtain

$$\text{(Newtonian)}\Gamma^0_{\mu\nu} p^\mu p^\nu = (E^2+q^2)\frac{\mathcal{H}}{a^2}(1-2\psi) + 2iEk^i q_i \frac{\psi}{a^2} - q^2\frac{\dot{\phi}}{a^2} + E^2\frac{\dot{\psi}}{a^2} \tag{13.142}$$

For the synchronous gauge the relevant Christoffel symbols are equation (13.49) and the successive, while the definition of q^i is (13.114) and $P^0 = E/a$. Putting everything together

$$\text{(Synchronous)}\Gamma^0_{\mu\nu} p^\mu p^\nu = (E^2+q^2)\frac{\mathcal{H}}{a^2} + \frac{q^2(n_i k^i)^2}{2k^2}\frac{\dot{h}}{a^2} + [\frac{3q^2(n_i k^i)^2}{k^2} - q^2]\frac{\dot{\eta}}{a^2} \tag{13.143}$$

Since $E = \sqrt{q^2+m^2}$ in either gauge, the derivative of p^0 can be expressed through the derivative $\frac{dq}{d\tau}$. In the synchronous gauge it is trivial, and so we work explicitly in the conformal-Newtonian

$$\text{(Newtonian)}\frac{dp^0}{ds} = \frac{dp^0}{d\tau}p^0 = p^0\left(\frac{dE}{d\tau} - E\frac{\mathcal{H}}{a}(1-\psi) - \frac{E}{a}\frac{d\psi}{d\tau}\right) \tag{13.144}$$

The derivative of ψ with respect to τ is the total derivative, since we are looking at its change along the world-line. We express it through partial derivatives. Since $\frac{dE}{d\tau} = \frac{q}{E}\frac{dq}{d\tau}$

$$\text{(Newtonian)}\frac{dp^0}{ds} = p^0\left(\frac{q}{E}\frac{dq}{d\tau} - E\frac{\mathcal{H}}{a}(1-\psi) - \frac{E}{a}(\frac{\partial\psi}{\partial\tau} + iqk_i n^i\frac{\psi}{a})\right) \tag{13.145}$$

where we inserted $\frac{dx^i}{d\tau} = p^i = \frac{qn^i}{a}$ at zero-order as the term multiplies ψ, a first order term. Putting together all the pieces, we obtain in the conformal-Newtonian gauge

$$\text{(Newtonian)}\frac{dq}{d\tau} = -q\mathcal{H} + q\dot{\phi} - iEk_i n^i\psi \tag{13.146}$$

And in the synchronous gauge

$$\text{(Synchronous)}\frac{dq}{d\tau} = -q\mathcal{H} - \frac{q}{2k^2}(n_i k^i)^2\dot{h} - 3q[\frac{(n_i k^i)^2}{k^2} - \frac{1}{3}]\dot{\eta} \tag{13.147}$$

We are now ready to write the full collisionless equation. Let's start by retaining only zero-order terms. f_0 depends on time through the temperature and on the momentum magnitude q, so at zero order

$$\frac{\partial f_0}{\partial\tau} - q\mathcal{H}\frac{\partial f_0}{\partial q} = 0 \tag{13.148}$$

Importantly, this is not the correct equation only in the absence of interactions, *but even when they exist*. In fact, any collision term must be proportional to the part of the distribution which deviates from equilibrium, and thus first order. So this zero-order equation is

exact even in the presence of interactions. For massless particles, $q \simeq E$ and since the τ dependence is only in the temperature, assuming a Bose-Einstein or Fermi-Dirac form for f_0, this equation implies that $T \propto a^{-1}$, which is exactly what we'd expect.

In the first order equation, a few terms proportional to the zero-order equation may appear, and they may be removed always (even if there were collisions). Then in the *conformal-Newtonian* gauge we obtain

$$f_0\left(\frac{\partial\Psi}{\partial\tau} + ik\mu\frac{q}{E}\Psi + \frac{\partial \ln f_0}{\partial \ln q}(\dot{\phi} - ik\mu\frac{E}{q}\psi) - q\mathcal{H}\frac{\partial\Psi}{\partial q}\right) = 0 \quad \text{(Newtonian)} \tag{13.149}$$

And in the synchronous gauge

$$f_0\left(\frac{\partial\Psi}{\partial\tau} + ik\mu\frac{q}{E}\Psi + \frac{\partial \ln f_0}{\partial \ln q}[-\frac{1}{6}\dot{h} - \frac{1}{2}(\mu^2 - \frac{1}{3})(\dot{h} + 6\dot{\eta})] - q\mathcal{H}\frac{\partial\Psi}{\partial q}\right) = 0 \quad \text{(Synchronous)} \tag{13.150}$$

Where we have defined the quantity $n^i k_i$ which repeatedly appears as

$$\mu \equiv n_i \hat{k}^i \tag{13.151}$$

This is the cosine of the angle between the momentum of the particle $\vec{q}$ and the Fourier wave-vector $\vec{k}$. Crucially, we observe that the direction of momentum n^i appears in the Boltzmann equations only through this angle. This is another aspect of the scalar-vector-tensor decomposition theorem. As we had discussed in the ending of section 13.6, the scalar modes are characterized by perturbations Ψ which do not have an azimuthal dependence around the wave-vector $\vec{k}$. On the other hand, tensor and vector perturbations do have a specific azimuthal dependence. If we expand Ψ in a Fourier sum over the azimuthal angle ϕ_q, then it is obvious from the above equation that any term which can contribute to the tensor perturbation of the metric does not enter the equation for the terms which contribute to scalar perturbation.

If we considered a tensor mode characterized by a perturbation h_{ij} as in (13.62) we would find, by repeating the above steps, a collisionless Boltzmann equation[80]

$$f_0\left(\frac{\partial\Psi}{\partial\tau} + ik\mu\frac{q}{E}\Psi - \frac{1}{2}\frac{\partial \ln f_0}{\partial \ln q}\dot{h}_{ij}n^i n^j - q\mathcal{H}\frac{\partial\Psi}{\partial q}\right) = 0 \tag{13.152}$$

In the tensor case, the azimuthal angle appears explicitly through the combination $\dot{h}_{ij}n^i n^j$. If we consider $\vec{k} \parallel \hat{z}$, then the tensor perturbation h_{ij} is such that $h_{11} = -h_{22} = h_+$ and $h_{12} = h_{21} = h_\times$, so the azimuthal dependence appears as

$$\dot{h}_{ij}n^i n^j = (1-\mu^2)\left(\dot{h}_+ \cos 2\phi_q + \dot{h}_\times \sin 2\phi_q\right) \tag{13.153}$$

which is exactly the azimuthal dependence we expected.

We have now found a form for the left-hand side of the Boltzmann equations. This equation is valid for Dark matter and neutrinos, however we will have to work on it a little more to arrive to a useful form.

13.9 Massless neutrinos

We shall now consider the Boltzmann equation for massless neutrinos. This is simply the collisionless Boltzmann equation which we have found for scalar modes in both the conformal-Newtonian gauge, equation (13.149), and the synchronous gauge, equation (13.150). In addition, we have derived it for a tensor perturbation (13.152). The formalism we introduce here is useful, and will be used in a more generalized manner for the photon Boltzmann equation, once an appropriate collision term is added.

The Boltzmann equation for a massless particle simplifies, since we can easily integrate out the q dependence and reduce the dimensionality of the distribution. We integrate the Liouville term over the magnitude of momentum, weighing by it and normalizing with the unperturbed distribution.

$$\frac{1}{\int q^2 dq \cdot q f_0} \int q^2 dq \cdot q \tag{13.154}$$

Then we can define the integral as

$$F_\nu(\vec{k}, \hat{n}, \tau) \equiv \frac{\int q^2 dq \cdot q f_0 \Psi(\vec{q}, \vec{k}, \hat{n}, \tau)}{\int q^2 dq \cdot q f_0} \tag{13.155}$$

Where the subscript ν indicates we are discussing neutrinos. Of course, the distributions f_0 and Ψ refer to neutrinos but we will not add a subscript to simplify the notation.

Since it will be of use, let explicitly find the derivative with respect to conformal-time. We recall that f_0 is time dependent and satisfies the zero-order Boltzmann equation(13.148) (which is valid even in the presence of collisions) $\frac{\partial f_0}{\partial \tau} = q\mathcal{H}\frac{\partial f_0}{\partial q}$.

$$\dot{F}_\nu = \frac{\partial F_\nu}{\partial \tau} = \frac{\int q^3 dq (f_0 \frac{\partial \Psi}{\partial \tau} + \frac{\partial f_0}{\partial \tau}\Psi)}{\int q^3 dq f_0} - \frac{\int q^{'3} dq^{'} f_0 \Psi \times \int q^3 dq \frac{\partial f_0}{\partial \tau}}{(\int q^3 dq f_0)^2} \tag{13.156}$$

Using the zero order equation, and integrating by parts in q, we find after some algebra

$$\dot{F}_\nu = \frac{\int q^3 dq f_0 (\frac{\partial \Psi}{\partial \tau} - \mathcal{H} q \frac{\partial \Psi}{\partial q})}{\int q^3 dq f_0} \tag{13.157}$$

This formula is very pleasing since it will allow us to absorb the two terms $\frac{\partial \Psi}{\partial \tau} - \mathcal{H} q \frac{\partial \Psi}{\partial q}$ which appear in every Liouville term, into the time derivative of F_ν.

When we perform the integration over the Liouville-term, due to the cancellation $q/E = 1$, every integral we obtain will be proportional to either F_ν or its time derivative. If the particle were massive, the integrals would be different and we would not be able to simply cast the equation into one for F_ν. For this reason, the treatment of massive neutrinos is more complicated.

Let's begin with only scalar modes present. From this point onward we will assume, when talking about scalar modes, that there is no azimuthal dependence in the distribution. If there were one, an integral over the azimuthal angle $d\phi_q$ would project onto the same equations. When using the conformal-Newtonian gauge we obtain, since the metric perturba-

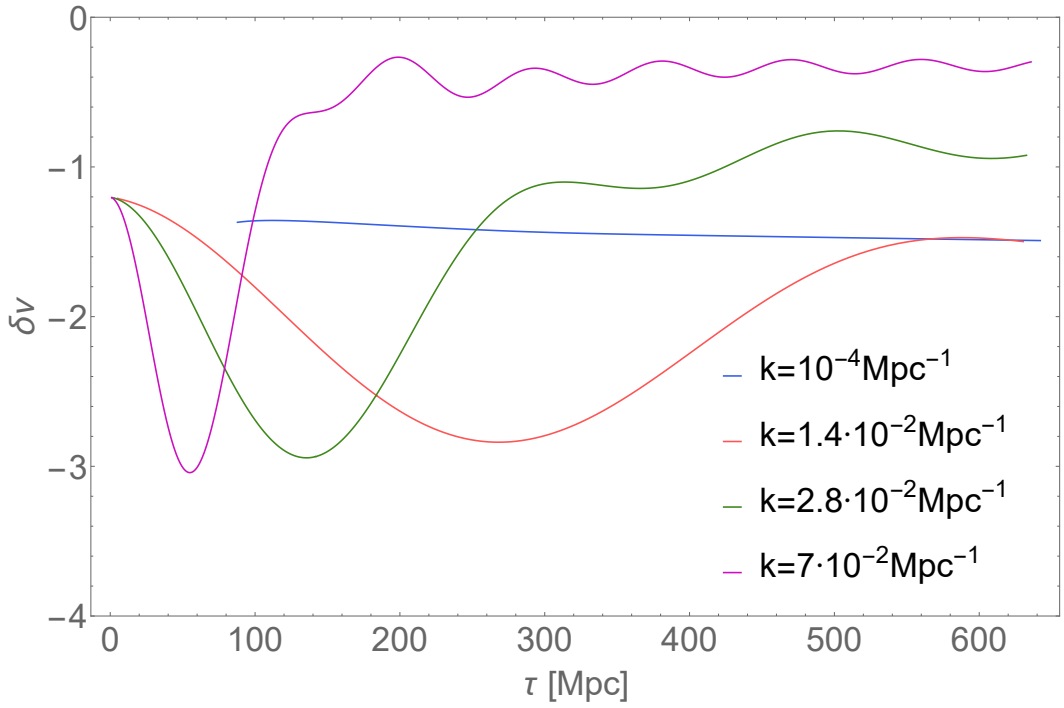

Figure 13.1: Evolution of neutrino density δ_ν for several values of k at early times. The calculation was performed by the CLASS[26] software using best fit ΛCDM parameters[60]. The numerical evolution begins later for the longer wavelength mode $k = 10^{-4} Mpc^{-1}$ since it enters the horizon much later. Its evolution before that time is analytical and described in section 13.15.

tions don't depend on q,

$$\text{(Newtonian)} \dot{F}_\nu + ik\mu F_\nu = 4(\dot{\phi} - ik\mu\psi) \tag{13.158}$$

When using the synchronous gauge

$$\text{(Synchronous)} \dot{F}_\nu + ik\mu F_\nu = -\frac{2}{3}\dot{h} - \frac{4}{3}(\dot{h} + 6\dot{\eta})P_2(\mu) \tag{13.159}$$

where we introduced the *Legendre polynomial of degree* 2 $P_2(\mu) = \frac{1}{2}(3\mu^2 - 1)$. The utility of this will be apparent shortly. First, let's remark how from an infinite number of equations for different values of q we have reduced ourselves to a smaller infinite amount of equations. One for each value of the *magnitude* k and μ.

For tensor perturbations, we separate out the azimuthal dependence which we have discussed at length. Choosing the vector $\vec{k} \parallel \hat{z}$, equation (13.152), we will assume henceforth when talking about tensor perturbations that the perturbed distribution Ψ contains a specific dependence on the azimuthal angle. Referring to the explicit form (13.64), the $+$ term will have a dependence as $\cos 2\phi_q$ and the $\times$ term a dependence as $\sin 2\phi_q$.

$$F_\nu^+(\vec{k}, \mu, \tau) \cdot (1 - \mu^2) \cos 2\phi_q \equiv \frac{\int q^2 dqq \cdot f_0 \Psi(\vec{k}, q, \hat{n}, \tau)}{\int q^2 dqq \cdot f_0} \tag{13.160}$$

$$F_\nu^\times(\vec{k},\mu,\tau)\cdot(1-\mu^2)\sin 2\phi_q \equiv \frac{\int q^2 dqq\cdot f_0\Psi(\vec{k},q,\hat{n},\tau)}{\int q^2 dqq\cdot f_0} \tag{13.161}$$

Again, if we did not assume the particular azimuthal dependence in Ψ, an integral over $\int d\phi_q \cos 2\phi_q$ or $\int d\phi_q \sin 2\phi_q$ would suffice to project onto the wanted component. Doing the integrals and separating the angular dependence, the equation for tensors becomes

$$\text{(Tensor)}\frac{\partial F_\nu^{(i)}}{\partial\tau} + ik\mu F_\nu^{(i)} = -2\dot{h}_i \tag{13.162}$$

where $i = +,\times$.

Separating out the various angular dependencies is a useful concept and essentially has shown us that different azimuthal dependencies in the distribution have completely decoupled from each other. We would like to make some similar simplification on the angle μ. We can project the F_ν (including the $F_\nu^{+,\times}$) on Legendre polynomials in μ. Definitions and properties of the Legendre polynomials are given in appendix B. In confronting with other resources, we should take note that the cosmology community, including us, use a different normalization than what can be found on most mathematical textbooks. The functions F_ν are expanded in μ independent terms as follows

$$F_\nu(k,\mu,\tau) = \sum_{\ell=0}^{\infty}(-i)^\ell(2\ell+1)F_{\nu\ell}(k,\tau)P_\ell(\mu) \tag{13.163}$$

since the Legendre polynomials are an orthogonal and complete base of functions on $\mu \in [-1,1]$. Conversely, by integrating over $\int d\mu P_\ell(\mu)$ one can project on any individual component.

$$F_{\nu\ell}(k,\tau) = \frac{1}{(-i)^\ell}\int_{-1}^{1}\frac{d\mu}{2}P_\ell(\mu)F_\nu(k,\mu,\tau) \tag{13.164}$$

These projections have an immediate physical meaning. In fact, the fractional density perturbation is (see (13.111))

$$\delta_\nu = \frac{\int q^2 dqd\Omega_q q f_0\Psi}{\int q^2 dqd\Omega_q q f_0} = \int_{-1}^{1}\frac{d\mu}{2}F_\nu(k,\mu,\tau) = F_{\nu 0} \tag{13.165}$$

where we used the fact that f_0 does not depend on the angle and $P_0(\mu)=1$. The velocity θ is (from equation (13.112) using $P=\rho/3$),

$$\theta_\nu = \frac{3}{4}\int_{-1}^{1}\frac{d\mu}{2}ik\mu F_\nu(k,\mu,\tau) = \frac{3}{4}kF_{\nu 1} \tag{13.166}$$

And the anisotropic stress term σ is (equation (13.113)), taking care for the $(-i)^2$ term in the definition,

$$\sigma_\nu = \frac{1}{2}F_{\nu 2} \tag{13.167}$$

The Legendre polynomials satisfy a recurrence relation

$$(\ell+1)P_{\ell+1}(x) - (2\ell+1)xP_\ell(x) + \ell P_{\ell-1}(x) = 0 \tag{13.168}$$

which we now use to write the equations for $F_{\nu\ell}$. We integrate equations (13.158) and (13.159) successfully with $\int \frac{d\mu}{2} P_\ell(\mu)$. Due to the orthogonality of the polynomials, this will select a specific μ dependence each time. Note that the metric perturbations don't have any μ dependence. In the *conformal-Newtonian gauge*

$$\text{(Newtonian)}\dot{\delta}_\nu = -\frac{4}{3}\theta_\nu + 4\dot{\phi} \tag{13.169}$$

$$\text{(Newtonian)}\dot{\theta}_\nu = k^2(\frac{1}{4}\delta_\nu - \sigma_\nu) + k^2\psi \tag{13.170}$$

$$\text{(Newtonian)}\dot{F}_{\nu\ell} = \frac{k}{(2\ell+1)}\left(\ell F_{\nu(\ell-1)} - (\ell+1)F_{\nu(\ell+1)}\right) \ \ell \geq 2 \tag{13.171}$$

In the *synchronous gauge*

$$\text{(Synchronous)}\dot{\delta}_\nu = -\frac{4}{3}\theta_\nu - \frac{2}{3}\dot{h} \tag{13.172}$$

$$\text{(Synchronous)}\dot{\theta}_\nu = k^2(\frac{1}{4}\delta_\nu - \sigma_\nu) \tag{13.173}$$

$$\text{(Synchronous)}\frac{1}{2}\dot{F}_{\nu 2} = \dot{\sigma}_\nu = \frac{8}{30}\theta_\nu - \frac{3}{10}kF_{\nu 3} + \frac{2}{15}\dot{h} + \frac{4}{5}\dot{\eta} \tag{13.174}$$

$$\text{(Synchronous)}\dot{F}_{\nu\ell} = \frac{k}{(2\ell+1)}\left(\ell F_{\nu(\ell-1)} - (\ell+1)F_{\nu(\ell+1)}\right) \ \ell \geq 3 \tag{13.175}$$

By this decomposition we have traded an infinite number of μ dependent equations for a coupled system of a countable number of equations, through ℓ. Note how the hierarchy of equations never closes, since every equation for $\dot{F}_{\nu\ell}$ depends on $F_{\nu(\ell+1)}$. In practice, these equations are numerically integrated and are cut-off at some high ℓ with some arbitrary choice, resulting in a numerical error. We will find a way to rewrite the solution of these equations in a manner that does not require cutting off at too high an ℓ.

For the tensor modes the same projection can be done and this results in the equations

$$\text{(Tensor)}\dot{F}^{(i)}_{\nu 0} + F^{(i)}_{\nu 1} = -2\dot{h}_i \tag{13.176}$$

$$\text{(Tensor)}\dot{F}^{(i)}_{\nu\ell} = \frac{k}{2\ell+1}[\ell F^{(i)}_{\nu(\ell-1)} - (\ell+1)F^{(i)}_{\nu(\ell+1)}]\,, \ell \geq 1 \tag{13.177}$$

where $i = +, \times$. We remember that because of the azimuthal dependence, which we have here simplified, the $F^{(i)}_{\nu 0,1}$ do not contribute to the over-density δ_ν or velocity θ_ν. They do, however, contribute to the anisotropic stress terms Σ_{11}, Σ_{12} and Σ_{22}, as we have seen in the discussion following (13.127).

We have now arrived at a set of coupled equations which describe the evolution of a fluid of relativistic massless neutrinos. If the universe contained all neutrinos, we would be

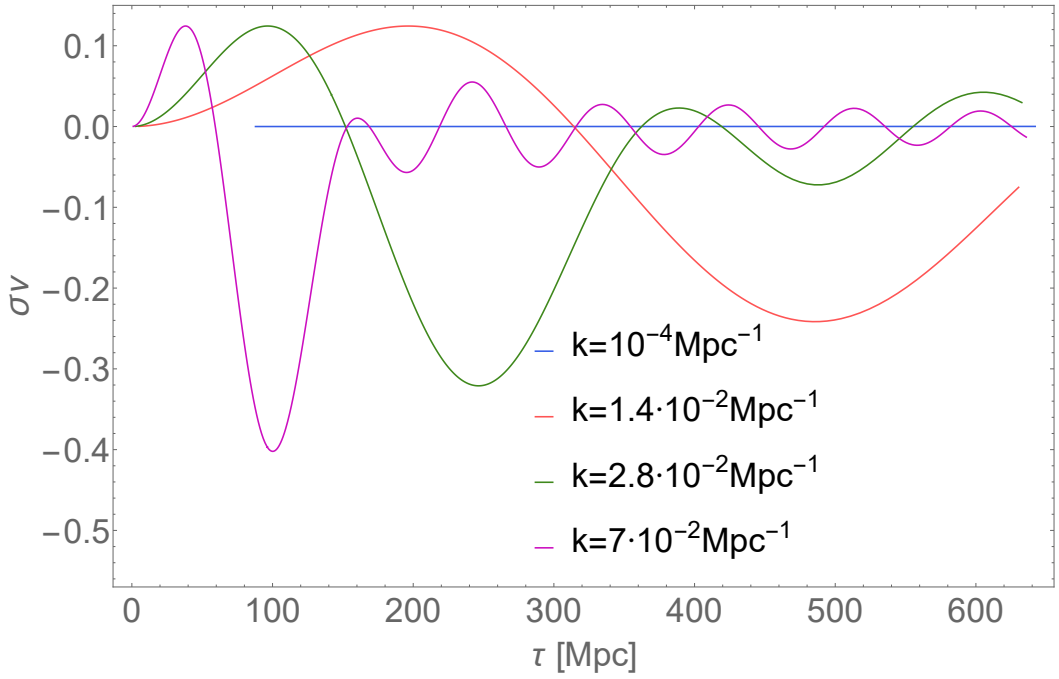

Figure 13.2: Calculation of the neutrino anisotropic stress σ_ν using the Boltzmann software CLASS[26]. The neutrinos are the species which contributes the most to the anisotropic stress.

done. For scalar modes, the equations (13.169)-(13.171), or (13.172)-(13.175), describe the matter content and the Einstein equations (13.117)-(13.118), or (13.119)-(13.120), the metric perturbations. Together, they form a closed system of equations for the variables ψ, ϕ (or h, η) and $F_{\nu\ell}$, for each k. Since the equations at different k's decouple, one would solve these equations numerically for a sequence of values of k. Of course, we haven't yet discussed what observables we can access.

For the tensor modes, in case only massless neutrinos exist, then equations (13.125)-(13.126) and (13.176)-(13.177) form a closed system of equations, for each k.

13.10 Photons polarization

Photons are a massless species. The left hand side of the Boltzmann equation will be the same as for massless neutrinos. We now study the right hand side, namely the collision term. As we shall see, the collision term has an important structure in terms of the polarization of the photon. This structure causes the radiation perturbations to become polarized. This polarization can be measured today in the CMB and is a crucial cosmological signal. Therefore at this point we must treat the dependence of the perturbed distribution $f_0\Psi$ on the polarization.

The polarization properties of electromagnetic radiation is described through the *Stoke parameters* I, Q, U, V[103]. These describe the intensity, I, and the polarization state of the radiation. Q and U indicate a linear polarization and V a circular one. Other polarization states, for example elliptical polarization can be described by a combination of these parameters. For a photon propagating in the direction $+\hat{z}$, we take a basis in the orthogonal plane and designate $\hat{x} - \hat{y}$ axes. The particular orientation of these axes in the orthogonal

plane is arbitrary, a fact we should keep in mind. Of course, the electric field vector $\vec{E}$ of the radiation lies in the orthogonal plane. The intensity of the radiation is given by

$$I =< |\vec{E} \cdot \hat{x}|^2 > + < |\vec{E} \cdot \hat{y}|^2 > \tag{13.178}$$

where the brackets $< \cdot >$ indicate an average over a period of oscillation of the field. The parameter Q is defined as the difference of the averages along the x and y axes.

$$Q =< |\vec{E} \cdot \hat{x}|^2 > - < |\vec{E} \cdot \hat{y}|^2 > \tag{13.179}$$

U is the difference along axes rotated by 45° degrees: $\hat{i} = \frac{1}{\sqrt{2}}(\hat{x} + \hat{y})$, $\hat{j} = \frac{1}{\sqrt{2}}(-\hat{x} + \hat{y})$

$$U =< |\vec{E} \cdot \hat{i}|^2 > - < |\vec{E} \cdot \hat{j}|^2 >= 2Re < E_x E_y^* > \tag{13.180}$$

V is the difference between the complex axes $\hat{l} = \frac{1}{\sqrt{2}}(\hat{x} + i\hat{y})$, $\hat{r} = \frac{1}{\sqrt{2}}(\hat{x} - i\hat{y})$ where the imaginary unit i adds a $\frac{\pi}{2}$ phase difference between the components, thus projecting on left or right circular polarization

$$V =< |\vec{E} \cdot \hat{l}|^2 > - < |\vec{E} \cdot \hat{r}|^2 > \tag{13.181}$$

The intensity I is always larger than zero, while Q, U and V may have any sign. When $Q = I$, $U = V = 0$, the radiation is linearly polarized along the $\hat{x}$ axis. When $Q = -I$ it linearly polarized along the $\hat{y}$ axis. If $|Q| < I$, $U = V = 0$ then the radiation is partially linearly polarized. In the same way, a pure value of U represents a linear polarization at forty-five degrees, while a pure value of V a circular polarization (left-handed if $V = I$ and right-handed if $V = -I$). The four parameters are collected into a polarization matrix

$$T = I\mathbf{1} + U\sigma_1 + V\sigma_2 + Q\sigma_3 = \begin{pmatrix} I + Q & U + iV \\ U - iV & I - Q \end{pmatrix} \tag{13.182}$$

where σ_i are the Pauli matrices and $\mathbf{1}$ is the identity matrix. The polarization matrix fully describes the polarization state and the intensity of the radiation. What is the take away point when we return to our Boltzmann equations? The distributions involved $f = f_0(1 + \Psi)$ describe a distribution in energy: either total intensity or polarized intensity (Q, U, V). In order to fully describe the distribution of photons, including their polarization, a total of *four distribution functions* are needed and their evolution must be tracked. We will use the notation

$$f(\vec{q}, \hat{\epsilon}) = f_0(\vec{q})(\frac{1}{2} + \Psi(\vec{q}, \hat{\epsilon})) \tag{13.183}$$

to indicate the distribution of intensity of radiation with wave-vector $\vec{q}$ and electric field projected along $\hat{\epsilon}$. Note that, by assumption, the zero-order distribution does not depend on $\hat{\epsilon}$, and we define f_0 with a $g_i = 2$ inside, to account for both polarization states. Naturally then we must sum over the different polarizations explicitly.

We can define the I, Q, U and V distributions by

$$f_I(\vec{q}) = f(\vec{q}, \hat{x}) + f(\vec{q}, \hat{y}) \tag{13.184}$$

$$f_Q(\vec{q}) = f(\vec{q},\hat{x}) - f(\vec{q},\hat{y}) = f_0(\vec{q})(\Psi(\vec{q},\hat{x}) - \Psi(\vec{q},\hat{y})) \tag{13.185}$$

and similarly for U and V.

As we pointed out, the choice of $\hat{x}$ and $\hat{y}$ axis is arbitrary, which means that the values of Q and U are dependent on a choice of base on the orthogonal plane. Usually, this problem is overcome by giving an unambiguous prescription for the axes. In cosmology we would like not to make such a prescription for two reasons. The first is due to the fact that isotropy plays an important part even in the perturbed equations, thus it would be preferable to work with quantities which are independent of the choice of basis. The second reason is that we will end up summing over photons arriving from different directions and the polarization quantities will have to be all rotated to a common basis. Again, it would be easier if we worked with basis independent quantities from the start, removing the need to use complex matrix operations at the end. It will also turn out that the basis-independent quantities have a more interesting physical meaning.

Suppose we rotate the frame of reference of the orthogonal plane *counterclockwise* (right-handed) , as viewed against the direction of propagation, by an angle ψ. The new basis is defined as $\hat{x}^{'} = \hat{x}\cos\psi + \hat{y}\sin\psi$, $\hat{y}^{'} = -\hat{x}\sin\psi + \hat{y}\cos\psi$. Then it is trivial to show from the definitions

$$Q^{'} = Q\cos 2\psi + U\sin 2\psi \tag{13.186}$$

$$U^{'} = -Q\sin 2\psi + U\cos 2\psi \tag{13.187}$$

In fact, the polarization is defined up to rotation of π around the axes. The circular polarization V does not change, as a fixed rotation cannot add or remove it. Because of these transformation laws, the following linear combinations are more useful

$$(Q \pm iU)^{'} = e^{\mp i2\psi}(Q + iU) \tag{13.188}$$

since they transform with a complex phase.

Let's take a step back and think of the broader problem we are dealing with, namely solving the Boltzmann equations. In that context, we may think of Q and U as functions of a wave-vector for the photon $\vec{q}$, and, implicitly, on a choice of basis orthogonal to $\vec{q}$, which we denote $\hat{e}_1$, $\hat{e}_2$. Indeed these functions would be $f_{Q,U}(\vec{q})$ as defined above. We have already separated out the magnitude q and the direction $\hat{n}$. To get a picture, we think of the unit vector $\hat{n}$ as applied to the origin. In our calculations $\hat{n}$ will span every possible direction and trace out a unit sphere. On each point of the sphere we therefore have a quantity $(Q \pm iU)(\hat{n})$. It would feel natural then to expand the angular dependence in terms of a complete basis of functions on the unit-sphere. To any physicist, the first idea that comes to mind must be the spherical harmonics $Y_\ell^m(\theta,\phi)$. This is possible and it is actually what we tacitly did when separating scalar and tensor modes in the massless neutrino distribution in section 13.9. It will also be the best idea when treating the intensity of the photon distribution. However, it is not the best manner of treating polarization since the spherical harmonics do not immediately encode the specific dependence of $(Q \pm iU)(\hat{n})$ on the choice of basis $\hat{e}_1, \hat{e}_2$ which, we point out, is not necessarily the coordinate basis $\hat{e}_\theta, \hat{e}_\varphi$. To encode this transformation it is better to use *spin-weighted spherical harmonics*[128, 115].

Useful properties of spin-weighted spherical harmonics are reviewed in appendix B. Consider a function ${}_sf(\theta,\phi,\{\hat{e}_1,\hat{e}_2\})$ on the unit-sphere, where θ and ϕ are the polar coordinates and $\hat{e}_1,\hat{e}_2$ is the basis on the tangent space at the point (θ,φ). The function ${}_sf$ is said to be spin-s if under a counterclockwise (right-handed) rotation by an angle ψ of the basis around the outgoing radial direction $\hat{n}$, the function changes as

$$ {}_sf(\theta,\phi,\{\hat{f}_1,\hat{f}_2\}) = e^{-is\psi} f(\theta,\phi,\{\hat{e}_1,\hat{e}_2\}) \tag{13.189} $$

where $\hat{f}_i$ are the rotated basis. Note that the radial direction is the of propagation of the photon, consistently with our to derivation of (13.188). Therefore we conclude that $(Q\pm iU)$ is a spin-2 (spin-(-2)) function. Henceforth we will not explicitly write the dependence on spin-s quantities on the tangent space basis to simplify the notation.

The spin-weighted spherical harmonics are denoted ${}_sY_\ell^m(\theta,\phi)$ and transform as (13.189) under rotations of the basis. They form a complete basis for spin-s functions

$$ {}_sf(\theta,\phi) = \sum_{\ell,m} a_{\ell m}^{(s)}\, {}_sY_\ell^m(\theta,\phi) \tag{13.190} $$

Note the sum is on ℓ, m and for spin-s spherical harmonics $\ell \geq |s|$ and $a_{\ell m}^{(s)}$ are complex quantities. Since the dependence on the basis change is encoded in the spin-weighted spherical harmonics, the coefficients $a_{\ell m}^{(s)}$ do not depend on the choice of basis. For our polarization problem, they are useful candidates to describe the basis independent quantities we need. The spin-weighted spherical harmonics are orthonormal and complete

$$ \int d\Omega\, {}_sY_\ell^{m*}\, {}_sY_{\ell'}^{m'} = \delta_{\ell,\ell'}\delta_{m,m'} \tag{13.191} $$

$$ \sum_{\ell m} {}_sY_\ell^{m*}(\theta,\phi)\, {}_sY_\ell^m(\theta',\phi') = \delta(\cos\theta-\cos\theta')\delta(\phi-\phi') \tag{13.192} $$

Note how these properties only pertain to harmonics with the same s. We also point out the parity property. Under a coordinate reflection, but not of the tangent space basis,

$$ {}_sY_\ell^m \to (-1)^\ell\, {}_{-s}Y_\ell^m \tag{13.193} $$

Only the regular spherical harmonics are parity eigenstates.

Given a spin-s function, one can get a spin-$(s+1)$ function by applying a *raising operator* ð[87, 115][11]

$$ ð_s f = -\sin^s\theta \left(\frac{\partial}{\partial\theta} + \frac{i}{\sin\theta}\frac{\partial}{\partial\phi}\right)(\sin^{-s}\theta\, {}_sf) \tag{13.194} $$

Under rotations of the tangent space basis the new function changes as

$$ (ð_s f)' = e^{-i(s+1)\psi}(ð_s f) \tag{13.195} $$

[11] ð(Eth) is a character used in Old and Middle English as well as in the modern Icelandic and Faroese languages. In Icelandic, as well as Old and Middle English, it is pronounced as 'th', such as in '**th**is'. Historically its use was dropped in english in favor of the 'd' or 'th'.

Conversely a spin lowering operator $\bar{\eth}$ exists

$$\bar{\eth}_s f = -\sin^{-s}\theta \left(\frac{\partial}{\partial\theta} - \frac{i}{\sin\theta}\frac{\partial}{\partial\phi} \right) (\sin^s\theta\, {}_s f) \tag{13.196}$$

which produces a spin-$(s-1)$ function from a spin-s one. This is indeed how the spin-weighted spherical harmonics are constructed, by repeatedly applying $\eth$ or $\bar{\eth}$ to the spin-0 harmonics Y_ℓ^m and then finding the appropriate normalization.

Let's return to the polarization of the photons. With all this mathematical machinery, it is clear that the phase-space distribution for the polarizations will have to be written in terms of the spin-weighted spherical harmonics in a form such as[221, 128, 135]

$$(Q \pm iU)(q,\hat{n}) = \sum_{\ell m} P^{\pm}_{\ell m}\, {}_{\pm 2}Y^m_\ell(\hat{n}) \tag{13.197}$$

We will not use exactly this form but one which makes more explicit the behavior under parity transformations of the momentum $\vec{q}$. By parity transformation, we mean a reflection of the coordinates *but not of the basis which is defined by the right handed rule.* When we transform $\hat{r} \to -\hat{r}$, Q does not change since it represents the polarization in the unchanged directions $\hat{e}_i$. However the U, as well as V, are defined via the right-handedness of the original basis (U was defined with a basis rotated counterclockwise around $\hat{r}$), thus, keeping the definition consistent, they pick up a minus sign. This means that under parity $(Q \pm iU) \to (Q \mp iU)$. Using the property (13.193) this implies that

$$P^{\pm}_{\ell m} = (-1)^\ell P^{\mp}_{\ell m} \tag{13.198}$$

It becomes simpler when we separate the real and imaginary terms. $P^{\pm}_{\ell m} = E_{\ell m} \pm iB_{\ell m}$, with $E_{\ell m}$ and $B_{\ell m}$ real. Then a parity transformation may be described by

$$E_{\ell m} \to (-1)^\ell E_{\ell m} \tag{13.199}$$

$$B_{\ell m} \to (-1)^{\ell+1} B_{\ell m} \tag{13.200}$$

So this decomposition has a simple interpretation under parity. In fact $E_{\ell m}$ has the same behavior under parity as the the coefficients of the intensity. If we take

$$I(q,\hat{n}) = \sum_{\ell m} a_{\ell m}(q) Y^m_\ell(\hat{n}) \tag{13.201}$$

Then by the same reasoning it is clear that under a parity transformation

$$a_{\ell m} \to (-1)^\ell a_{\ell m} \tag{13.202}$$

13.11 Photon collision term

Photons, being massless, have the same Liouville term as for neutrinos, which we calculated explicitly in section 13.9. Unlike neutrinos, photons frequently scatter on electrons

and protons and thus their collision term must be taken into account in the treatment of the Boltzmann equation. The precise treatment of this term is essential. The decoupling of photons with baryons determines the form of the CMB fluctuations. Because the $e^- + \gamma$ scattering has a polarization dependency, this will give rise to a polarization in the CMB, which is a powerful cosmological observable.

In the early universe, we must consider scattering of photons on protons and electrons. The matrix element $|\mathcal{M}|^2$ can be calculated from field theory[25, 144]. We are interested in temperatures $T \ll m_e$, so we are safely in the non-relativistic limit. In this limit, the cross section of the scattering of $e^- + \gamma \leftrightarrow e^- + \gamma$ is proportional to m_e^{-2}. This implies that the scattering rate of photons on protons is suppressed by a factor $(m_e/m_p)^2$ relative to that on electrons[80]. We will neglect the scattering of photons on protons in the following. There is another important reason this is a good approximation. In fact, the protons and electrons interact with each other much more frequently and indeed, at relevant times, so frequently that they can be often considered as a single fluid. Because the $e + p$ interaction is so fast, any non-thermal distribution produced in the electron fluid is rapidly transferred to the proton fluid and vice versa. There are also other scattering processes which we may consider, such as Compton scattering with Bremsstrahlung $e^- + \gamma \leftrightarrow e^- + 2\gamma$. The cross section of these other processes is smaller by at least a factor α relative to the Compton scattering.

Before diving into the calculation of the full collision term we will discuss the dependence of $|\mathcal{M}|^2$ on polarization. If we consider an incoming photon with linear polarization $\hat{\epsilon}$ and an outgoing photon with polarization $\hat{\epsilon}^{'}$, the transition matrix element is proportional to[12]

$$|\mathcal{M}^2| \propto |\hat{\epsilon} \cdot \hat{\epsilon}^{'}|^2 \tag{13.203}$$

We wish to eventually understand this structure in a general manner, so we will pass to the Stoke parameters I, Q and U. We will not discuss V since circular polarization can not be generated in Compton scattering and therefore does not appear in this setting. The photon scattering happens in a plane spanned by the incoming and outgoing photon momenta $\vec{q}$ and $\vec{p}$. The angle between $\vec{q}$ and $\vec{p}$ is known as the scattering angle and we denote it as β:

$$\vec{q} \cdot \vec{p} = \cos\beta \tag{13.204}$$

The initial polarization pseudo-vector $\hat{\epsilon}$ lies in the plane orthogonal to $\vec{q}$. A natural basis on this orthogonal plane is $\{\hat{e}_\parallel, \hat{e}_\perp\}$ where the parallel direction $\hat{e}_\parallel$ is taken to lie in the scattering plane while the direction $\hat{e}_\perp$ is orthogonal to it. In the case of forward or backward scattering there is no unique scattering plane, $\beta = 0, \pi$, and the choice of basis is arbitrary but it will turn out no problems arise from this. We choose the directions basis such that $\hat{e}_\parallel - \hat{e}_\perp - \hat{q}$ is a right-handed coordinate system. Therefore, the Stoke factor Q is positive when the incident radiation is polarized in the parallel direction. In the same manner, we

[12] In the non-relativistic limit the cross section is proportional to the projection of the initial polarization onto the final one, squared. This is true in general for elastic non-relativistic scattering of photons with particles much smaller than its wavelength, for example Rayleigh scattering. Rayleigh scattering happens extensively in our atmosphere and is the cause of the color of the sky. The polarization dependence is well known to photographers. When the sun is at the zenith, skylight coming from close to the horizon is polarized parallel to the hoizon. It has also been suggested that insects may use the polarization dependence of the skylight to navigate[145]. This polarization is well described by the geometrical argument we use here.

define a basis in the plane orthogonal to the outgoing momentum $\vec{p}$, taking $\hat{e}^{'}_{\perp} = \hat{e}_{\perp}$.

We consider the incident photon to have parameters I, Q and U. Then, using the basic definitions (13.178), (13.179), (13.180), it is trivial to find that the outgoing photon will have parameters

$$I^{'} = I\frac{1+\cos^2\beta}{2} - Q\frac{\sin^2\beta}{2} \tag{13.205}$$

$$Q^{'} = -I\frac{\sin^2\beta}{2} + Q\frac{1+\cos^2\beta}{2} \tag{13.206}$$

$$U^{'} = U\cos\beta \tag{13.207}$$

To make the notation more compact we define the column vector

$$T = (I, Q+iU, Q-iU)^T \tag{13.208}$$

making use of the quantities $Q \pm iU$. We have seen in the discussion following (13.188), this will allow us to obtain basis-independent quantities. The vectors T before and after scattering are then [128]

$$T_{f,s} = \begin{pmatrix} \frac{1+\cos^2\beta}{2} & -\frac{\sin^2\beta}{4} & -\frac{\sin^2\beta}{4} \\ -\frac{\sin^2\beta}{2} & \frac{(1+\cos\beta)^2}{4} & \frac{(1-\cos\beta)^2}{4} \\ -\frac{\sin^2\beta}{2} & \frac{(1-\cos\beta)^2}{4} & \frac{(1+\cos\beta)^2}{4} \end{pmatrix} T_{i,s} \tag{13.209}$$

We may also include an overall normalization factor, but we will forego this until we write the full cross section. We indicate with a subscript s that we are in the scattering basis.

The basis we have used to define the Stoke parameters are useful in the scattering problem but, precisely due to this fact, it is less useful in our Boltzmann equation. Eventually we will have to integrate over all over angles β and all scattering planes, thus we need to rewrite this scattering matrix in a common basis. Given the direction of the momentum $\hat{q}$ (which we called $\hat{n}$ in the derivation of the Liouville terms, eg. (13.149)) the most natural basis on which to define the Stoke vectors is the coordinate basis on the tangent space of the unit sphere $\{\hat{e}_\theta, \hat{e}_\phi\}$ where (θ, ϕ) are the polar coordinates of $\hat{q}$, which is itself the radial direction. The three vectors $\hat{q} - \hat{e}_\theta - \hat{e}_\phi$ form a right-handed coordinate system. The two basis $\{\hat{e}_\theta, \hat{e}_\phi\}$ and $\{\hat{e}_\parallel, \hat{e}_\perp\}$ are related by a right handed rotation around $\hat{q}$ (counterclockwise when viewed against $\hat{q}$) by an angle α such that $\hat{e}_\parallel = \hat{e}_\theta\cos\alpha + \hat{e}_\phi\sin\alpha$, $\hat{e}_\perp = -\hat{e}_\theta\sin\alpha + \hat{e}_\phi\cos\alpha$. Under such a rotation the quantities $(Q \pm iU) \to e^{\mp 2i\alpha}(Q \pm iU)$ which means the T vector between the two frames transforms as

$$T_{i,s} = \begin{pmatrix} 1 & 0 & 0 \\ 0 & e^{-2i\alpha} & 0 \\ 0 & 0 & e^{2i\alpha} \end{pmatrix} T_i \tag{13.210}$$

The precise value of α could be calculated but it will turn out to be unimportant. In the same manner, after the scattering we want to rotate back to the common basis $\{\hat{e}_{\theta'}, \hat{e}_{\phi'}\}$. Again the rotation is around $\hat{q}$. The rotation angle $-\gamma$ is taken to be negative because we are rotating *back to* the common basis, and its value will be in general different from α since

m	$Y_2^m(\theta,\phi)$	${}_2Y_2^m(\theta,\phi)$
2	$\frac{1}{4}\sqrt{\frac{15}{2\pi}}\sin^2\theta e^{2i\phi}$	$\frac{1}{8}\sqrt{\frac{5}{\pi}}(1-\cos\theta)^2 e^{2i\phi}$
1	$\sqrt{\frac{15}{8\pi}}\sin\theta\cos\theta e^{i\phi}$	$\frac{1}{4}\sqrt{\frac{5}{\pi}}\sin\theta(1-\cos\theta)e^{i\phi}$
0	$\frac{1}{2}\sqrt{\frac{5}{4\pi}}(3\cos^2\theta-1)$	$\frac{3}{4}\sqrt{\frac{5}{6\pi}}\sin^2\theta$
-1	$-\sqrt{\frac{15}{8\pi}}\sin\theta\cos\theta e^{-i\phi}$	$\frac{1}{4}\sqrt{\frac{5}{\pi}}\sin\theta(1-\cos\theta)e^{i\phi}$
-2	$\frac{1}{4}\sqrt{\frac{15}{2\pi}}\sin^2\theta e^{-2i\phi}$	$\frac{1}{8}\sqrt{\frac{5}{\pi}}(1+\cos\theta)^2 e^{-2i\phi}$

Table 13.1: A few (spin-weighted) spherical harmonics. The general formula is given in (B.15). Spin-(-2) harmonics can be obtained by complex conjugation using ${}_sY_l^{m*} = (-1)^{s+m}\,{}_{-s}Y_l^m$.

the $\hat{e}_{i'} \neq \hat{e}_i$, as we are at a different point on the unit sphere. Putting everything together we can relate the vector T encoding the Stoke parameters Q and U in the common basis *before and after the scattering as*

$$T_f = \begin{pmatrix} \frac{1+\cos^2\beta}{2} & -\frac{\sin^2\beta}{4}e^{-2i\alpha} & -\frac{\sin^2\beta}{4}e^{2i\alpha} \\ -\frac{\sin^2\beta}{2}e^{-2i\gamma} & \frac{(1+\cos\beta)^2}{4}e^{-2i(\alpha+\gamma)} & \frac{(1-\cos\beta)^2}{4}e^{2i(\alpha-\gamma)} \\ -\frac{\sin^2\beta}{2}e^{2i\gamma} & \frac{(1-\cos\beta)^2}{4}e^{-2i(\alpha-\gamma)} & \frac{(1+\cos\beta)^2}{4}e^{2i(\alpha+\gamma)} \end{pmatrix} T_{i,s} \tag{13.211}$$

At this point we wish to cast the terms in to quantities which have defined properties under rotations, with our mind on having to integrate over them later. The obvious choice is the use of spin-weighted spherical harmonics ${}_sY_\ell^m(\beta,\alpha)$. For $s=0$ of course these are regular spherical harmonics. Luckily we only need spherical harmonics with $|\ell,s,m| \leq 2$, these are tabulated in 13.1. Doing the substitutions we arrive at

$$T_f = \frac{1}{3}\sqrt{\frac{4\pi}{5}} \begin{pmatrix} Y_2^0(\beta,\alpha) + 2\sqrt{5}Y_0^0(\beta,\alpha) & -\sqrt{\frac{3}{2}}Y_2^{-2}(\beta,\alpha) & -\sqrt{\frac{3}{2}}Y_2^2(\beta,\alpha) \\ -\sqrt{6}\,{}_2Y_2^0(\beta,\alpha)e^{-2i\gamma} & 3\,{}_2Y_2^{-2}(\beta,\alpha)e^{-2i\gamma} & 3\,{}_2Y_2^2(\beta,\alpha)e^{-2i\gamma} \\ -\sqrt{6}\,{}_{-2}Y_2^0(\beta,\alpha)e^{2i\gamma} & 3\,{}_{-2}Y_2^{-2}(\beta,\alpha)e^{2i\gamma} & 3\,{}_{-2}Y_2^2(\beta,\alpha)e^{2i\gamma} \end{pmatrix} T_i \tag{13.212}$$

Now comes the finer point. We notice that (α,β,γ) are the Euler angles which rotate the coordinate system $\{\hat{e}_\theta,\hat{e}_\phi,\hat{q}\}$ into the system $\{\hat{e}_{\theta'},\hat{e}_{\phi'},\hat{p}\}$, using the zyz convention[116, 185]. Euler angles define the rotation between two $x-y-z$ coordinate axes . In our case, we take the z axes to be $\hat{q}$ and $\hat{p}$. The angle β is precisely the angle between the z axes. The other two angles are defined by the angle between the y axes and a *line of nodes* $\hat{N}$. The line of nodes is the line perpendicular to *both the z axes*. In our case, this is direction perpendicular to the scattering plane. Thus, $\cos\alpha = \hat{e}_\phi \cdot \hat{e}_\perp$ and $\cos\beta = \hat{e}_{\phi'} \cdot \hat{e}_\perp$. The angles are *defined as positive if* the y axes goes into the line of nodes with a right handed rotation around $\hat{z}$, which is our case. Therefore we may use the generalized addition theorem (B.21)

$$\sum_m {}_{s_1}Y_\ell^{m*}(\theta,\phi)\,{}_{s_2}Y_\ell^{m*}(\theta',\phi') = \sqrt{\frac{2\ell+1}{4\pi}}\,{}_{s_2}Y_\ell^{-s_1}(\beta,\alpha)e^{-is_2\gamma} \tag{13.213}$$

which arises due to the fundamental connection between spin-weighted spherical harmon-

ics and rotations. With this, the polarization structure in the common basis is

$$T_f = \frac{4\pi}{15}\sum_{m=-2}^{2}\begin{pmatrix} 10\delta_{m0}Y_0^0Y_0^{0'} + Y_2^mY_2^{m'} & -\sqrt{\frac{3}{2}}\,{}_2Y_2^mY_2^{m'} & -\sqrt{\frac{3}{2}}\,{}_{-2}Y_2^mY_2^{m'} \\ -\sqrt{6}Y_2^m\,{}_2Y_2^{m'} & 3\,{}_2Y_2^m\,{}_2Y_2^{m'} & 3\,{}_{-2}Y_2^m\,{}_2Y_2^{m'} \\ -\sqrt{6}Y_2^m\,{}_{-2}Y_2^{m'} & 3\,{}_2Y_2^m\,{}_{-2}Y_2^{m'} & 3\,{}_{-2}Y_2^m\,{}_{-2}Y_2^{m'} \end{pmatrix} T_i \quad (13.214)$$

where the unprimed spherical harmonics imply $Y_\ell^m = Y_\ell^m(\theta,\phi)$ and the primed ones have a complex conjugation $Y_\ell^{m'} = Y_\ell^{m*}(\theta',\phi')$.

This is a main result. We have related the Stoke parameters before the scattering at a momentum $\vec{q}$ with polar coordinates (θ,ϕ) to those after the scattering at (θ',ϕ').

The mathematics often hides the physics but here we have actually discovered, once again, an aspect of the scalar-vector-tensor decomposition. Indeed, the structure of the scattering has neatly separated out the different azimuthal dependencies. We recall, in fact, that ${}_sY_\ell^m \propto e^{im\phi}$ where $\phi$is the azimuthal angle. We notice that we have defined the position of the pole on the sphere which is spanned by $\hat{q}$. Our construction so far has only depended on the rotation connecting (θ,ϕ) to (θ',ϕ'). We still have a freedom to choose where the pole is, with respect to what axis measure θ and where to set the zero of the azimuthal angle. As we will be working in Fourier space with a mode $\vec{k}$ is makes sense to define $\hat{z} \parallel \vec{k}$ so that the polar coordinates described here have the same meaning as those we used elsewhere. With this choice it becomes fully clear that the different values of m in the sum of the scattering matrix refer to scalar modes ($m = 0$), vector modes ($m = \pm 1$) and tensor modes ($m \pm 2$). This decomposition is the same we had achieved less systematically with neutrinos in the definitions (13.160), (13.161).

How can unpolarized radiation become polarized due to Compton scattering? Our freshly calculated matrix gives us a hint, since $(Q+iU)' \ni Y_2^m I$. This means that, once we project on definite angular dependencies, polarized radiation is produced by Compton scattering of radiation with a quadrupole distribution, that is an angular dependency in the incoming radiation which has its maximum and minimum separated by an angle $\frac{\pi}{2}$, for example $Y_2^0 \propto 1 - 3\cos^2\theta$.

To understand why this must be, consider a simple example of four unpolarized beams incoming at $\vec{x} = 0$ from the directions $\pm\hat{x}$ and $\pm\hat{y}$ and scattering towards $+\hat{z}$. The polarization which survives this scattering, is the one orthogonal to $\hat{z}$, but it is different depending on the initial direction of the beam. The rays incoming at $\pm\hat{x}$ will scatter into fully polarized beam with polarization along $\hat{y}$ while the incoming rays at $\pm\hat{y}$ will produce polarization in the $\hat{x}$ direction. If the four beams have the same intensity, a *uniform or monopole distribution*, the outgoing radiation will be unpolarized, as there is the same amount of intensity in both $\hat{x}$ and $\hat{y}$ directions, thus $Q = 0$. A *dipole distribution* does not work as well. In this case the incoming rays from $+\hat{x}$ and $+\hat{y}$ will be more intense than those incoming from $-\hat{x}$, $-\hat{y}$, but the difference between the two perpendicular directions is the same and therefore again the scattering will result in $Q = 0$. Now we understand that to have a polarized final state we need a *quadrupole distribution*, with the intensity coming from $\pm\hat{x}$ higher than that from $\pm\hat{y}$. This will result in the transmission of radiation preferentially polarized along $\hat{y}$. We develop a $Q < 0$ polarization.

Let's now begin calculating the full collision term. We consider the Compton scattering,

happening both ways,

$$e^-(s) + \gamma(q, \hat{\epsilon}) \leftrightarrow e^-(t) + \gamma(p, \hat{\epsilon}^{'}) \tag{13.215}$$

in the non-relativistic limit, when it reduces to Thompson scattering. The energies of the electrons are denoted $E_{s,t}$. Electrons in the non-relativistic limit at temperatures $T \sim 1 \div 10^3 eV$ are distributed according to a Maxwell-Boltzmann distribution and their typical momentum is given by $\frac{p^2}{2m} = \frac{3}{2}T$ which implies $p_e \sim \sqrt{mT} \sim 10^2 \div 10^4 eV$. On the other hand, the average energy of a photon is $E_\gamma \sim T$. This allows us to estabilish a hierarchy of quantities $m_e \sim E_e \gg p_e \gg E_\gamma = p_\gamma$ and we will use these scale differences to approximate our result.

We have made explicit the polarization of the photon before and after and we shall proceed by fully specifying both the initial and final polarization state. The full collision term would have to account for scattering into both polarizations $\hat{\epsilon}^{'}$. A general collision term is given by (9.7) and we may neglect the Bose-Einstein enhancement or Fermi-Dirac suppression terms $(1 \pm f)$.

$$\begin{aligned} C_{\epsilon,\epsilon'}[f] = (\frac{\partial f}{\partial t})_C \quad &= \quad \frac{1}{2q} \int \frac{d^3s}{2E_s(2\pi)^3} \frac{d^3t}{2E_t(2\pi)^3} \frac{d^3p}{2p(2\pi)^3} (2\pi)^4 \delta^4(s^\mu + q^\mu - t^\mu - p^\mu) \times \\ &\quad |\mathcal{M}|^2 \left(f_e(\vec{t}) f(\vec{p}, \epsilon^{'}) - f_e(\vec{s}) f(\vec{q}, \epsilon) \right) \end{aligned}$$

where $f_e(t)$ is the electron distribution, which may be perturbed, and $f(q, \epsilon)$ is the photon distribution. The integral in d^3t can be done by using the spatial part of the Dirac delta. This fixes $\vec{t} = \vec{s} + \vec{q} - \vec{p}$ below

$$C_{\epsilon,\epsilon'}[f] = \int \frac{d^3s d^3p}{16(2\pi)^5 q p E_s E_t} |\mathcal{M}|^2 \delta(E_s + q - E_t - p) \left(f_e(\vec{t}) f(\vec{p}, \epsilon^{'}) - f_e(\vec{s}) f(\vec{q}, \epsilon) \right) \tag{13.216}$$

We now make a few approximations since we are working in the non-relativistic limit[80]. The phase space factors are $E_{s,t} \simeq m_e$ in the denominator and we pull them out of the integral. The momentum difference between $\vec{t} - \vec{s}$ is sub-dominant, being of order of the momenta of the photons which we have remarked is much smaller than that of the electrons. Therefore we approximate $f_e(\vec{t}) \simeq f_e(\vec{s})$. Next we eye the Dirac delta. In the non-relativistic limit

$$\delta(E_t + p - E_s - q) = \delta(\frac{|\vec{s} + \vec{q} - \vec{p}|^2}{2m_e} - \frac{s^2}{2m_e} + p - q) \tag{13.217}$$

Of all the terms in the argument of the Dirac delta, we may neglect the terms which are quadratic in the momenta of the photon

$$\delta(E_t + p - E_s - q) = \delta(p - q - \frac{\vec{s}}{m_e} \cdot (\vec{p} - \vec{q})) \tag{13.218}$$

The term $\frac{\vec{s}}{m_e} \cdot (\vec{p} - \vec{q})$ is small, which implies that there is not much momentum exchanged in non-relativistic scattering. Because the term is small we may Taylor expand to first order the Dirac delta

$$\delta(E_t + p - E_s - q) = \delta(p - q) - \frac{\vec{s} \cdot (\vec{p} - \vec{q})}{m_e} \frac{\partial}{\partial p} \delta(p - q) \tag{13.219}$$

In case there are doubts about the validity of this, one can verify it works when acting on a test function[13]. With these developments

$$\begin{aligned} C_{\epsilon,\epsilon'}[f] &= \frac{\pi}{k}\frac{1}{4m_e^2}\int \frac{d^3s}{(2\pi)^3}\frac{d^3p}{(2\pi)^3}|\mathcal{M}|^2 f_e(\vec{s})(f(\vec{p},\epsilon^{'}) - f(\vec{q},\epsilon)) \\ &\times \left[\delta(p-q) - \frac{\vec{s}\cdot(\vec{p}-\vec{q})}{m_e}\frac{\partial}{\partial p}\delta(p-q)\right] \end{aligned}$$

The integral on the electron incoming momentum d^3s is now trivial. We note that

$$\int \frac{d^3s}{(2\pi)^3} f_e(s) = n_e + \delta n_e \tag{13.220}$$

and

$$\int \frac{d^3s}{(2\pi)^3} f_e(s)\frac{\vec{s}}{m_e} = n_e\vec{v}_b \tag{13.221}$$

where n_e is the zero-order free electron density, δn_e the perturbation and $\vec{v}_b$ the common velocity of the electron-proton fluid (the "baryons"). We don't count the number of baryons which may be bound in hydrogen atoms as their scattering cross section is much smaller. In doing these integrals we are implicitly assuming that the matrix element $\mathcal{M}$ does not depend on the momentum $\vec{s}$. Indeed, in the non-relativistic limit, it does not.

We separate the photon momentum integral into the magnitude and the angular parts. We may use the Dirac delta to remove the integral on the magnitude. In doing so note that the difference

$$f(q\hat{p},\epsilon^{'}) - f(q\hat{q},\epsilon) = f_0(q)(\Psi(q\hat{p},\epsilon^{'}) - \Psi(q\hat{q},\epsilon)) \tag{13.222}$$

is first order due to the fact that the zero-order distribution is isotropic and unpolarized. The term with the derivative of the Dirac delta is already first order in $\vec{v}_b$. We keep only the zero-order term. These last facts imply that we can neglect the density perturbation of the electrons δn_e as it would be multiplying terms which are already first order.

$$C_{\epsilon,\epsilon'}[f] = \frac{n_e\pi}{8m_e^2}\int \frac{d\Omega_p}{(2\pi)^3}|\mathcal{M}|^2\left[f_0(\Psi(q\hat{p},\epsilon^{'}) - \Psi(q\hat{q},\epsilon)) + \frac{1}{2}q\frac{df_0}{dq}\vec{v}_b\cdot(\hat{p}-\hat{q})\right] \tag{13.223}$$

The factor $\frac{1}{2}$ in the last term comes from our definition of the polarization specific distribution as $f = f_0(\frac{1}{2} + \Psi)$ as in (13.184). Separating the terms which still depend on the direction of $\hat{p}$

$$\begin{aligned} C_{\epsilon,\epsilon'}[f] &= -\frac{n_e\pi}{8m_e^2}(f_0\Psi(\vec{q},\epsilon) + \frac{1}{2}\frac{df_0}{dq}\vec{v}_b\cdot\vec{q})\int\frac{d\Omega_p}{(2\pi)^3}|\mathcal{M}|^2 \\ &+\frac{n_e\pi}{8m_e^2}\int\frac{d\Omega_p}{(2\pi)^3}|\mathcal{M}|^2(f_0\Psi(q\hat{p},\epsilon^{'}) + \vec{v}_b\cdot\hat{p}\frac{df_0}{dq}q) \end{aligned} \tag{13.224}$$

Let's now introduce the polarization structure encoded in the matrix (13.214). The collision term we have calculated is for specific incoming and outgoing linear polarizations. We now obtain the collision term for the total intensity. We must take the sum for two mutually or-

[13] Take $\int dx f(x)\delta(x - x_0 + ax - b)$ for small a, b. This equals $f(\frac{b+x_0}{1+a})\frac{1}{|1+a|}$ which can be Taylor expanded to first order in a, b. The result is the same that is obtained by doing a Taylor expansion of the Dirac Delta first: $\int dx f(x)(\delta(x-x_0) + (ax-b)\delta^{'}(x-x_0))$

thogonal initial state polarizations $\hat{\epsilon}_{1,2}$, as in (13.184). We define the intensity perturbation distribution as $\Psi_I = \Psi(\vec{q}, \hat{e}_\theta) + \Psi(\vec{q}, \hat{e}_\phi)$ where $\hat{e}_i$ are the basis vectors defined above. When summing the various collision terms, we must remember that $\mathcal{M}$ is a function of the polarizations and the outgoing polarization $\hat{\epsilon}^{'}$ is a function of the direction $\hat{p}$ to be integrated upon. Explicitly

$$|\mathcal{M}|^2 = 24\pi m_e^2 \sigma_T |\hat{\epsilon} \cdot \hat{\epsilon}^{'}|^2 \tag{13.225}$$

where $\sigma_T = 6.652 \cdot 10^{-25} cm^2$ is the Thompson cross section.

Let's look at the first term of (13.224) when we sum over ϵ and $\epsilon^{'}$. It is easy to show, for example by explicitly writing out the polarizations $\hat{\epsilon}_{1,2}$, $\hat{\epsilon}^{'}_{1,2}$ in some coordinate system and doing the integral that

$$\sum_{\epsilon^{'}=1}^{2} \int d\Omega_p |\mathcal{M}|^2 = 64\pi m_e^2 \sigma_T \tag{13.226}$$

for a fixed incoming polarization $\hat{\epsilon}$. This result can also be understood by noticing that if we had an unpolarized initial and final state $\sum_{\epsilon,\epsilon^{'}} |\hat{\epsilon} \cdot \hat{\epsilon}^{'}|^2 = (1 + \cos^2\beta)$, where β is the scattering angle. This is of course the unpolarized dependence on the scattering angle. If we only include one initial polarization, we divide by 2. We have already seen this quantity: it was expressed through spherical harmonics of the incoming and outgoing directions in the $I - I$ term of the matrix in(13.214). Since only the Y_0^0 spherical harmonics contribute to the integral, we immediately read off $\int \frac{1}{2}(1 + \cos^2\beta) d\Omega = 8\pi/3$. Of course, the integral can be simply evaluated in polar coordinates, but it is less illuminating. With this in mind the first term, summed over all polarizations, becomes simply

$$-n_e \sigma_T (f_0 \Psi_I + \frac{df_0}{dq} \vec{v}_b \cdot \vec{q}) \tag{13.227}$$

The second term can be expressed via our scattering matrix (13.214). Indeed, in the vector T we can replace $I \to f_0 \Psi_I$, $(Q \pm iU) \to f_0 \Psi_\pm$ where $\Psi_\pm = \Psi_Q \pm i\Psi_U$, $\Psi_Q = \Psi(\vec{q}, \hat{e}_\theta) - \Psi(\vec{q}, \hat{e}_\phi)$, $\Psi_U = \Psi(\vec{q}, \frac{\hat{e}_\theta + \hat{e}_\phi}{\sqrt{2}}) - \Psi(\vec{q}, \frac{\hat{e}_\theta - \hat{e}_\phi}{\sqrt{2}})$. First, we note that the scattering matrix is expressed only in terms of spherical harmonics with $\ell = 0, 2$, while the dot product between a fixed vector $\vec{v}_b$ and the direction $\hat{p}$ is a sum of spherical harmonics with $\ell = 1$. Due to the orthogonality of spherical harmonics, this term does not contribute to the integral. We obtain[14]

$$\begin{aligned} C_I[f] \quad = \quad & -n_e \sigma_T \left(f_0 \Psi(\vec{q}, \epsilon) + \frac{df_0}{dq} \vec{v}_b \cdot \vec{q} \right) \\ & + \frac{n_e \sigma_T}{10} \sum_{m=-2}^{2} \int d\Omega_p [10 \delta_{m0} Y_0^0(\hat{q}) Y_0^{0*}(\hat{p}) f_0 \Psi_I(q, \hat{p}) \\ & + Y_2^m(\hat{q}) Y_2^{m*}(\hat{p}) f_0 \Psi_I(q, \hat{p}) - \sqrt{\frac{3}{2}} Y_2^m(\hat{q})\, {}_2Y_2^{m*}(\hat{p}) f_0 \Psi_+ \\ & - \sqrt{\frac{3}{2}} Y_2^m(\hat{q})\, {}_{-2}Y_2^{m*}(\hat{p}) f_0 \Psi_-] \end{aligned} \tag{13.228}$$

[14] It can be tricky to relate get the overall factor in the second term. To get it right note that $\sum_{\epsilon,\epsilon^{'}} |\hat{\epsilon}_i \cdot \hat{\epsilon}^{'}_j|^2 \Psi(\hat{\epsilon}^{'}_j) = (\Psi(\hat{\epsilon}^{'}_1) + \Psi(\hat{\epsilon}^{'}_2)) \frac{1+\cos^2\beta}{2} + (\Psi(\hat{\epsilon}^{'}_1) - \Psi(\hat{\epsilon}^{'}_2)) \frac{1-\cos^2\beta}{2} = \Psi_I \frac{1+\cos^2\beta}{2} + \Psi_Q \frac{\sin^2\beta}{2}$ so the polarization term alone generates exactly the term in the $I - I$ term in the scattering matrix. Plugging in all the numerical factors leads to the correct expression.

In a similar fashion one can arrive at the collision terms for the $\Psi_{Q,U}$ distributions $C_{Q,U}[f]$ by subtracting the terms with differing initial polarization states, while summing on the final polarization states, ie.

$$C_Q[f] = \sum_{\epsilon'} C_{\epsilon_1,\epsilon'}[f] - C_{\epsilon_2,\epsilon'}[f] \tag{13.229}$$

The terms $C_\pm[f] = C_Q[f] \pm iC_U[f]$ are

$$\begin{aligned} C_\pm[f] &= -n_e\sigma_T f_0 \Psi_\pm(q,\hat{q}) \\ &+\frac{n_e\sigma_T}{10}\sum_{m=-2}^{2} d\Omega_p[-\sqrt{6}\,{}_{\pm2}Y_2^m(\hat{q})Y_2^{m*}(\hat{p})f_0\Psi_I(q,\hat{p}) \\ &+3\,{}_{\pm2}Y_2^m(\hat{q})\,{}_2Y_2^{m*}(\hat{p})f_0\Psi_+(q,\hat{p}) + 3\,{}_{\pm2}Y_2^m(\hat{q})\,{}_{-2}Y_2^{m*}(\hat{p})f_0\Psi_-(q,\hat{p})] \end{aligned} \tag{13.230}$$

13.12 Boltzmann equation for photons

It took us some time but we are now ready to write the full Boltzmann equations for photons. The left hand side is given by the Liouville terms which we have calculated in the conformal-Newtonian (13.149) or synchronous (13.150) gauge, for scalar perturbations, and for tensor perturbations (13.152). We recall that the variables being used are the vector $\vec{k}$ of the Fourier expansion, the magnitude of the momenta of a photon q, its direction $\hat{n}$ and conformal time τ. Also, we had defined $\mu \equiv \hat{k}\cdot\hat{n}$, the angle between the Fourier mode and the photon momentum.

The Liouville term for the total intensity, or energy, of the photon distribution Ψ_I is the one already described. We can think of obtaining it by simply summing over two terms relating the distributions of orthogonal polarizations. On the other hand, the Liouville terms describing the polarizations $\Psi_\pm$ are obtained by taking the difference of the terms relating to orthogonal polarizations $\hat{\epsilon}$. A side effect of this is to remove the metric perturbations from the Boltzmann equation of polarization. We could have expected this, since by no physical mechanism can metric perturbations generate polarization. The Liouville term for polarizations are

$$f_0\left(\frac{\partial\Psi_\pm}{\partial\tau} + ik\mu\frac{q}{E}\Psi_\pm - q\mathcal{H}\frac{\partial\Psi_\pm}{\partial q}\right) \tag{13.231}$$

in both the synchronous and conformal-Newtonian gauges, as well as for tensor modes.

The right hand side of the Boltzmann equations are given by the collision terms (13.228) and (13.230). We note they do not depend on the metric perturbations, a consequence of having used $\vec{q}$ instead of the canonical momenta p_μ. Since we are working in conformal time, we must add multiply the calculated collision terms by a, as it was calculated using coordinate time. The collision terms we have derived are expressed though spherical harmonics which make the azimuthal dependence of each term obvious. Due to integrations with the perturbed distribution $\Psi_{I,\pm}$, these terms will have the effect of projecting out the specific azimuthal dependencies of the perturbed matter distribution. In particular, $m = 0$ relates to scalar modes (azimuthal independence), $m = \pm1$ to vector modes, and $m = \pm2$ to tensor modes. This fact completes our understanding of the scalar-vector-tensor decomposition. We will work for scalar or tensor modes one at the time and ignore vector

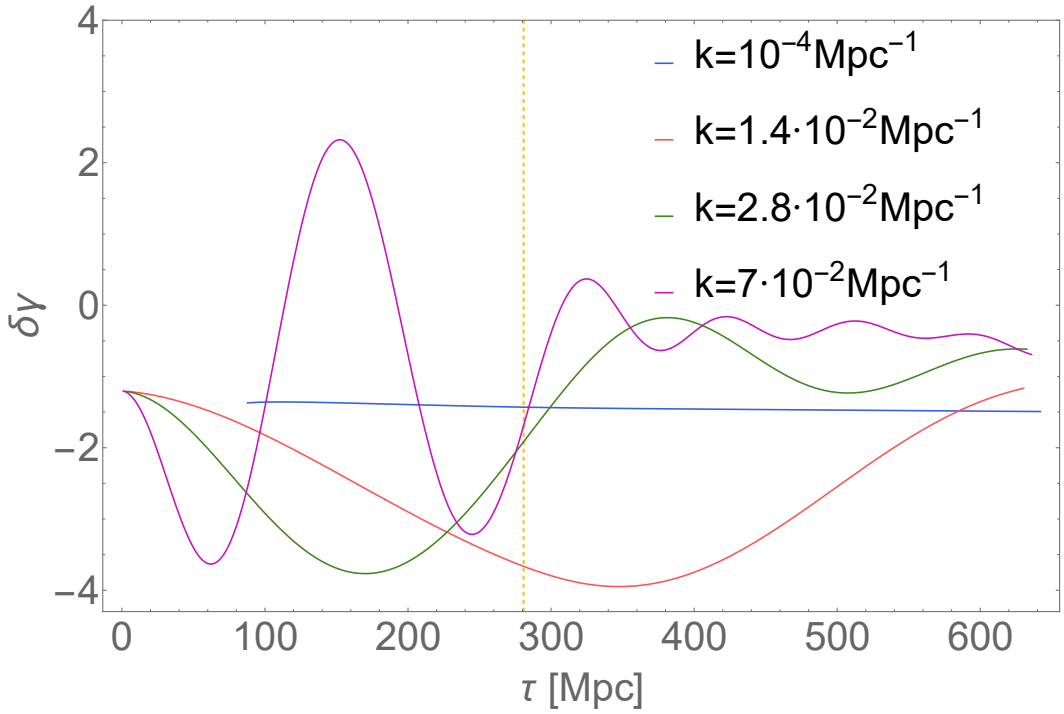

Figure 13.3: Evolution of the photon density δ_ν at early times, for various modes k. The dashed yellow line represents the last scattering surface at $\tau_{LSS} \simeq 280 Mpc$. Anisotropies on the CMB are determined by the value of the photon density at decoupling, plus a metric term which Doppler shifts the photons, since they must escape the gravitational potential well. The oscillatory nature of the solutions is imprinted in the CMB anisotropies. . The calculation was performed with the Boltzmann software CLASS[26] assuming the best fit ΛCDM model[60].

modes, since they do not appear in standard cosmology. However, the equations for vector perturbations are easily obtained using the same formalism which we now present.

As we did for neutrinos in (13.155), we may integrate over the magnitude of the photon momentum q and define[155]

$$F_\gamma(\vec{k}, \hat{n}, \tau) = \frac{\int q^2 dqq \cdot f_0 \Psi_I(q, \hat{n}, \tau, \vec{k})}{\int q^2 dqq \cdot f_0} \tag{13.232}$$

And similarly for the polarizations

$$F_\pm(\vec{k}, \hat{n}, \tau) = \frac{\int q^2 dqq \cdot f_0 \Psi_\pm(q, \hat{n}, \tau, \vec{k})}{\int q^2 dqq \cdot f_0} \tag{13.233}$$

As we had shown in (13.165), F_γ is related to the local fractional density change of the photons. Some texts define the photon perturbation through its local fractional *temperature* change $\Theta_\gamma = \frac{\delta T}{T}$. Because $\rho \propto T^4$, the two quantities are related by $F_\gamma = 4\Theta_\gamma$.

When we integrate the collision terms in this manner, we must pay attention to the term $-n_e \sigma_T \frac{df_0}{dq} \vec{v}_b \cdot \vec{q}$. Of course, we integrate by parts, but care must be taken with the baryon velocity. This velocity can be decomposed into an irrotational term and a vorticity term :

$$\vec{v}_b = \vec{\nabla} v_b + \vec{v}_v \tag{13.234}$$

with $\vec{v}_v$ being divergenceless, so $\vec{v}_v \cdot \vec{k} = 0$. As explained in section 13.7, vorticity is related to vector perturbations. For scalar perturbations, $\vec{v}_b \parallel \vec{k}$ so that $\vec{v}_b \cdot \vec{q} = v_b q\mu$. For tensor perturbations this term does not appear.

Using the formula (13.157) for the derivative of $\dot{F}_{\gamma,\pm}$, we obtain the Boltzmann equations for the total intensity for scalar modes

$$\begin{aligned} &\text{(Synchronous)} \\ \dot{F}_\gamma + ik\mu F_\gamma + \frac{2}{3}\dot{h} + \frac{4}{3}(\dot{h} + 6\dot{\eta})P_2(\mu) &= \\ &\text{(Newtonian)} \\ \dot{F}_\gamma + ik\mu F_\gamma - 4(\dot{\phi} - ik\mu\psi) &= \\ & -an_e\sigma_T F_\gamma + 4an_e\sigma_T v_b\mu + \frac{an_e\sigma_T}{10}\int d\Omega_p \times \\ & [10Y_0^0(\hat{n})Y_0^{0*}(\hat{p})F_\gamma(\hat{p}) + Y_2^0(\hat{n})Y_2^{0*}(\hat{p})F_\gamma(\hat{p}) \\ & -\sqrt{\frac{3}{2}}Y_2^0(\hat{n})\,{}_2Y_2^{0*}(\hat{p})F_+(\hat{p}) \\ & -\sqrt{\frac{3}{2}}Y_2^0(\hat{n})\,{}_{-2}Y_2^{0*}(\hat{p})F_-(\hat{p})] \end{aligned} \tag{13.235}$$

Where we'd chooses the left hand side according to the gauge we're working in (Synchronous or conformal-Newtonian), the right hand side being the same. For tensor modes

$$\begin{aligned} &\text{(Tensor)} \\ \dot{F}_\gamma + ik\mu F_\gamma + \frac{1-\mu^2}{2}e^{2i\phi}(h_+ - ih_\times) & \\ +\frac{1-\mu^2}{2}e^{-2i\phi}(h_+ + ih_\times) &= \\ & -an_e\sigma_T F_\gamma + \frac{an_e\sigma_T}{10}\sum_{m=\{-2,2\}}\int d\Omega_p \\ & [Y_2^m(\hat{n})Y_2^{m*}(\hat{p})F_\gamma(\hat{p}) \\ & -\sqrt{\frac{3}{2}}Y_2^m(\hat{n})\,{}_2Y_2^{m*}(\hat{p})F_+(\hat{p}) \\ & -\sqrt{\frac{3}{2}}Y_2^m(\hat{n})\,{}_{-2}Y_2^{m*}(\hat{p})F_-(\hat{p})] \end{aligned} \tag{13.236}$$

where the sum on the right extends *only* to the values $m = \pm 2$. We have made explicit the terms with definite azimuthal dependence on the left hand side since shortly we will be turning them into spherical harmonics.

The Boltzmann equation for polarization of scalar modes is

$$\begin{aligned} \text{(Scalar)}\dot{F}_\pm + ik\mu F_\pm &= \\ & -an_e\sigma_T F_\pm + \frac{an_e\sigma_T}{10}\int d\Omega_p \times \\ & [-\sqrt{6}\,{}_{\pm2}Y_2^0(\hat{n})Y_2^{0*}(\hat{p})F_\gamma(\hat{p}) \\ & +3\,{}_{\pm2}Y_2^0(\hat{n})\,{}_2Y_2^{0*}(\hat{p})F_+(\hat{p}) \\ & +3\,{}_{\pm2}Y_2^0(\hat{n})\,{}_{-2}Y_2^0(\hat{p})F_-(\hat{p})] \end{aligned} \tag{13.237}$$

It does not depend explicitly on the choice of gauge, since the metric perturbations have dropped out. For tensors

$$\begin{aligned}(\text{Tensor})\dot{F}_\pm + ik\mu F_\pm \quad = \quad & \\ & -an_e\sigma_T F_\pm + \frac{an_e\sigma_T}{10}\sum_{m=\{-2,2\}}\int d\Omega_p \times \\ & [-\sqrt{6}\,{}_{\pm2}Y_2^m(\hat{n})Y_2^{m*}(\hat{p})F_\gamma(\hat{p}) \\ & +3\,{}_{\pm2}Y_2^m(\hat{n})\,{}_2Y_2^{m*}(\hat{p})F_+(\hat{p}) \\ & +3\,{}_{\pm2}Y_2^m(\hat{n})\,{}_{-2}Y_2^m(\hat{p})F_-(\hat{p})]\end{aligned} \tag{13.238}$$

Next we expand the $F_{\gamma,\pm}$ in (spin-weighted) spherical harmonics

$$F_\gamma(\vec{k},\hat{n},\tau) = \sum_{\ell,m}\sqrt{2\ell+1}\sqrt{4\pi}(-i)^\ell F_{\gamma\ell}^{(m)}(\vec{k},\tau)Y_\ell^m(\hat{n}) \tag{13.239}$$

For polarization

$$F_\pm(\vec{k},\hat{n},\tau) = \sum_{\ell,m}\sqrt{2\ell+1}\sqrt{4\pi}(-i)^\ell(E_\ell^{(m)} \pm iB_\ell^{(m)})\,{}_{\pm2}Y_\ell^m(\hat{n}) \tag{13.240}$$

Since $(Q \pm iU)$ are spin-2 (-2) quantities it is more natural to decompose them into spin-weighted spherical harmonics. We also separate out the two E and B modes of polarization. As we had shown in (13.199) and (13.200), these quantities have definite behavior under parity transformations. Due to the realness of $F_\pm$, $E_\ell^{(-m)} = E_\ell^{(m)}$ and $B_\ell^{(-m)} = -B_\ell^{(m)}$. The normalization chosen is, of course, arbitrary. With this choice, since $Y_\ell^0(\theta,\phi) = \sqrt{\frac{2\ell+1}{4\pi}}P_\ell(\cos\theta)$, we obtain the same normalization as we had when expanding the neutrinos (13.163) in Legendre polynomials. Thus (13.165), (13.166), (13.167) hold the same for photons:

$$\delta_\gamma = F_{\gamma0}^{(0)} \tag{13.241}$$

$$\theta_\gamma = \frac{3}{4}F_{\gamma1}^{(0)} \tag{13.242}$$

$$\sigma_\gamma = \frac{1}{2}F_{\gamma2}^{(0)} \tag{13.243}$$

Integrating over $F_{\gamma,\pm}$, by weighing with a (spin-weighted) spherical harmonic, projects onto the relevant component

$$\int d\Omega_n Y_\ell^{m*}(\hat{n})F_\gamma(\vec{k},\hat{n},\tau) = (-i)^\ell\sqrt{4\pi(2\ell+1)}F_{\gamma\ell}^{(m)}(\vec{k},\tau) \tag{13.244}$$

$$\int d\Omega_n\,{}_{\pm2}Y_\ell^{m*}(\hat{n})F_\pm(\vec{k},\hat{n},\tau) = (-i)^\ell\sqrt{4\pi(2\ell+1)}F_{\pm\ell}^{(m)}(\vec{k},\tau) \tag{13.245}$$

We will now take the above Boltzmann equations for intensity (13.235), (13.236) and integrate them over $\int d\Omega_n Y_\ell^{m*}(\hat{n})$. We take $m = 0$ for the scalar modes and $m = \pm2$ for the

tensors, and increasing values of $\ell \geq 0$. This projects out all the possible angular dependencies.

In order to do the projections systematically it is useful to write all the spherical harmonics in the above equations explicitly

$$\mu = \sqrt{\frac{4\pi}{3}} Y_1^0(\hat{n}) \tag{13.246}$$

$$P_2(\mu) = \sqrt{\frac{4\pi}{5}} Y_2^0(\hat{n}) \tag{13.247}$$

$$\frac{1-\mu^2}{2} e^{\pm 2i\phi} = \sqrt{\frac{8\pi}{15}} Y_2^{\pm 2}(\hat{n}) \tag{13.248}$$

The outer product $Y_1^0 \, {}_sY_\ell^m$ appears and is expressed in terms of Clebsch-Gordan coefficients (see appendix B) as

$$\begin{aligned} \sqrt{\frac{4\pi}{3}} Y_1^0 \, {}_sY_\ell^m &= \frac{\sqrt{(\ell^2-m^2)(1-\frac{s^2}{\ell^2})}}{\sqrt{(2\ell+1)(2\ell-1)}} \, {}_sY_{\ell-1}^m \\ &\quad - \frac{ms}{\ell(\ell+1)} \, {}_sY_\ell^m + \frac{\sqrt{((\ell+1)^2-m^2)(1-\frac{s^2}{(\ell+1)^2})}}{\sqrt{(2\ell+1)(2\ell+3)}} \, {}_sY_{\ell+1}^m \end{aligned}$$

This allows us to write out the expansion μF_γ as

$$\begin{aligned} \mu F_\gamma &= -i\sqrt{4\pi} F_{\gamma 1}^{(0)} Y_0^0 + \sum_{\ell=1,m} (-i)^\ell \sqrt{\frac{4\pi}{2\ell+1}} Y_\ell^m \times \\ &\quad \left[-i\sqrt{(\ell+1)^2-m^2} F_{\gamma,\ell+1}^{(m)} + i\sqrt{\ell^2-m^2} F_{\gamma,\ell-1}^{(m)} \right] \end{aligned}$$

Because it will be of use in the equations, we define the polarization source

$$\Pi^{(m)} = \frac{1}{10}(F_{\gamma 2}^{(m)} - \sqrt{6} E_2^{(m)}) \tag{13.249}$$

Then we obtain for scalar modes in the *synchronous gauge* [15]

$$\textit{(Synchronous)}\dot{\delta}_\gamma = -\frac{4}{3}\theta_\gamma - \frac{2}{3}\dot{h} \tag{13.250}$$

$$\text{(Synchronous)}\dot{\theta}_\gamma = \frac{k^2}{4}\delta_\gamma - \frac{k^2}{2} F_{\gamma 2}^{(0)} - a n_e \sigma_T (\theta_\gamma - \theta_b) \tag{13.251}$$

$$\text{(Synchronous)}\dot{F}_{\gamma 2}^{(0)} = \frac{8}{15}\theta_\gamma - \frac{3k}{5} F_{\gamma 3}^{(0)} + \frac{4}{15}(\dot{h} + 6\dot{\eta}) - a n_e \sigma_T F_{\gamma 2}^{(0)} + a n_e \sigma_T \Pi^{(0)} \tag{13.252}$$

[15] Recall $\theta_b = ikv_b$ since $\vec{k} \parallel \vec{v}_b$ for scalar modes.

For scalar modes in the *conformal-Newtonian gauge*

$$\text{(Newtonian)}\dot{\delta}_\gamma = -\frac{4}{3}\theta_\gamma + 4\dot{\phi} \tag{13.253}$$

$$\text{(Newtonian)}\dot{\theta}_\gamma = \frac{k^2}{4}\delta_\gamma - \frac{k^2}{2}F^{(0)}_{\gamma 2} + k^2\psi - an_e\sigma_T(\theta_\gamma - \theta_b) \tag{13.254}$$

$$\text{(Newtonian)}\dot{F}^{(0)}_{\gamma 2} = \frac{8}{15}\theta_\gamma - \frac{3k}{5}F^{(0)}_{\gamma 3} - an_e\sigma_T F^{(0)}_{\gamma 2} + an_e\sigma_T\Pi^{(0)} \tag{13.255}$$

For $\ell \geq 3$ scalar modes, in either gauge,

$$\dot{F}^{(0)}_{\gamma\ell} = k\frac{\ell}{2\ell+1}F^{(0)}_{\gamma(\ell-1)} - k\frac{\ell+1}{2\ell+1}F^{(0)}_{\gamma(\ell+1)} - an_e\sigma_T F^{(0)}_{\gamma\ell} \;\; \ell \geq 3 \tag{13.256}$$

The polarization equations are the same in either gauge.

$$\dot{E}^{(0)}_2 = -\frac{k}{\sqrt{5}}E^{(0)}_3 - an_e\sigma_T E^{(0)}_2 - \sqrt{6}an_e\sigma_T\Pi^{(0)} \tag{13.257}$$

$$\dot{B}^{(0)}_2 = -\frac{k}{\sqrt{5}}B^{(0)}_3 - an_e\sigma_T B^{(0)}_2 \tag{13.258}$$

And for $\ell \geq 3$

$$\dot{E}^{(0)}_\ell = k\frac{\sqrt{\ell(1-\frac{4}{\ell^2})}}{2\ell+1}E^{(0)}_{\ell-1} - k\frac{\sqrt{(\ell+1)^2(1-\frac{4}{(\ell+1)^2})}}{2\ell+1}E^{(0)}_{\ell+1} - an_e\sigma_T E^{(0)}_\ell \;\; \ell \geq 3 \tag{13.259}$$

$$\dot{B}^0_\ell = k\frac{\sqrt{(\ell(1-\frac{4}{\ell^2})}}{2\ell+1}B^{(0)}_{\ell-1} - k\frac{\sqrt{(\ell+1)^2(1-\frac{4}{(\ell+1)^2})}}{2\ell+1}B^{(0)}_{\ell+1} - an_e\sigma_T B^{(0)}_\ell \;\; \ell \geq 3 \tag{13.260}$$

For scalar modes, the B polarization has completely decoupled from the other perturbations.

For tensor modes the intensity perturbations satisfy

$$\dot{F}^{(\pm 2)}_{\gamma 2} = -kF^{(\pm 2)}_{\gamma 3} + \frac{1}{5}\sqrt{\frac{2}{3}}(h_+ \mp ih_\times) - an_e\sigma_T F^{(\pm 2)}_{\gamma 2} + an_e\sigma_T P^{(2)} \tag{13.261}$$

$$\dot{F}^{(\pm 2)}_{\gamma\ell} = k\frac{\sqrt{\ell^2-4}}{2\ell+1}F^{(\pm 2)}_{\gamma(\ell-1)} - k\frac{\sqrt{(\ell+1)^2-4}}{2\ell+1}F^{(\pm 2)}_{\gamma(\ell+1)} - an_e\sigma_T F^{(\pm 2)}_{\gamma\ell} \;\; \ell \geq 3 \tag{13.262}$$

Note that $F^{(\pm 2)}_{\gamma\ell}$ is a complex function, while $h_{+,\times}$and $E^{(2)}_2$ are not. So each of these equations is actually two equations which one can separate into a real and imaginary part. The equations for $m = \pm 2$ are not independent, so there are only two equations per ℓ.

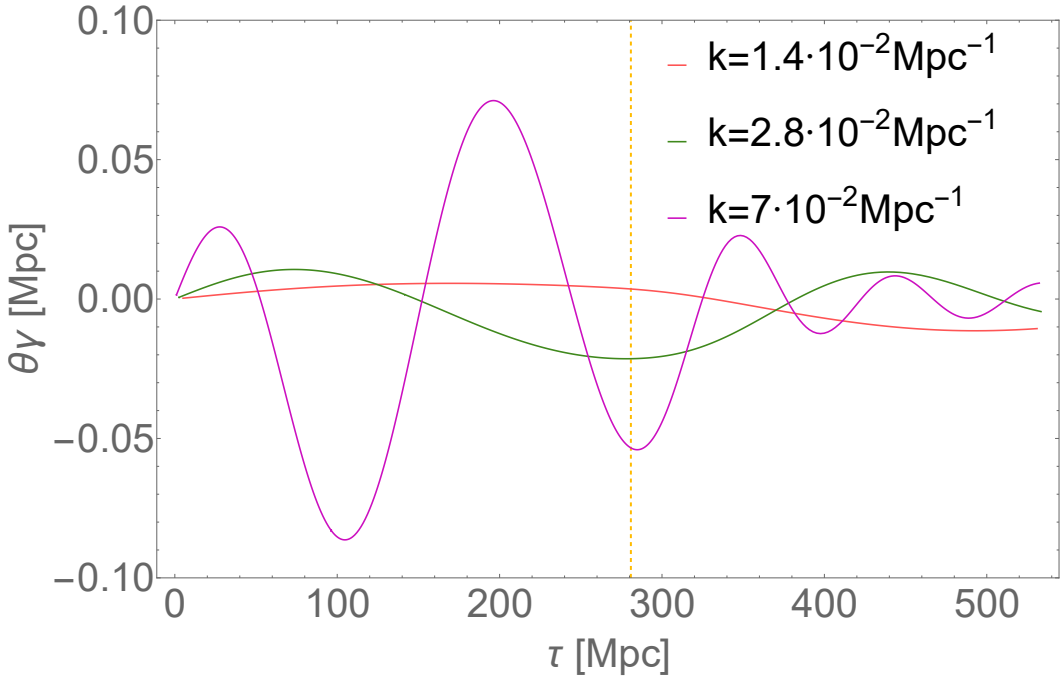

Figure 13.4: Evolution of the velocity (dipole) term of the photon distribution at early times for several ks. The dashed yellow line is the last scattering time. The oscillatory nature of the solution, via the different values at decoupling with changing k, will be impressed in the form of the CMB anisotropies. The calculation is performed with the Boltzmann code CLASS[26] using the best fit ΛCDM parameters[60].

Finally we have the polarization for tensors, for $m = 2$,

$$\dot{E}_2^{(m)} = -k\frac{\sqrt{\frac{5}{9}(9-m^2)}}{5}E_3^{(m)} - \frac{km}{3}B_2^{(m)} - an_e\sigma_T E_2^{(m)} - \sqrt{6}an_e\sigma_T\Pi^{(m)} \tag{13.263}$$

$$\begin{aligned}\dot{E}_\ell^{(m)} &= k\frac{\sqrt{(\ell^2-m^2)(1-\frac{4}{\ell^2})}}{2\ell+1}E_{\ell-1}^{(m)} - \frac{2km}{\ell(\ell+1)}B_\ell^{(m)} - an_e\sigma_T E_\ell^{(m)} \\ &\quad -k\frac{\sqrt{((\ell+1)^2-m^2)(1-\frac{4}{(\ell+1)^2})}}{2\ell+1}E_{\ell+1}^{(m)}\end{aligned} \tag{13.264}$$

For tensors $m = 2$ in the above, but we kept the m explicit because the same formula is valid for vector perturbations. Furthermore, keeping the m explicit shows that for $m \neq 0$ E and B are coupled.

We have derived the complete set of Boltzmann equations for photon scalar and tensor modes, including their polarization. As with neutrinos, we have found an infinite hierarchy of equations, parametrized by the quantum number ℓ. In a practical calculation this would need to be cutoff. Unlike neutrinos, these equations can't be solved on their own. We need two ingredients to go along with them. One is the evolution of the number density of free electrons n_e, which as we discussed in section 11 is a quantity that must be determined numerically as well during recombination. Furthermore, we need to write the perturbation equations for baryons as they appear in their velocity parameter θ_b.

An interesting point to be made is that the B polarization modes never appear as a source,

nor are they sourced, in a term proportional to $an_e\sigma_T$, to the other perturbations. They only couple through the Clebsch-Gordan coefficients linking one angular component to the next. In particular, scalar B modes are completely decoupled from the rest of the perturbations. If there were any initial scalar B modes, they would rapidly decay due to the terms $-an_e\sigma_T B_\ell^{(0)}$ present in all equations, and due the fact that there is no E or F_γ mode which can excite them. So far, only primordial scalar modes have been detected in cosmological observables, this conclusion is linked precisely to the fact that B modes are *not observed in the cosmic microwave background*. If they exist, B modes would signal the presence of tensor or vector modes. Currently, there is a strong belief that primordial inflation excites tensor modes, gravitational waves, which would leave an imprint in the CMB in the form of B modes. The detection of B modes is therefore an interesting avenue to pursue new physics.

13.13 Baryon Boltzmann equations

The final component of the universe which we must discuss are the "baryons". In this cosmological context we mean electrons and protons. Due to the relatively large cross section $e^- + p \leftrightarrow e^- + p$ the electrons and protons can be thought as forming a single fluid. As we had pointed out, scattering between electrons and photons is much larger than that between photons and protons. There will be a large energy exchange between electrons and photons. This energy exchange will be quickly thermalized in the baryon fluid so that, indirectly, the photons exchange energy just as quickly with protons as they do with electrons.

To derive the Boltzmann equations for baryons, we could use the Liouville terms we've already calculated, write the collision terms for the various scatterings and integrate. There is a much faster route which we will employ here.

We can use the continuity equations (13.81)-(13.84). Those equations, which descend from the Einstein equations and the Bianchi identities to make the energy-momentum tensor divergenceless $T^{\mu\nu}_{\ ;\nu} = 0$, were derived for either the sum of all components of the universe or a single component which interacts only gravitationally. We had used this to derive the Boltzmann equations for dark matter 13.95-(13.98). For baryons this is not suitable, since the interaction with photons is non-negligible, however conservation of momentum comes to our aide.

Let's consider the continuity equations for the baryon fluid. It is a non-relativistic fluid and $P = w\rho \ll \rho$ so that we can neglect the pressure and the anisotropic stress σ. We shall however keep a term proportional to the sound speed $c_s^2 = \frac{\delta P}{\delta \rho} \simeq w$ in the equation for the velocity θ_b, to which we will return. Thus the continuity equations are approximated by

$$\dot{\delta}_b + \theta_b - 3\dot{\phi} + \frac{\dot{h}}{2} = (\frac{\partial \delta_b}{\partial \tau})_C \tag{13.265}$$

$$\dot{\theta}_b + \mathcal{H}\theta_b - k^2\psi - k^2 c_s^2 \delta_b = (\frac{\partial \theta_b}{\partial \tau})_C \tag{13.266}$$

We have abused notation by mixing the synchronous and conformal-Newtonian gauges. Obviously, ψ and ϕ terms only appear when using the conformal-Newtonian gauge, while h appears only when using the synchronous gauge. We have also introduced the collision

terms which are due to energy exchange, through δ_b, and momentum exchange, through θ_b, with the photons. In principle, we may consider higher moments for the baryon distribution but, as with dark matter, these terms are negligible for a non-relativistic species.

Looking at the equations for the perturbations of the photons , (13.250) and the following, we see that the photons don't exchange energy with the baryons via the density. There is no baryon related term δ_b in the equation for δ_γ. Thus, there is no collision term $(\frac{\partial \delta_b}{\partial \tau})_C$ and the equation for $\dot{\delta}_b$ is the collisionless one. On the other hand, momentum is exchanged through θ_b as can be seen in equations (13.251) and (13.254). The fact that baryons don't appear in equations for higher moments ($\ell \geq 2$) of the photons is due to the fact they are non-relativistic, which we had used extensively in deriving the collision term for $e^- + \gamma \leftrightarrow e^- + \gamma$ scattering. Higher moments of the proton distribution are negligible.

The correct form of the collision term $(\frac{\partial \theta_b}{\partial \tau})_C$ can be understood by recalling that, by definition (13.76),

$$(\rho + P)\theta = ik^i \delta T^0_i \tag{13.267}$$

so conservation of momentum implies

$$(\rho_b + P_b)(\frac{\partial \theta_b}{\partial \tau})_C + (\rho_\gamma + P_\gamma)(\frac{\partial \theta_\gamma}{\partial \tau})_C = 0 \tag{13.268}$$

Using $P_b = 0$ and $P_\gamma = \frac{\rho_\gamma}{3}$, together with the fact that

$$(\frac{\partial \theta_\gamma}{\partial \tau})_C = -an_e\sigma_T(\theta_\gamma - \theta_b) \tag{13.269}$$

we may now write the equations for the scalar modes of baryon perturbations

$$\text{(Synchronous)}\dot{\delta}_b = -\theta_b - \frac{1}{2}\dot{h} \tag{13.270}$$

$$\text{(Synchronous)}\dot{\theta}_b = -\mathcal{H}\theta_b + k^2 c_s^2 \delta_b + \frac{4}{3}\frac{\rho_\gamma}{\rho_b} an_e\sigma_T(\theta_\gamma - \theta_b) \tag{13.271}$$

$$\text{(Newtonian)}\dot{\delta}_b = -\theta_b + 3\dot{\phi} \tag{13.272}$$

$$\text{(Newtonian)}\dot{\theta}_b = -\mathcal{H}\theta_b + k^2 c_s^2 \delta_b + k^2\psi + \frac{4}{3}\frac{\rho_\gamma}{\rho_b} an_e\sigma_T(\theta_\gamma - \theta_b) \tag{13.273}$$

What happens during recombination? Electrons and protons bind together, forming hydrogen atoms, and we may think these equations lose their validity. Actually they do not. In fact, neutral atoms still scatter off free electrons and protons very quickly, so that electrons, protons and neutral atoms can be still thought to form one fluid. In particular, the factor ρ_b in the denominator, having come from conservation of momentum, is still the total amount of baryons, including those bound in atoms. What *is* important to consider during recombination is the amount of free electrons n_e, an amount that must be calculated using the equations for recombination.

The sound speed c_s^2 is defined through $c_s^2 = \frac{\partial P}{\partial \rho} = \frac{\dot{P}}{\dot{\rho}}$. Since it is proportional to $w \simeq 0$ for matter, the term with the sound speed can be safely neglected unless $k \simeq c_s^{-1}$, which

turns out to be quite large for practical needs. Nevertheless, we take a look at it. Assuming constant $w \simeq 0$ in the equation of state $P = w\rho$, we can use the non-relativistic formulas for P and ρ given in table 8.1. $\rho_b \simeq \mu_b n_b$, $P_b = n_b T_b$ where

$$\mu_b = \sum_i m_i \frac{n_i}{n_b} \tag{13.274}$$

is the mean molecular weight of the baryon fluid. The sum is taken over baryon species i (including electrons and atoms) with their number fractions n_i/n_b, such that $\sum_i n_i = n_b$. With these definitions the sound speed is

$$c_s^2 = \frac{\dot{n}_b T_b + \dot{T}_b n_b}{\dot{\mu}_b n_b + \dot{n}_b \mu_b} \tag{13.275}$$

We neglect the change of the mean molecular weight over time $\dot{\mu}_b$ since c_s^2, a term originating from matter pressure, is already small and $\dot{\mu}_b$ represents a small correction to the denominator even during recombination, when its change is fastest. Using the fact $n_b \propto a^{-3}$

$$c_s^2 = \frac{T_b}{\mu_b}(1 - \frac{1}{3}\frac{d\ln T_b}{d\ln a}) \tag{13.276}$$

The baryon temperature T_b will be equal to the photon temperature so long as the scattering rate is fast enough. During recombination this scattering becomes inefficient and $T_b \neq T_\gamma$. Using the first law of thermodynamics in a comoving volume $V = a^3$

$$\partial Q = d(\rho a^3) + P da^3 \tag{13.277}$$

Together with the non-relativistic thermodynamic quantities P, ρ , we get

$$\dot{T}_b = -2\mathcal{H}T + \frac{2}{3}\frac{1}{N_b}\frac{\partial Q}{\partial \tau} \tag{13.278}$$

where N_b is the number of baryons and $\frac{\partial Q}{\partial \tau}$ is the heat in exchange per unit time in the volume. This heat exchange can be calculated with the Boltzmann equation in the usual way. By considering Compton scattering on the electrons one finds

$$\dot{T}_b = -2\mathcal{H}T + \frac{8}{3}\frac{\mu_b}{m_e}\frac{\rho_\gamma}{\rho_b} a n_e \sigma_T (T_\gamma - T_b) \tag{13.279}$$

At this point we have derived all the differential equations needed to solve, in principle, for the small perturbations around a homogeneous universe. The next steps will be understanding the meaning of these equations and relating them to actual experimental observables.

13.14 Tight coupling approximation

We have found the photon intensity equations for scalar modes in the synchronous gauge (equation (13.250) and the following) and in the conformal-Newtonian gauge (equation (13.253) and the following). We have put them in a form of an infinite hierarchy parametrized

by the moment ℓ, every $F^{(0)}_{\gamma\ell}$ depends on the successive moment $F^{(0)}_{\gamma(\ell+1)}$. In a numerical calculation, to determine the form of the anisotropies in the CMB, we would actually need to know the values of $F^{(0)}_{\gamma\ell}$ for up to $\ell \sim 10^3$, which implies a very large number of differential equations to track, increasing the computational time. Fortunately, there are a few techniques to reduce this complexity. We will now be discussing the tight coupling approximation. A more precise approximation will be used in section 17.2, here we focus on understanding the phenomenon of *acoustic oscillations.*

We introduce the *optical depth[80]*

$$\tau_c \equiv \int_{\tau}^{\tau_0} d\tau^{'} a n_e \sigma_T \tag{13.280}$$

where τ_0 is the conformal time today. By the limits of integration used,

$$\dot{\tau}_c = -a n_e \sigma_T \tag{13.281}$$

The optical depth appears in every Boltzmann equation which contains Thompson scattering and is a measure of how far back in the universe we may see freely. Today, since the free electron density is very small, $\tau_c \ll 1$, while at early times it is very large. The idea of the *tight coupling limit* is that at early enough times $\dot{\tau}_c$, being proportional to the scattering rate, is much larger than the Hubble factor $\tau_c \gg \mathcal{H}$, which is $\mathcal{H} \sim \tau^{-1}$ by the Friedmann equation. This means that the baryon and photon fluid are very tightly coupled and, in some limit, may be thought of as a single fluid. Although they will eventually decouple, understanding the motion of the single fluid approximation gives a good picture of what is going on.

Because the baryons are non-relativistic, we have neglected their anisotropic stresses(and higher moments $\ell \geq 2$), since these are proportional to at least $(\frac{v}{c})^2$, v being the average velocity. For a massless species such as photons this is not possible. However, if the baryons and photons interact quickly enough with one another, and can be treated as single fluid, the moments of the two species should become the same. This is evident in equations (13.271) and (13.251), where the Compton scattering term is proportional to $\theta_\gamma - \theta_b$. If $|\dot{\tau}_c| \gg 1$ then this term dominates and $\theta_\gamma - \theta_b \to 0$. In a similar fashion, we expect all the $\ell \geq 2$ moments $F^{(0)}_{\gamma\ell}$ to go to zero. This has an ulterior implication.

Polarization of the photons is generated by the quadrupole $F^{(0)}_{\gamma 2}$. While the photons and baryons are tightly coupled we expect no polarization to be generated. Since recombination and decoupling happens fairly quickly, this implies the polarization of the cosmic microwave background will be very small. After decoupling in fact there is no Compton scattering which can generate it.

Getting into the details, we find that the photon intensity equations for $\ell \geq 3$ *(13.256)* reduce to the form

$$\dot{F}^{(0)}_{\gamma\ell} = k\frac{\ell}{2\ell+1}F^{(0)}_{\gamma(\ell-1)} - k\frac{\ell+1}{2\ell+1}F^{(0)}_{\gamma(\ell+1)} + \dot{\tau}_c F^{(0)}_{\gamma\ell} \tag{13.282}$$

Indeed, even for $\ell < 3$ the equations have these terms, in addition to others coupling the

photons to the metric and the baryons, as well as their polarization. Understanding the $\ell \geq 3$ equations therefore is a good start to understand the $\ell < 3$ equations. Notice that $\dot{\tau}_c$ is negative and we look at the limit where it is very large in absolute value. If the terms $F^{(0)}_{\gamma(\ell\pm1)}$ were not present, the moment would be very quickly exponentially damped. The derivative can be thought approximately as $\dot{F}^{(0)}_{\gamma\ell} \sim F^{(0)}_{\gamma\ell}/\tau$ while the right hand side term $\dot{\tau}_c F^{(0)}_{\gamma\ell} \sim \tau_c F^{(0)}_{\gamma\ell}/\tau$. Since τ_c is very large, we may neglect the derivative and the equation reduces to an algebraic one. Forgetting the $F^{(0)}_{\gamma(\ell+1)}$ for a moment we see that

$$F^{(0)}_{\gamma\ell} \sim \frac{k\tau}{\tau_c} F^{(0)}_{\gamma(\ell-1)} \tag{13.283}$$

So long as $k\tau < \tau_c$, we find that higher moments of the photon distribution are suppressed by increasing powers of $\tau_c/\tau \simeq \dot{\tau}$. This justifies our neglecting $F^{(0)}_{\gamma(\ell+1)}$ a moment ago. In the tight coupling limit we will now neglect all photon moments with $\ell \geq 2$, keeping only δ_γ and θ_γ. Before we proceed with a few calculations, we note that a more precise approximation would track the anisotropic stress of the photons $\sigma_\gamma = \frac{1}{2}F^{(0)}_{\gamma 2}$ as well.

We take the equation for $\dot{\theta}_b$ in either the synchronous (13.271) or conformal-Newtonian (13.273) gauge. As a shorthand, we define[16]

$$R = \frac{4}{3}\frac{\rho_\gamma}{\rho_b} \tag{13.284}$$

The baryon equations can be written as

$$-R\dot{\tau}_c(\theta_\gamma - \theta_b) = \dot{\theta}_b + \mathcal{H}\theta_b - k^2\psi \tag{13.285}$$

where we neglected the sound speed term $k^2 c_s^2 \delta_b$, since it is very small for matter and k not large. Obviously, the above equation is valid in the conformal-Newtonian gauge, for the equation in the synchronous gauge, the $k^2\psi$ term is absent. Now we notice that $\theta_b = \theta_\gamma + o(\dot{\tau}_c^{-1})$. We want to neglect terms $o(\dot{\tau}_c^{-2})$ which means we can substitute $\theta_b = \theta_\gamma$ in the left-hand side above. Rearranging the terms

$$\theta_b \simeq \theta_\gamma + \frac{1}{R\dot{\tau}_c}(\dot{\theta}_\gamma + \mathcal{H}\theta_\gamma - k^2\psi) \tag{13.286}$$

Next, we take the equations for θ_γ in either the synchronous (13.251) or conformal-Newtonian (13.254) gauge. We neglect the higher photon moment $F^{(0)}_{\gamma 2}$ and substitute the above value for θ_b. After rearranging the terms

$$\dot{\theta}_\gamma = \frac{R}{1+R}\frac{k^2}{4}\delta_\gamma - \frac{1}{1+R}\mathcal{H}\theta_\gamma + k^2\psi \tag{13.287}$$

in the conformal-Newtonian gauge. The equations is the same in the synchronous gauge once the $k^2\psi$ term is removed. This equation, together with one for δ_γ, form a set of two equations for the photon fluid which depend only on the metric. The baryon perturbations can then be determined algebraically using (13.286).

A more intuitive equation can be obtained combining the two first order photon equa-

[16] Care must be taken when comparing across the literature. Some sources define the same quantity as R^{-1}.

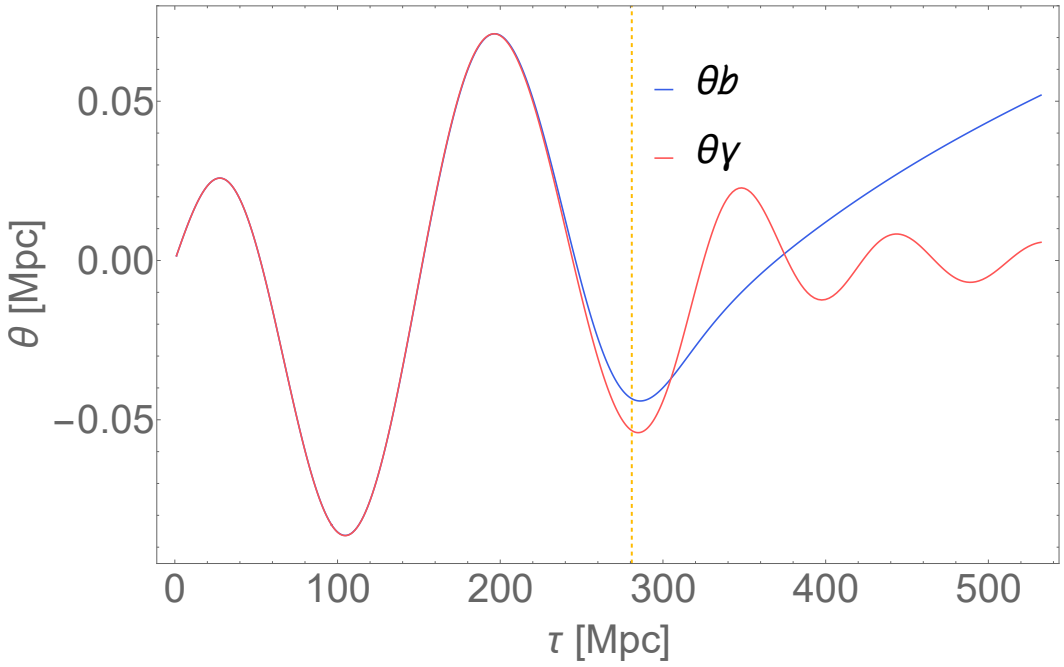

Figure 13.5: Numerical solution of θ_b and θ_γ for $k = 0.7 \cdot 10^{-2} Mpc^{-1}$. The dashed yellow line is the last scattering time, or decoupling. At early enough times the evolution is indistinguishable. The tight coupling approximation is justified (and indeed was used in this calculation as well for small τ). As the free electron density decreases, so does $\dot{\tau}_c$, and the two solutions diverge. The calculation is performed with the CLASS[26] software using the best fit ΛCDM parameters[60].

tions into a second order one. Differentiating the equation for δ_γ in either the synchronous (13.250) or conformal-Newtonian (13.253) gauge, and substituting the approximated $\dot{\theta}_\gamma$, we obtain the equation, of the same form in both gauges,

$$\ddot{\delta}_\gamma + \frac{\mathcal{H}}{1+R}\dot{\delta}_\gamma + \frac{R}{1+R}\frac{k^2}{3}\delta_\gamma = F(k,\tau) \tag{13.288}$$

where we have defined the *forcing function*

$$(Synchronous) F(k,\tau) = -\frac{2}{3}\frac{\mathcal{H}}{1+R}\dot{h} - \frac{2}{3}\ddot{h} \tag{13.289}$$

$$(Newtonian) F(k,\tau) = -\frac{4}{3}k^2\psi + 4\frac{\mathcal{H}}{1+R}\dot{\phi} + 4\ddot{\phi} \tag{13.290}$$

We define the *speed of sound of the coupled fluid*

$$c_s^2 = \frac{1}{3}\frac{R}{1+R} \tag{13.291}$$

The photon-baryon coupled fluid acts as a single fluid which satisfies a damped and forced oscillating equation with wave number $\omega = kc_s$. The damping term is proportional to the Hubble factor $\mathcal{H}$. This is common to find in oscillatory equations in an expanding universe and is known as *Hubble friction*. A quickly expanding universe damps or freezes

oscillations.

What we have found is the phenomena of *acoustic oscillations*. Sound waves propagate through the early universe and the density of photons and baryons is compressed and decompressed in an oscillatory manner. This will have a striking consequence on the spectrum of CMB fluctuations. Indeed, the oscillatory nature of the fluid will leave an oscillatory imprint on the CMB today. Roughly speaking, the oscillations continue until decoupling when they freeze out. Thus the pattern of over and under-densities of the *CMB will follow that of the state of oscillations at decoupling.*

We can extract an approximate solution from (13.288). We notice that in the conformal-Newtonian gauge the derivatives of 4ϕ appear in the same way as those of δ_γ. In the synchronous gauge the same can be said for $\frac{2}{3}h$. This allows us to recast the equation into the form

$$(\frac{d^2}{d\tau^2} + \frac{\mathcal{H}}{1+R}\frac{d}{d\tau} + k^2c_s^2)(\delta_\gamma - 4\phi) = -4k^2c_s^2\phi - \frac{4}{3}k^2\psi \tag{13.292}$$

and similarly for the synchronous gauge. We will now derive the approximate solution in the conformal-Newtonian gauge, neglecting the damping term $\frac{\mathcal{H}}{1+R}\frac{d}{d\tau}$. The same steps can be used in the synchronous gauge. The reason it is possible to neglect the damping term is that for modes k which are "within the horizon", in the sense $k\tau \gtrsim 1$, the friction is small: $k^2 \gtrsim \mathcal{H}^2$, since $\mathcal{H} \sim \tau^{-1}$ in a matter or radiation dominated era. The solutions to the homogeneous equations $(\frac{d^2}{d\tau^2} + k^2c_s^2)S = 0$ are

$$S_1(k,\tau) = \sin kr_s(\tau) \tag{13.293}$$

$$S_2(k,\tau) = \cos kr_s(\tau) \tag{13.294}$$

where

$$r_s(\tau) \equiv \int_0^\tau d\tau^{'} c_s(\tau) \tag{13.295}$$

is the sound horizon, or the maximum comoving distance a sound wave may have traveled during a time τ. The general solution is a linear combination of the homogeneous solutions with a particular solution. This can be found using the retarded Green function for the differential operator $\frac{d^2}{d\tau^2} + k^2c_s^2$. We will postpone derivation of the explicit Green function until the end, to avoid obfuscating the discussion. Its form is given by (13.308). The general solution is[126]

$$\begin{aligned}\delta_\gamma(\tau) - 4\phi(\tau) &= A_1S_1(\tau) + A_2S_2(\tau) \\ &-\frac{4k^2}{3}\int_0^\tau d\tau^{'}(\frac{R}{1+R}\phi(\tau^{'}) + \psi(\tau^{'}))\times \\ &\frac{S_1(\tau^{'})S_2(\tau) - S_1(\tau)S_2(\tau')}{S_1(\tau^{'})\dot{S}_2(\tau^{'}) - \dot{S}_1(\tau^{'})S_2(\tau^{'})}\end{aligned} \tag{13.296}$$

This equation would have the same form if we had included the damping term, but with different homogeneous solutions. A numerical solution, which includes the tight coupling

approximation at the earliest time, is found in figure 13.3, and presents the oscillatory phenomena we have predicted here.

Let's determine the initial conditions at $\tau_0 = 0$ (this may be the Big Bang *or* the end of inflation phase and the beginning of a radiation dominated era). We will discuss initial conditions to the perturbations properly in section (13.15). Here we note that, from experiments measuring the CMB and thus precisely these oscillations, the initial conditions are such that δ_γ and ϕ (as well as ψ) are a constant value at very early time. These are known as adiabatic, or isentropic, initial conditions. In the synchronous gauge, the same initial conditions are not given by constants δ_γ and h, however $\delta_\gamma + \frac{2}{3}h$ is, and the same arguments apply. Due to these initial conditions

$$\delta_\gamma(0) - 4\phi(0) = A_2 \tag{13.297}$$

$$\dot{\delta}_\gamma(0) - 4\dot{\phi}(0) = A_1 = 0 \tag{13.298}$$

Therefore, only the cosine mode is excited. Using some trigonometry, we obtain the solution

$$\begin{aligned}\delta_\gamma(\tau) - 4\phi(\tau) &= (\delta_\gamma(0) - 4\phi(0))\cos kr_s(\tau) + \\ &+\frac{4k}{3c_s}\int_0^\tau d\tau'\left(\frac{R}{1+R}\phi(\tau') + \psi(\tau')\right)\sin(kr_s(\tau') - kr_s(\tau))\end{aligned}$$

Although the potentials are still present, this solution is really neat. We have, with some approximation, passed from a large number of coupled Boltzmann equations to this shorter one. Indeed, another numerical approximation that can be used is to treat the metric terms as only generated by dark matter. We could decouple the metric+dark matter equations from the photon-baryons and calculate the potentials separately, then plug them in the above equation. This might seem at first to be too much simplification, but it turns out to get the location of the acoustic peaks correctly although it consistently overestimates the height[80]. This was to be expected since we neglected the damping.

In particular, the time τ_p of the acoustic peaks, in absolute value, for a mode k can be estimated by simply looking at the homogeneous solution. It turns out that it is usually the dominating term.

$$r_s(\tau_p) = \frac{n\pi}{k}\ n = 1, 2, \ldots \tag{13.299}$$

Since r_s depends only on background quantities ρ_γ and ρ_b we have actually gained a huge amount of knowledge about the evolution of a large number of coupled differential equations.

Another important information we obtain is about the dipole of the photon fluid θ_γ (which is approximately equal to θ_b in the tight coupling limit). By equation (13.253) we find that $\theta_\gamma = -\frac{3}{4}(\dot{\delta}_\gamma - 4\dot{\phi})$ therefore

$$\begin{aligned}\theta_\gamma &= \frac{3}{4}c_s(\tau)(\delta_\gamma(0) - 4\phi(0))\sin kr_s(\tau) + \\ &+k\int_0^\tau d\tau'\left(\frac{R}{1+R}\phi(\tau') + \psi(\tau')\right)\cos(kr_s(\tau') - kr_s(\tau))\end{aligned} \tag{13.300}$$

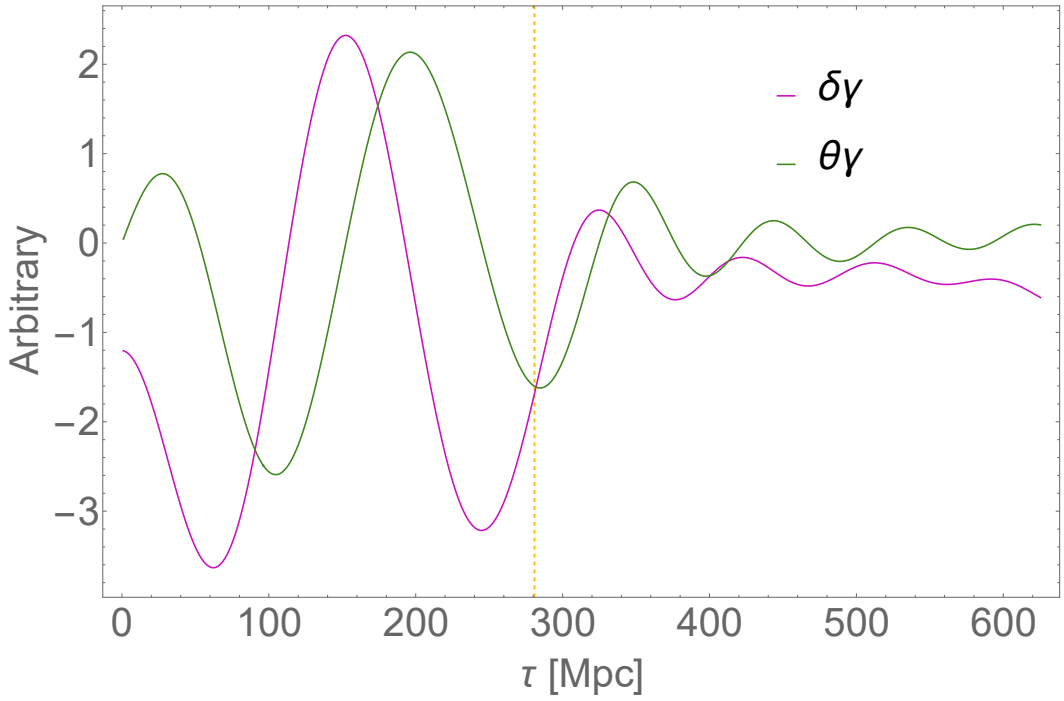

Figure 13.6: A numerical solution, for $k = 0.7 \cdot 10^{-4} Mpc$, showing how the monopole and dipole of the photon distribution are out of phase with one another. This result was deduced from the tight coupling approximation and remains valid even after decoupling (depicted by the dashed yellow line). The calculation was done with the CLASS Boltzmann code[26] using the best fit ΛCDM parameters[60].

Again, assuming the homogeneous term dominates, we find that the dipole, or velocity, of the fluid is out of phase with the monopole. This feature remains true even when solving the complete Boltzmann equations and is illustrated in the numerical solution in figure 13.6. A numerical solution for the photon velocity for different modes k is shown in figure 13.4.

13.14.1 Determination of the Green function

To conclude this section, let's derive the retarded Green function we used in (13.296). The Green function $G(\tau, \tau^{'})$ is a function of τ parametrized by $\tau^{'}$. It is defined as the inverse of the differential operator in the sense of the distributions

$$\left(\frac{d^2}{d\tau^2} + \beta(\tau)\frac{d}{d\tau} + \alpha(\tau)\right) G(\tau, \tau^{'}) = \delta(\tau - \tau^{'}) \tag{13.301}$$

If we had such a function, then a particular solution of the non-homogeneous equation with forcing term $F(\tau)$ would be given by

$$S_p(\tau) = \int_0^{\tau_0} d\tau^{'} G(\tau, \tau^{'}) F(\tau^{'}) \tag{13.302}$$

Where τ_0 is an upper value of conformal time which we can take to be arbitrarily large so long as α, β and F are smooth in $[0, \tau_0]$.

With the general solution being a linear combination of S_p with the two solutions of the homogeneous equation $S_{1,2}(\tau)$.

The following construction is valid for any second-order linear differential equation with variable coefficients[199]. The retarded Green function is constructed by imposing initial conditions $G(0,\tau^{'}) = \dot{G}(0,\tau^{'}) = 0$. Other possible Green functions can be obtained by choosing different initial conditions. These conditions are those we want to impose on the particular solution. By inspecting our differential equation (13.301), we find that $G(\tau,\tau^{'})$ satisfies the homogeneous equation for every τ except $\tau = \tau^{'}$. Therefore, the Green function must be a linear combination of homogeneous solutions in the intervals $[0,\tau^{'})$ and $(\tau^{'},\tau_0]$. The constants of the linear combination may be different in the two intervals and *may depend on* $\tau^{'}$. Thus

$$G(\tau,\tau^{'}) = \begin{cases} A_1(\tau^{'})S_1(\tau) + A_2(\tau^{'})S_2(\tau) & \tau < \tau^{'} \\ B_1(\tau^{'})S_1(\tau) + B_2(\tau^{'})S_2(\tau) & \tau > \tau^{'} \end{cases} \tag{13.303}$$

By imposing the initial conditions, we obtain an equation for the coefficients

$$\begin{pmatrix} S_1(0) & S_2(0) \\ \dot{S}_1(0) & \dot{S}_2(0) \end{pmatrix} \begin{pmatrix} A_1 \\ A_2 \end{pmatrix} = 0 \tag{13.304}$$

The determinant of the matrix of coefficients is the Wronskian. It is zero since $S_{1,2}$ are linearly independent. Therefore the only solution is

$$A_{1,2} = 0 \tag{13.305}$$

The retarded Green function is zero for $\tau < \tau^{'}$. Looking at the particular solution (13.302), the physical significance of this is that the solution is only affected by values of the forcing function in the past. It is causal.

To obtain the values of the coefficients $B_{1,2}(\tau^{'})$ we must match the function at $\tau = \tau^{'}$. Being a distribution $G(\tau,\tau^{'})$ may be discontinuous, even infinitely so, at $\tau = \tau^{'}$. Suppose the weakest discontinuity, given by the step function $\theta(\tau - \tau^{'})$. Then $\frac{d}{d\tau}G = \delta(\tau - \tau^{'})$ and $\frac{d^2}{d\tau^2}G = \delta^{'}(\tau - \tau^{'})$. Looking at the definition (13.301), assuming α and β are smooth functions, we see that there is no way that a derivative of the Dirac delta may appear to cancel out the one coming from the second derivative of the Green function. We conclude $G(\tau,\tau^{'})$ must be continuous at $\tau = \tau^{'}$. Its derivative may not, but it may at most contain a step function at $\tau^{'}$.

Now we take the definition and integrate it on a very small interval around $\tau^{'}$

$$\int_{\tau^{'}-\epsilon}^{\tau^{'}+\epsilon} d\tau \ddot{G}(\tau,\tau^{'}) + \beta(\tau)\dot{G}(\tau,\tau^{'}) + \alpha(\tau)G(\tau,\tau^{'}) = \int_{\tau^{'}-\epsilon}^{\tau^{'}+\epsilon} d\tau\delta(\tau - \tau^{'}) = 1 \tag{13.306}$$

Since $G(\tau,\tau^{'})$ is continuous, the integral of the term αG is zero as $\epsilon \to 0$. The same is true for the integral of the term $\beta\dot{G}$. Even if $\dot{G}$ where a step function, the integral of the step is zero, and β is a continuous function. Only the second derivative term may contribute. By

the fundamental theorem of calculus

$$\dot{G}(\tau)_{\tau\to\tau'_+} - \dot{G}(\tau)_{\tau\to\tau'_-} = 1 \tag{13.307}$$

Applying the gluing conditions on the G and its derivative on both sides we obtain

$$\begin{aligned} B_1 S_1(\tau^{'}) + B_2 S_2(\tau^{'}) &= 0 \\ B_1 \dot{S}_1(\tau^{'}) + B_2 \dot{S}_2(\tau^{'}) &= 1 \end{aligned}$$

The system is simply solved and we can write the final form of the retarded Green function

$$G(\tau,\tau^{'}) = \theta(\tau-\tau^{'}) \left(\frac{S_1(\tau^{'})S_2(\tau) - S_1(\tau)S_2(\tau^{'})}{S_1(\tau')\dot{S}_2(\tau') - S_2(\tau')\dot{S}_1(\tau')} \right) \tag{13.308}$$

13.15 Initial conditions

Any calculations of matter perturbations today must be given some suitable initial conditions. These initial conditions are, a priori, unknown. On paper, we may only give useful parametrizations of the initial conditions. Indeed, this is a deep philosophical issue in cosmology. By what mechanism, if any, are the primordial perturbations seeded? The Boltzmann equations we have formulated are homogeneous in the perturbation variables. Perturbations cannot grow out of a perfectly uniform universe. If the universe appears to be nearly uniform, which is the nature of the horizon problem, what broke this uniformity at very early stages? As with the horizon problem, the solution may come from the theory of primordial inflation. In fact, the initial conditions we *deduce from observation* are consistent with those that may be generated by the quantum fluctuations of the scalar field which is thought to drive inflation. Here we will not prove this, and we will study the form of the Boltzmann equations at very early time.

It must be pointed out that the initial conditions, as deduced from observation of the CMB spectrum, seem to be statistical in nature. We cannot point to a single "correct" initial condition, but rather to its statistical properties, which we encode in the statistical correlations between the values of different fields. Furthermore, we notice that, since the equations are linear and first order, if we were to multiply every initial value by a constant, the same solution would hold after multiplication by the same constant. For the same reason we can separate out different "type" of initial conditions, for example corresponding to different kinds of statistical correlations, and solve the equations for each initial condition separately. Once we have a general linear combination of initial conditions, we may combine linearly the solutions.

To understand the form of the initial conditions, we analyze the Boltzmann equations for modes k that are well out of the horizon τ. In fact, the conformal time τ is the maximum comoving distance a photon may have traveled since the radiation era has begun, whether we call that Big Bang or the end of inflation. Thus, if the state of a Fourier mode is set at the beginning, with wavelength $\sim k^{-1}$, we don't expect any causal physics may affect it until $k\tau \sim 1$. The causality structure is, of course, already included in the Boltzmann equations, we just make it more evident. It will turn out, as we shall soon see, that the behavior of

a mode k well outside the horizon is gauge dependent. This is not a problem, since by causality it cannot affect any physical observables.

Let's work with scalar modes in the synchronous gauge explicitly. We take an initial time τ such that any mode which may be of interest *today* is well outside the horizon, $k\tau \ll 1$. Clearly, we must be deep in the radiation era and so the total density of neutrinos and photons dominate the Friedmann equations. The conformal Hubble factor is exactly given by, see(7.46),

$$\mathcal{H} = \frac{1}{\tau} \tag{13.309}$$

We will use this substitution freely here. Due to the Friedmann equation

$$\frac{8\pi G a^2}{3} = \frac{1}{\tau^2 \rho} \tag{13.310}$$

where $\rho = \rho_\nu + \rho_\gamma$ is the total density. With these equalities, we combine the time-time (13.119) and spatial trace (13.122) equations to eliminate the term in η and obtain

$$\tau^2 \ddot{h} + \tau \dot{h} + 6\left((1 - R_\nu)\delta_\gamma + R_\nu \delta_\nu\right) = 0 \tag{13.311}$$

where

$$R_\nu = \frac{\rho_\nu}{\rho_\nu + \rho_\gamma} \tag{13.312}$$

is the ratio of neutrino density to total density. After electron-positron annihilation this value is constant, given by (8.46). We will now work at the lowest order, neglecting any term proportional to k^2 in the Boltzmann equations. If we assume the tight coupling limit between photons and baryons then we immediately obtain

$$\dot{\theta}_\gamma = \dot{\theta}_b = 0 \tag{13.313}$$

$$\theta_\gamma = \theta_b \tag{13.314}$$

at the lowest order in $k\tau$. Differentiating (13.311) twice we obtain[155]

$$\tau \frac{d^4 h}{d\tau^4} + 5\frac{d^3 h}{d\tau^3} = 0 \tag{13.315}$$

which has solutions of the form $h = A_n (k\tau)^n$ for some constant A_n, and the k is included to make the A_n dimensionless. The possible values of n can be found by substitution and are $n = 0, 1, 2, -2$. The most general form of h outside of horizon is of the form

$$h = A_0 + A_1 k\tau + A_2 k^2 \tau^2 + A_{-2}(k\tau)^{-2} \tag{13.316}$$

The constants need to be set by initial conditions. It can be shown that the solution with $n = 0, -2$ are non-physical and can be eliminated by a gauge transformation. Regardless, we are prone to choose the mode with the largest power of τ. In fact, if any mode were excited in the earliest moments after the Big Bang, the ones which enter the horizon, much

later, will be dominated by the A_2 mode. Thus we will write

$$h = A_2 k^2 \tau^2 \tag{13.317}$$

At the same order we can immediately find the initial densities for photons, baryons, neutrinos and dark matter. Indeed, for each of these the equation for the density perturbation is proportional to $\dot{h}$, such that

$$-\frac{2}{3}\dot{h} = \dot{\delta}_\gamma = \dot{\delta}_\nu = \frac{4}{3}\dot{\delta}_b = \frac{4}{3}\dot{\delta}_{cdm} \tag{13.318}$$

at very early times. Using the time-time Einstein equation (13.119) we can show that at this order

$$\eta = 2A_2 \tag{13.319}$$

Curiously, this analysis of modes much larger than the horizon has provided the fact that at early times the derivatives of the densities are all proportional to one another and to the derivative of the metric. They evolve together. What is not fixed is the actual ratio of values. In fact if we integrate $\dot{\delta}_\gamma = \frac{4}{3}\dot{\delta}_b$ we obtain

$$\delta_{\gamma,i} = \frac{4}{3}\delta_{b,i} + \text{const} \tag{13.320}$$

This constant of integration may be different for each species. Initial conditions are classified as *adiabatic, or isentropic*, if all these constants vanish. If one or more are non-zero they are classified as *isocurvature perturbations*. The set of all adiabatic initial conditions is a one-dimensional vector space. Since all the constants of integrations are set to zero, the only freedom is the value A_2, or alternatively of the initial photon density $\delta_{\gamma,i}$. On the other hand there are many more possibilities when dealing with isocurvature perturbations. We will not discuss the parametrization of isocurvature perturbation. In fact, it has been experimentally verified from analysis of the CMB that primordial perturbations are adiabatic. If there are any isocurvature perturbations they play a much smaller, but not less interesting role. We will come back to adiabatic perturbations momentarily.

First, let's return to the study of initial conditions. Having found the zero order terms we can now look at first order initial conditions for $\theta_{\nu,b,\gamma}$. Neglecting higher moments the equation for the photon dipole is $\dot{\theta}_\gamma = \frac{k^2}{4}\delta_\gamma$ which implies

$$\theta_\gamma = \theta_b = -h\frac{k^2\tau}{18} \tag{13.321}$$

We removed any constant of integration, so we are now working in the adiabatic choice. Getting the neutrino velocity θ_ν is a little more involved. It involves differentiating the equation for $\dot{\theta}_\nu$ (13.173) and substituting the equations for $\dot{\delta}_\nu$ (13.172) and $\dot{F}_{\nu 2}$ (13.174) , while ignoring higher moments $F_{\nu 3}$. Then one obtains a second order equation for θ_ν. Assuming that $\theta_\nu \sim k^4\tau^3$, the same dependence of θ_γ, and working through the algebra one obtains

$$\theta_{\nu,i} = \frac{23 + 4R_\nu}{15 + 4R_\nu}\theta_\gamma \tag{13.322}$$

having removed the integration constants. Unlike the photon quadrupole, $F_{\nu 2}$ is not negli-

gible, since it is not washed out by Compton scattering.

$$F_{\nu 2,i} = \frac{2}{3}\frac{h}{15+4R_\nu} \tag{13.323}$$

Finally, using the time-space Einstein equation (13.124) we can derive a better approximation for η

$$\eta = 2A_2 - \frac{5+4R_\nu}{6(15+4R_\nu)}h \tag{13.324}$$

Repeating the same calculation for the conformal-Newtonian gauge we can find, after some not particularly illuminating algebra,

$$\psi = \frac{20A_2}{15+4R_\nu} \tag{13.325}$$

$$\phi = (1+\frac{2}{5}R_\nu)\psi \tag{13.326}$$

$$-2\psi = \delta_\gamma = \delta_\nu = \frac{4}{3}\delta_b = \frac{4}{3}\delta_{cdm} \tag{13.327}$$

$$\theta_\gamma = \theta_b = \theta_\nu = \theta_{cdm} = \frac{1}{2}k^2\tau\psi \tag{13.328}$$

$$F_{\nu 2} = \frac{1}{30}(k\tau)^2\psi \tag{13.329}$$

These equations are for adiabatic initial conditions. A_2 is the same coefficient as defined for the fastest growing mode in the synchronous gauge. These equations can be easily used to relate the initial conditions in both gauges through A_2. Note that the fastest growing mode in the conformal-Newtonian gauge is $\psi \sim$ const. The metric potentials in the conformal-Newtonian gauge are constant well outside the horizon. This can be easily shown to be the same mode of $h = A_2k^2\tau^2$ using the gauge transformations (13.38) and (13.39).

Let's return to discussing these initial adiabatic conditions. In both gauges it is clear the ratios between the various densities and the metric, as well as ratios among the velocities, are constant since early times. Indeed, as we had anticipated, all the initial values can be related to a single free parameter, which we can choose to be, arbitrarily, the photon overdensity δ_γ or the metric h, ψ. These choices are common since they are relatable to other more experimentally viable quantities with easy.

An understanding of the adiabatic perturbations can be gained by returning to real space. We have been working so much in Fourier space it is easy to forget this. In real space, at every point in space there is a common over or under-density given by the size, and sign, of δ_γ. The relative composition is size independent. Since usual equilibrium evolution does not change relative composition, we may think that the thermal evolution of every point in space is slightly early or in advance of the uniform background, as if the same *exact initial conditions* where set the same everywhere but at slightly different times. We may write this as[17]

$$\delta\rho_i(\vec{x},\tau) = \rho_i(\vec{x},\delta\tau(\vec{x})) - \rho_i(\vec{x}) = \dot{\rho}_i(\vec{x})\delta\tau(\vec{x}) \tag{13.330}$$

This pointlike common time shift $\delta\tau(\vec{x})$ parametrizes the adiabatic perturbations. Thermal evolution is isentropic, or adiabatic, and so the perturbation to entropy density is zero.

Since the shift is common to all species, the ratio

$$\frac{\delta\rho_i}{\dot{\rho}_i} = \text{const} \tag{13.331}$$

is independent of species. Using the continuity equation $\dot{\rho}_i = -3\mathcal{H}(1+w)\rho_i$ we obtain that

$$\frac{\delta_i}{1+w_i} = \text{const} \tag{13.332}$$

Using this equation we get back the initial conditions already found (13.327). This formula is useful to quickly grasp what the adiabatic initial condition might be for an exotic species.

Adiabatic perturbations are, so far, the only measured perturbations in cosmology. The CMB and structures detected originating from the primordial universe can be described very successfully by them. The picture that every point has a time-shift is suggestive in the context of the theory of inflation. Inflation is generally described by a scalar field with a high potential energy, thus mimicking for some time a constant energy density causing the universe to expand exponentially (like the late dark energy phase). Eventually, the field falls to the bottom of the potential and the accelerated expansion stops, allowing the normal evolution begin. During the evolution of the field, quantum fluctuations may change its value in a probabilistic way. In the sense of the uncertainty principle, the field may be at a point higher or lower in the potential than classical evolution dictates. At various points in space, the field reaches the minimum at slightly different times. Assuming the field can be efficiently converted to standard model particles, the same evolution begins everywhere in the universe but at slightly different times, giving rise to adiabatic perturbations.

So are adiabatic perturbations a sign of inflation? Perhaps. However it cannot be conclusive. In fact, we don't know what theory gives the initial conditions ab initio. It may be that the initial conditions just turn out to be this way. In a certain sense, they look "natural".

We close this section by describing the statistical nature of adiabatic perturbations. We know we can solve the Boltzmann equations provided some initial value of the metric potential ψ (as we have seen we can relate initial conditions in the synchronous gauge to this value as well). It does not matter what value we choose as long as the rest of the conditions are fixed, so we may choose the initial value of, for example, $\psi(\vec{k},\tau_i) = 1$. At the end of the calculation we may multiply all the solutions by any constant.

This is useful since the initial perturbations are statistical in nature. For any Fourier mode

$$< \psi(\vec{k},\tau_i) >= 0 \tag{13.333}$$

This is equivalent to taking the perturbation average to be zero in real space as well, ie $< \psi(\vec{x},\tau_i) >= 0$ and therefore is simply the statement that the zero-order metric is the average metric. In fact the only value this average may take if we assume the cosmological principle is a constant, which may then well be zero. Recall way back in section 5, we had discussed that the cosmological perturbations, the basis of structures, obviously violate the principles of homogeneity and isotropy, which must be then understood in a statistical

sense. This is exactly the statistical sense we were talking about.

To understand the statistical properties of $\psi(\tau_i)$ we look at moments of the distribution of initial values.

$$< \psi(\vec{k},\tau_i)\psi^*(\vec{k}^{'},\tau_i) >= \frac{1}{(2\pi)^3}P_\psi(k)\delta^3(\vec{k}-\vec{k}^{'}) \tag{13.334}$$

where we defined the *power spectrum* $P_\psi(k)$. The power spectrum has dimension of k^{-3}. Care should be taken in confronting formulas among the literature since another common definition is to make the k^3 explicit with the *dimensionless power spectrum* $P_\psi^{dimensionless} = P_\psi(k)\cdot k^3$. Other definitions may or may not include the factor $(2\pi)^3$. Let's discuss the form of the above correlator. First of all, we write this simply because experiment confirms it. In fact the form of $P_\psi(k)$ has been measured for a cosmologically relevant range of k. *A priori* It can only depend on the magnitude of $\vec{k}$ due to *statistical isotropy* of the cosmological principle. Actually, checking if this is the case is a robust test for standard cosmology. Next, we note that there is no correlation between modes with different wave vectors $\vec{k}$ due to the Dirac delta. Every mode is independent from every other not only during evolution, due to the equations being linear, but in their initial conditions. This can be linked again to statistical isotropy in real space. In fact, the correlator

$$\begin{aligned} < \psi(\vec{x},\tau_i)\psi^*(\vec{x}^{'},\tau_i) > &= \int d^3k d^3k^{'} e^{i\vec{k}\cdot\vec{x}-i\vec{k}^{'}\cdot\vec{x}^{'}} < \psi(\vec{k},\tau_i)\psi^*(\vec{k}^{'},\tau_i) > \\ &= \frac{1}{2\pi^2}\int kdk\frac{\sin|\vec{x}-\vec{x}^{'}|}{|\vec{x}-\vec{x}^{'}|}P_\psi(k) = \xi(|\vec{x}-\vec{x}^{'}|) \end{aligned}$$

As it should be, since the spatial correlator can only depend on the magnitude of the separation of two points. In principle there may be a correlator between more than two fields $\psi(\vec{k},\tau_i)$. These should not be discarded theoretically, although they have not been experimentally detected but may arise in more exotic models. Thus we will just keep the two point correlator henceforth.

13.16 Line of sight solution

The higher moments $F^{(m)}_{(\nu,\gamma)\ell}$ of neutrino and photons cannot be neglected in a calculation. Indeed these moments are present even today and, in the case of photons from CMB, are directly measurable. Not only are the photon moments not negligible, we'd like to calculate them from theory with a great degree of accuracy. It seems this comes at a severe computational cost. In order to accurately describe the CMB fluctuations we observe today, we must calculate moments up to $\ell \sim 3000$. As we have seen in the set of equations following (13.250), the equations to solve are one per value of l, for each mode ($m = 0$ scalars, $m = 1$ vectors or $m = 2$ tensors). Eventually, we should cut off the hierarchy at some very high $\ell \gtrsim 3000$. We will now describe an alternative exact analytical solution to get the value of $F^{(m)}_{\gamma\ell}$ at high ℓ once the moments at low ℓs, usually up to $\ell \sim 10$, are well calculated. If we can do this, we can truncate the hierarchy of equations at very low ℓ , since we only need to know exactly the lower moments to find the higher ones. This method is known as the *line of sight solution* to the Boltzmann equations[192]. It is also useful to arrive to a more physical picture of what determines the CMB perturbations today.

We start with the full Boltzmann equation for the intensity of photons for scalar or tensor

modes given by (13.235) and (13.236). These are the equations we had obtained before projecting on the various angular dependencies with the spherical harmonics. We rearrange the terms as follows

$$\dot{F}_\gamma + ik\mu F_\gamma - \dot{\tau}_c F_\gamma = S_\gamma(\vec{k},\tau,\hat{n}) \tag{13.335}$$

where $\dot{\tau}_c = -an_e\sigma_T$ is the optical depth we had defined in (13.280). S_γ is a *source term* whose form is given by the Boltzmann equations and can be simply be seen to be

$$\begin{aligned}(\text{Newtonian})S_\gamma(\vec{k},\tau,\hat{n}) &= 4\dot{\phi}\sqrt{4\pi}Y_0^0 - 4ik\psi\sqrt{\frac{4\pi}{3}}Y_1^0 - 4\dot{\tau}_c v_b\sqrt{\frac{4\pi}{3}}Y_1^0 \\ &\quad -\dot{\tau}_c F_{\gamma 0}^{(0)}\sqrt{4\pi}Y_0^0 + \dot{\tau}_c\Pi^{(0)}\sqrt{4\pi}\sqrt{5}Y_2^0\end{aligned} \tag{13.336}$$

$$\begin{aligned}(\text{Synchronous})S_\gamma(\vec{k},\tau,\hat{n}) &= -\frac{2}{3}\dot{h}\sqrt{4\pi}Y_0^0 - \frac{4}{3}(\dot{h}+6\dot{\eta})\sqrt{\frac{4\pi}{5}}Y_2^0 - 4\dot{\tau}_c v_b\sqrt{\frac{4\pi}{3}}Y_1^0 \\ &\quad -\dot{\tau}_c F_{\gamma 0}^{(0)}\sqrt{4\pi}Y_0^0 + \dot{\tau}_c\Pi^{(0)}\sqrt{4\pi}\sqrt{5}Y_2^0\end{aligned} \tag{13.337}$$

$$\begin{aligned}(\text{Tensor})S_\gamma(\vec{k},\tau,\hat{n}) &= -(h_+ - ih_\times)\sqrt{\frac{8\pi}{15}}Y_2^2 \\ &\quad +\dot{\tau}_c\Pi^{(2)}\sqrt{4\pi}\sqrt{5}Y_2^2\end{aligned} \tag{13.338}$$

where we have kept only the $m = 2$ terms in the tensor source for simplicity. The $m = -2$ should always be considered as well, they have a similar form with $h_+ - ih_\times \to h_+ + ih_\times$. We recall that the polarization source $\Pi^{(m)} = \frac{1}{10}(F_{\gamma 2}^{(m)} - \sqrt{6}E_2^{(m)})$ was defined in (13.249). We have made all the angular dependence explicit through spherical harmonics, since we will want to project onto them at then end.

Now we take (13.335) and note that the left hand side can be written as

$$\frac{d}{d\tau}(F_\gamma e^{-\tau_c}e^{ik\mu\tau})e^{\tau_c}e^{-ik\mu\tau} \tag{13.339}$$

We integrate both sides in τ from $\tau = 0$, or some very early time, to τ_0, the conformal time today. Since $\tau_c \gg 1$ at early times, one extreme of integration on the left hand side vanishes. With $\tau_c(\tau_0) = 0$ we obtain

$$F_\gamma(\vec{k},\tau_0,\hat{n}) = \int_0^{\tau_0} d\tau e^{ik\mu(\tau-\tau_0)}e^{-\tau_c}S_\gamma(\vec{k},\tau,\hat{n}) \tag{13.340}$$

Recall that $\mu = \hat{k}\cdot\hat{n}$. We define $r = \tau_0 - \tau$ which is the comoving coordinate of a photon at time τ, if it is received at $r = 0$ today. We note that a plane wave may be expanded as follows in spherical harmonics with $m = 0$

$$e^{-i\vec{k}\hat{n}r} = \sum_\ell (-i)^\ell\sqrt{4\pi(2\ell+1)}j_\ell(kr)Y_\ell^0(\hat{n}) \tag{13.341}$$

Where $j_\ell(x)$ are the spherical Bessel functions of the first kind. The angular expansion of F_γ was given by (13.240) and we define the spherical expansion of S_γ as

$$S_\gamma = \sum_{\ell,m}(-i)^\ell \frac{\sqrt{4\pi}}{\sqrt{2\ell+1}} S_{\gamma\ell}^{(m)}(\vec{k},\tau) Y_\ell^m(\hat{n}) \tag{13.342}$$

Note the different normalization in the definition of the angular coefficients, $(2\ell+1)^{-\frac{1}{2}}$ instead of $(2\ell+1)^{\frac{1}{2}}$ defined in F_γ . This is done in order to simplify some factors in the formulas that follow. The source only contains terms with $\ell \leq 2$. With this normalization they are

$$\begin{aligned} S_{\gamma,0}^{(0)} &= 4\dot{\phi} - \dot{\tau}_c F_{\gamma 0}^{(0)} \text{(Newtonian)} \\ &= -\frac{2}{3}\dot{h} - \dot{\tau}_c F_{\gamma 0}^{(0)} \text{(Synchronous)} \end{aligned} \tag{13.343}$$

$$\begin{aligned} S_{\gamma,1}^{(0)} &= 4k\psi - 4\dot{\tau}_c \frac{\theta_b}{k} \text{(Newtonian)} \\ &= -4\dot{\tau}_c \frac{\theta_b}{k} \text{(Synchronous)} \end{aligned} \tag{13.344}$$

$$\begin{aligned} S_{\gamma,2}^{(0)} &= -5\dot{\tau}_c \Pi^{(0)} \text{(Newtonian)} \\ &= \frac{4}{3}(\dot{h}+6\dot{\eta}) - 5\dot{\tau}_c \Pi^{(0)} \text{(Synchronous)} \end{aligned} \tag{13.345}$$

$$S_{\gamma,2}^{(\pm 2)} = \sqrt{\frac{3}{2}}(h_+ \mp h_\times) - 5\dot{\tau}_c \Pi^{(2)} \text{(Tensor)} \tag{13.346}$$

We now project (13.340) on the angular modes by integrating both sides with $\int d\Omega Y_L^{M*}(\hat{n})$. On the right hand side, we will have integrals of three spherical harmonics. It is possible to rewrite one outer product through Clebsch-Gordan coefficients. The result can be expressed through Wigner-3j symbols using (see Appendix B)

$$\begin{aligned} \int d\Omega \, {}_S Y_L^{M*}(\hat{n})\, {}_{s1}Y_{\ell 1}^{m_1}(\hat{n})\, {}_{s_2}Y_{\ell_2}^{m_2}(\hat{n}) &= (-1)^{S+M}\sqrt{\frac{(2L+1)(2\ell_1+1)(2\ell_2+1)}{4\pi}} \\ &\quad \begin{pmatrix} L & \ell_1 & \ell_2 \\ -M & m_1 & m_2 \end{pmatrix} \begin{pmatrix} L & \ell_1 & \ell_2 \\ S & -s_1 & -s_2 \end{pmatrix} \end{aligned}$$

where

$$\begin{pmatrix} j_1 & j_2 & j_3 \\ m_1 & m_2 & m_3 \end{pmatrix} \tag{13.347}$$

are the Wigner-3j symbols, related to the Clebsch-Gordan coefficients. For regular spherical

harmonics, the spins $S = s_1 = s_2 = 0$. We define the *radial functions*[128]

$$j_L^{(\ell m)}(x) = \sum_{\ell'}(-i)^{\ell+\ell'-L}(-1)^m(2\ell'+1)j_{\ell'}(x)\times \begin{pmatrix}\ell & \ell' & L\\ m & 0 & -m\end{pmatrix}\begin{pmatrix}\ell & \ell' & L\\ 0 & 0 & 0\end{pmatrix}$$

which is a sum that appears when doing our angular projection. In fact using this definition we obtain the simple relation

$$F_{\gamma L}^{(m)}(k,\tau_0) = \int_0^{\tau_0} d\tau e^{-\tau_c}\sum_\ell S_{\gamma\ell}^{(m)}(k,\tau)j_L^{\ell m}(kr) \tag{13.348}$$

This is a powerful formula. It directly givs the values of the moments today in terms of a total of five source terms, three for the scalar modes and two for the tensor modes. The problem of finding $F_{\gamma L}^{(m)}$ has separated into a physical part and a geometrical term $j_L^{(\ell m)}(kr)$ which *does not depend in any way on the physics.* This means that the radial functions can be precalculated and stored for use. It is clear why it is known as a line of sight integral, the physics is contained in $S_{\gamma\ell}^{(m)}$ and integrated over the trajectory of the photon.

Note how there is no sum on m, which is due to the fact, in the parlance of quantum mechanics, that angular moment along an axis is conserved, so the Clebsch-Gordan only link states with $m+m^{'} = M$. The plane wave expansion has only given a term with $m = 0$. Again, scalar and tensor modes, with different values of m, don't mix.

Finally, we notice, as anticipated, that the source terms only depend on the lowest photons modes with $\ell \leq 2$. We can cut off the Boltzmann hierarchy in ℓ at a value $\ell \sim 10$ to get accurate solutions for the lower modes and then find the higher modes by integration. This gives a huge advantage in computational cost.

The radial functions, $j_L^{(\ell m)}(x)$, are a linear combination of the Bessel functions with $|L-\ell| \leq \ell^{'} \leq L+\ell$. The spherical Bessel equations satisfy the recurrence relations

$$j_\ell(x) = \frac{x}{2\ell+1}(j_{\ell-1}(x)+j_{\ell+1}(x)) \tag{13.349}$$

$$j_\ell^{'}(x) = \frac{1}{2\ell+1}(\ell j_{\ell-1}(x)-(\ell+1)j_{\ell+1}(x)) \tag{13.350}$$

which implies that the $j_L^{(\ell m)}(x)$ may be rewritten as a linear combination of $j_L(x)$ and its derivatives. The computation of these functions is lengthy but straightforward so we give the results[128]

$$j_L^{(00)} = j_L \tag{13.351}$$

$$j_L^{(10)} = j_L^{'} \tag{13.352}$$

$$j_L^{(20)} = \frac{3}{2}j_L^{''}+\frac{1}{2}j_L \tag{13.353}$$

$$j_L^{(22)} = j_L^{(2,-2)} = \sqrt{\frac{3}{8}\frac{(L+2)!}{(L-2)!}}\frac{j_L(x)}{x^2} \tag{13.354}$$

We don't need other radial functions, as the source function has only $\ell \leq 2$.

We repeat the same calculation with the photon polarization. Starting with the Boltzmann equations with the full angular dependence (13.237) and (13.238), we again put the equation in the form (13.335). The source terms have the same form for scalar, in both gauges, and tensor modes. Its angular dependence is proportional to ${}_{\pm 2}Y_2^m$ and so its angular projection is trivial (we use the same normalization as for the photon intensity source (13.342)).

$$S_{\pm 2}^{(m)}(\vec{k},\tau) = 5\sqrt{6}\dot{\tau}_c \Pi^{(m)} \tag{13.355}$$

where $m = 0$ for scalars and $m = \pm 2$ for tensor modes. The angular expansion for the polarization $F_\pm$ was given through E and B modes in (13.240). To obtain the line of sight integral we employ the same procedure as before, integrating and defining the real radial functions $\epsilon_L^{(m)}(x)$ and $\beta_L^{(m)}(x)$.

$$(E_L^{(m)} \pm iB_L^{(m)})(\vec{k},\tau_0) = \int_0^{\tau_0} d\tau e^{-\tau_c} S_{\pm 2}^{(m)} (\epsilon_L^{(m)} \pm i\beta_L^{(m)}) \tag{13.356}$$

So the $\epsilon_L^{(m)}$ function applies to the E mode and the $\beta_L^{(m)}$ to the B mode.

The definition of the radial functions is

$$\begin{aligned} \epsilon_L^{(m)}(x) \pm i\beta_L^{(m)}(x) &= \sum_{\ell'} (-i)^{\ell'-L}(-1)^m (2\ell'+1) j_{\ell'}(x) \times \\ & \begin{pmatrix} 2 & \ell' & L \\ m & 0 & -m \end{pmatrix} \begin{pmatrix} 2 & \ell' & L \\ \mp 2 & 0 & \pm 2 \end{pmatrix} \end{aligned}$$

Notice there is one less index compared to the $j_L^{(\ell m)}$, since the source term has only $\ell = 2$. The spherical Bessel functions $j_{\ell'}(x)$ and the Wigner-3j symbols are real, therefore the right hand side is pure imaginary when $\ell' - L$ is odd and pure real when $\ell' - L$ is even. This allows to clearly see which terms of the sum contribute to $\epsilon_L^{(m)}$ and which to $\beta_L^{(m)}$. For scalar modes, $m = 0$ and the Wigner-3j symbol is *non-vanishing* only if the sum of the integers in the first row is even. $\ell' + L$ is even whenever $\ell' - L$ is, therefore

$$\beta_L^{(0)} = 0 \tag{13.357}$$

implying there are no B modes for scalars. We had already noticed this when discussing the Boltzmann equations for the polarization of tensor modes (13.264). *The B modes are not sourced nor do they source the other perturbations.* We may think that there might have a been a primordial scalar B mode generated by some initial conditions. In fact, decoupling does not necessarily mean they are not there. Since the line of sight integral (13.356) is exact, the only place the primordial scalar B mode may contribute is in the integration constant we dropped. We had dropped this initial value constant $F_\pm(\vec{k},\hat{n},\tau=0)e^{-\tau_c(\tau=0)}$ since it is exponentially suppressed by the optical depth. If there were some scalar B initial condition, for some reason, it would be long gone.

The radial functions may also be calculated in a lengthy but straightforward manner using

the Wigner-3j symbols and the spherical Bessel functions recurrence relations. The necessary results are given by

$$\epsilon_L^{(0)}(x) = \sqrt{\frac{3}{8}\frac{(L+2)!}{(L-2)!}}\frac{j_L(x)}{x^2} \tag{13.358}$$

$$\epsilon_L^{(1)} = \frac{1}{2}\sqrt{(L-1)(L+2)}(\frac{j_L(x)}{x^2} + \frac{j_L^{'}(x)}{x}) \tag{13.359}$$

$$\epsilon_L^{(2)} = \frac{1}{4}\left(-j_L(x) + j_L^{''}(x) + 2\frac{j_L(x)}{x^2} + 4\frac{j_L^{'}(x)}{x}\right) \tag{13.360}$$

$$\beta_L^{(1)} = \frac{1}{2}\sqrt{(L-1)(L+2)}\frac{j_L(x)}{x} \tag{13.361}$$

$$\beta_L^{(2)} = \frac{j_L^{'}(x)}{2} + \frac{j_L(x)}{x} \tag{13.362}$$

Let's conclude this section by discussing the structure temperature scalar modes $F_{\gamma\ell}^{(0)}(\vec{k},\tau_0)$. First, note that this is the Fourier transform of the actual distribution today, thus it is directly related to observables, particularly to the CMB anisotropies today, viewed from Earth. The CMB anisotropies are none other than $F_{\gamma\ell}^{(0)}$. In the conformal-Newtonian gauge the line of sight integral (13.348) evaluates to

$$F_{\gamma\ell}^{(0)}(\vec{k},\tau_0) = \int_0^{\tau_0} d\tau e^{-\tau_c}\left[(4\dot{\phi} - \dot{\tau_c}F_{\gamma 0}^{(0)})j_\ell(kr) + (4k\psi - 4\dot{\tau_c}\frac{\theta_b}{k})j_\ell^{'}(kr) - 5\dot{\tau_c}\Pi^{(0)}j_\ell^{(20)}\right] \tag{13.363}$$

The prime denotes the derivative of j_ℓ with respect to its parameter $x = kr$. Let's drop the polarization source $\Pi^{(0)}$ for our discussion. The actual polarization of the CMB is $1 \div 10\%$ of the intensity of the perturbations. We can integrate one terms by parts

$$4k\psi j_\ell^{'}(kr)e^{-\tau_c} \to 4e^{-\tau_c}(\dot{\psi} - \dot{\tau_c}\psi)j_\ell(kr) \tag{13.364}$$

Indeed the boundary term at $\tau \to 0$ is proportional to $e^{-\tau_c} \to 0$, while at $\tau = \tau_0$ we have $j_\ell(0)$ which vanishes for any $\ell \neq 0$. The case $\ell = 0$ would be the monopole which, unfortunately, is unobservable from the Earth. We would not be able to distinguish it from the average temperature of the CMB. To detect the monopole, we would have to measure the CMB uniform temperature on the sky at different locations in the universe. So for every $\ell \neq 0$ we obtain

$$F_{\gamma\ell}^{(0)}(\vec{k},\tau_0) \simeq \int_0^{\tau_0} d\tau g(\tau)\left[(F_{\gamma 0}^{(0)} + 4\psi)j_\ell(kr) + 4\frac{\theta_b}{k}j_\ell^{'}(kr)\right] + 4e^{-\tau_c}(\dot{\phi} + \dot{\psi})j_\ell(kr) \tag{13.365}$$

In the above we introduce the *visibility function (figure 13.7)*

$$g(\tau) \equiv -\dot{\tau_c}e^{-\tau_c} \tag{13.366}$$

The signs make it positive. The visibility function is a fundamental quantity in our understanding of the CMB. Physically, it represents the probability distribution of the last scattering time of a photon. It follows a bell-shape and we may define the moment of de-

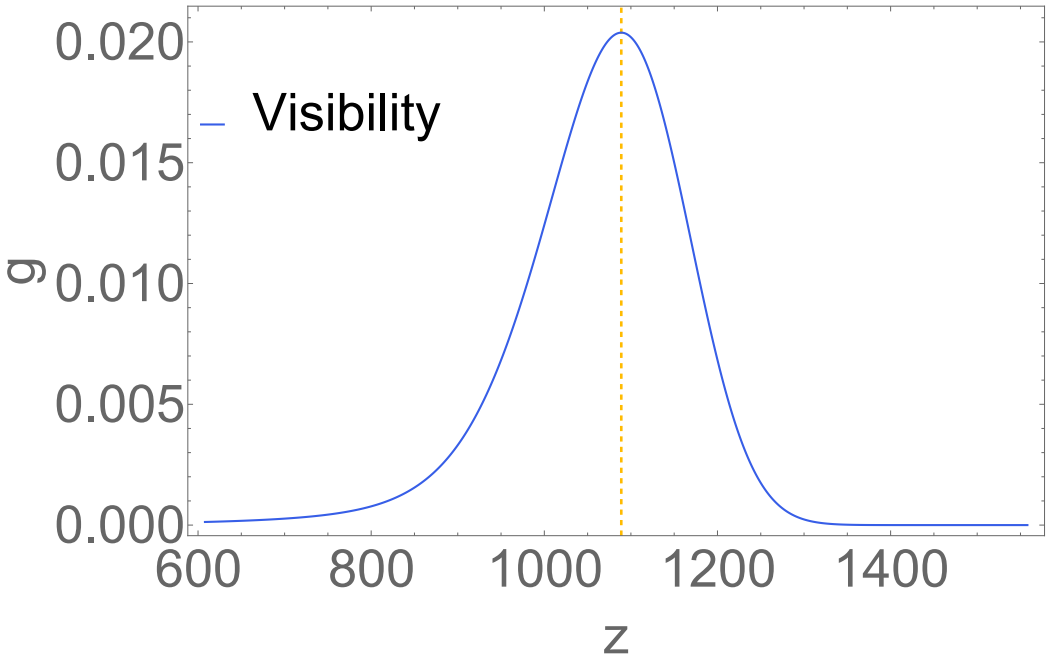

Figure 13.7: The visibility function $g(\tau)$ calculated by the CLASS Boltzmann code[26], which uses the RECFAST algorithm[187, 190, 191].. The peak is at $z = 1088$, which corresponds to $\tau \simeq 280 Mpc$. The best fit ΛCDM parameters were used in the calculation.

coupling through its peak. The photon moment $F^{(0)}_{\gamma\ell}$ contains two terms with different meanings.

Let's start with the second term $4e^{-\tau_c}(\dot{\phi} + \dot{\psi})$. Since $\tau_c \lesssim 1$ after decoupling, this represents the change in the mode when the photons are free streaming. They are affected by the change in the potentials between here and last scattering. It can be shown that in a matter dominated era $\phi \simeq \psi \simeq$ const. These terms do not change or alter the anisotropies in a fundamental way, but they are necessary for accurate predictions.

The first term is instead the integral of a source weighed by the probability distribution $g(\tau)$ of the last scattering time. We recall that the intensity perturbation $F_{\gamma\ell} = 4\Theta_\ell$ with Θ_ℓ being the perturbation in temperature of the photon distribution. So what appears is actually a term

$$g(\tau)(\Theta_0 + \psi) \tag{13.367}$$

The moment today depends on the monopole *plus* metric potential at recombination. We could have expected this. In fact, any photon we observe on the CMB must have climbed out its potential well to reach us and been Doppler-shifted in the process. ψ plays the role of a Newtonian potential. Repeating: the anisotropies we observe on the CMB are not *only the photon over or under-density at decoupling but the sum of the density and the gravitational potential*. Very curiously, a more detailed analysis of the equations could show that where there is a *matter over-density at decoupling the sum* $\Theta_0 + \psi < 0$! The hot, red, spots on the CMB seen today correspond to locations in the universe with an under-density of matter (mostly dark matter)[80].

The other term proportional to $g(\tau)$, contains the velocity of the baryons. This can be interpreted as a Doppler shift due to the photons being coupled to baryons which have a

non-negligible velocity. As we had shown when discussing the tight coupling approximation, the common baryon and photon velocity (13.300) is out of phase with the photon density oscillations. Of course the tight coupling approximation cannot be valid during decoupling, but this qualitative feature remains true.

We note tha,t for many estimates, it may be a good approximation to think of $g(\tau)$ as a Dirac delta centered around the last scattering surface (the peak, or decoupling)

$$g(\tau) \simeq \delta(\tau - \tau_{LSS}) \tag{13.368}$$

This shows that the CMB anisotropies today can be well estimated through (13.365) by knowing the photon monopole, the metric perturbation and the baryon velocity at recombination. The baryon velocity may also be taken to be similar to the photon dipole term θ_γ.

Finally, let's look at the geometrical term $j_\ell(k(\tau_0 - \tau))$. The spherical Bessel functions of the first kind are oscillating functions. Qualitatively for large enough ℓ, $j_\ell(x)$ peaks at $x \sim \ell$ and then oscillates with decreasing amplitude. Roughly speaking, this creates a "window" making the integral be non-zero only at times $(\tau_0 - \tau) \sim \frac{\ell}{k}$. We can turn this around. Since $g(\tau)$ is very peaked at τ_{LSS}, for a fixed value of ℓ the parts of the integral proportional to $g(\tau)$, thus physically due to decoupling, would only be non-zero for values of the Fourier mode

$$k \sim \frac{\ell}{(\tau_0 - \tau_{LSS})} \tag{13.369}$$

Since $\tau_0 - \tau_{LSS}$ is constant, in order to measure the perturbation with Fourier mode k we would have to measure the right ℓ. This last approximation is also telling us that observing larger values of ℓ means observing larger wave-vectors and therefore smaller wavelengths of the perturbations.

Combining this with our estimation of the peaks of the acoustic oscillations (13.299), we understand that measuring which $F_{\gamma l}$ is largest, as a function of ℓ, is a direct measurement of the sound horizon $r_s(\tau_{LSS})$ at last scattering.

13.17 Non-linearity and reionization

The machinery we have introduced in this section allows us to compute with a high degree of precision the anisotropies in the CMB today. Before concluding, we mention two important issues which must be considered when performing the calculation, in addition to the Boltzmann and Einstein equations: *matter perturbations non-linearity and reionization.*

Radiation perturbations, such as photons and ultra-relativistic neutrinos, cannot collapse and form the gravitationally bound structures, such as clusters and galaxies that we see today. The reason is that ultra-relativistic particles will free stream out of any gravitational well that forms. For this reason, no structure grows during a radiation dominated universe. As the universe becomes matter dominated, the dark matter perturbations will begin to grow. The perturbative under and over-densities $\delta_{c,b}$ which we have considered so far, are the seeds for large scale structure formation. As can be readily understood, the density of any large scale structure is much larger than the average density of the universe. At

some point in time, the perturbations $\delta_{cdm,b}$ grow so large that the perturbative, and linear, approach is no longer valid. We say the matter perturbations are becoming *non-linear.*

Non-linearity means that different Fourier modes couple and the simple decoupled differential equations we have derived are longer valid. A more precise and complex approach is needed. We will not describe it here, but one approach involves running N-body simulations in order to derive the shape of the late time matter perturbations given those at some early time. The relationship is parametrized and a fitting formula is given. This formula can then be used in conjunction with a Boltzmann code to determine the late time form of matter perturbations. A successful approach of this type is the HALOFIT model[207].

In practice, one can use the linear regime for all wavelengths $k \lesssim 0.1hMpc^{-1}$, especially when calculating CMB anisotropies. Shorter wavelengths have had enough time to collapse and enter the non-linear regime. The matter non-linearities can induce small corrections to the final CMB anisotropies.

The second aspect one needs to consider is *reionization[223, 218].* Following decoupling, the universe entered what is known as the *dark ages*[158]. The only radiation in the transparent universe was that of the cosmic microwave background which, at the time, had a spectrum peaked in the red-infrared, quickly redshifting. The subsequent hundreds of millions of years the universe would appear to us humans as very dark. This era lasts until the smallest wavelengths of matter perturbations condense enough to form the first stars, which light up the universe again. The radiation spectrum of stars is nearly thermal, but with a very different temperature than the now redshifted uniform universe. This new radiation will gradually ionize the neutral hydrogen present in the intergalactic void. Such neutral hydrogen forms what is part of the *intergalactic medium* (IGM). The particular process of reionization is not well constrained theoretically nor experimentally. In fact, some constraints come from analysis of the cosmic microwave background itself. Indeed if the IGM becomes ionized, the CMB photons scatter off the IGM and their anisotropy is washed out. Reionization depletes higher modes of the photon anisotropies.

Evidence of reionization comes from the Lyman-α forest[45] and the detection of the Gunn-Peterson trough[119], or lack thereof, in Quasi-Stellar Objects (QSO) at high z. The Lyman-α forest is a series of very close absorption lines observed in the electromagnetic spectra of distant objects due to the photons exciting resonant transitions in neutral hydrogen atoms along its worldline.

As a photon redshifts, it will eventually have the right resonant energy to be absorbed by a neutral hydrogen. A group of photons with different wavelenghts will all be absorbed at different times and the spectra we receive at Earth from distant objects will show a series, the forest, of very close absorption lines throughout. The position of the forest in the spectra depends on the intrinsic temperature of the QSO and its redshift. The strength of the absorption depends on the integrated density of neutral hydrogen along the photon world-line. A simple calculation can show that, if the IGM were completely neutral, the absorption would be complete and spectrum of distant objects would present a trough, known as the Gunn-Peterson trough. For objects at $z \lesssim 6$, the Lyman-α forest is detected but the trough is not present. This implies the IGM, since at least $z < 6$ until today, is nearly completely ionized, the neutral hydrogen fraction being $x_H \sim 10^{-4}$. For objects at $z \gtrsim 6$, the trough is detected, showing that the IGM is neutral at earlier times. It can also be

observed that the trough does not appear everywhere in the universe at the same redshift, implying reionization is a process that happens inhomogeneously as the radiation from stars streams out into the universe, but is pretty much complete by $z \sim 6$. Analysis of the CMB confirms that reionization happened at around those redshifts.

14 CMB Power Spectrum

14.1 Temperature

A great deal of calculatios has gone into finding the Einstein and Boltzmann equations for small perturbations. For photons, we obtained the Boltzmann equations (13.250)-(13.264) which we formally solved through the line of sight integrals (13.348) and (13.356). We now wish to compare these with an observable.

The anisotropies in the Cosmic Microwave Background have been successfully measured by several experiments, the latest and most comprehensive being the PLANCK experiment. The goal of these experiments is to measure the CMB fluctuations across the sky, or a fraction of it. Measurements are done at many frequencies in order to eliminate contamination from astrophysical foregrounds between Earth and the last scattering surface. At each point in the sky the radiation has a blackbody spectrum with some temperature $T_{CMB}(1 + \Theta_\gamma(\hat{n}))$, where T_{CMB} is the average of the temperature across the sky and Θ_γ the fluctuation depending on the position. We want to now connect this temperature fluctuation to the photon modes $F^{(m)}_{\gamma\ell}$ we have defined previously (see (13.232) and (13.342)).

First let's recall that $F^{(m)}_{\gamma\ell}$ represents a fluctuation in intensity of the photon distribution, in fact $F^{(0)}_{\gamma 0} = \delta_\gamma$ is the density fluctuation. Then we may define the temperature fluctuations as

$$\Theta^{(m)}_{\gamma\ell} = \frac{F^{(m)}_{\gamma\ell}}{4} \tag{14.1}$$

Some authors prefer to work from the outset by characterizing the distribution perturbations with the temperature fluctuations instead of intensity, as we have done. The only real difference in the two approaches is the factor of 4.

Now, suppose an experiment has measured the intensity (or equivalently temperature) fluctuations on all the sky. In terms of the variables we have defined in the Boltzmann treatment, what is being measured is

$$F_\gamma(\vec{x}_0, \tau_0, \hat{n}) \tag{14.2}$$

the real space version of $F_\gamma(\vec{k}, \tau_0, \hat{n})$ we have used so extensively. $\vec{x}_0$ is the position of the Earth which we may take to be zero, thanks to the cosmological principle. τ_0 is the time today. Importantly, the direction $\hat{n}$ of the incoming photon in our formalism is the opposite of the direction in the sky being looked at. The measured fluctuations can be expressed by the Fourier transform

$$F_\gamma(\vec{x}_0, \tau_0, \hat{n}) = \int d^3k e^{i\vec{k}\cdot\vec{x}_0} F_\gamma(\vec{k}, \tau_0, \hat{n}) \tag{14.3}$$

Where we kept $\vec{x}_0$, for pedagogical clarity. The fluctuations across the sky can be expressed

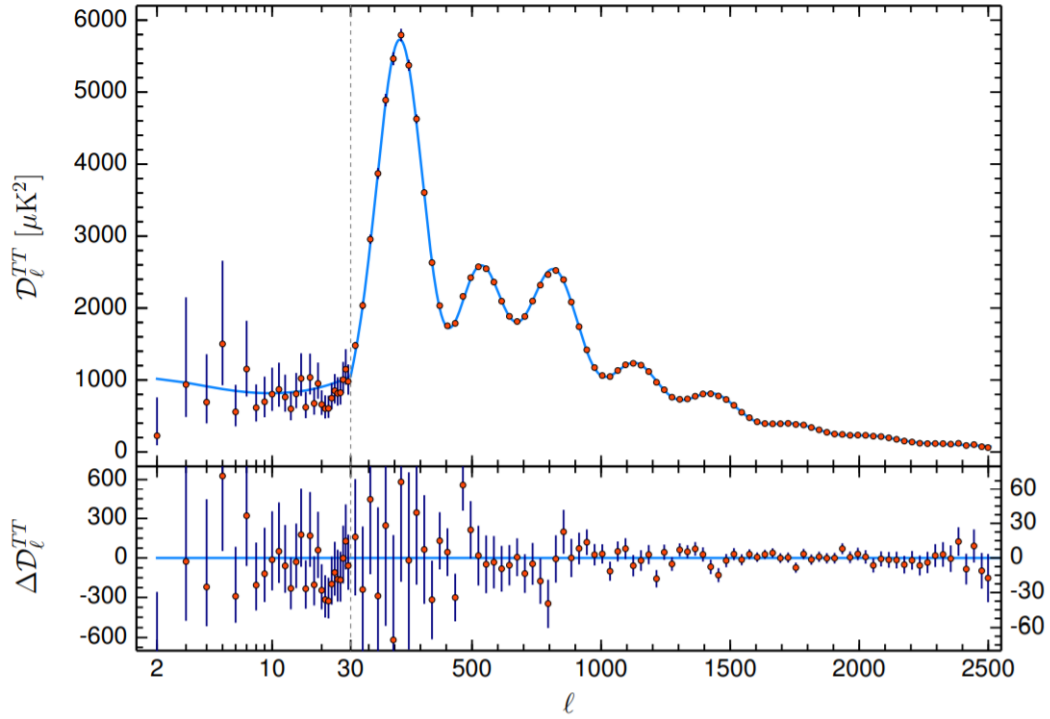

Figure 14.1: CMB temperature power spectrum $D_\ell^{TT} = \frac{1}{2\pi}\ell(\ell+1)C_\ell^{TT}$ reported by the PLANCK experiment. The blue line is the global best fit (including polarization data) and the residuals are shown. The error bars are $\pm 1\sigma$ Gaussian uncertainties. Notice that the scales change at $\ell = 30$. The acoustic oscillations of the plasma are striking. The damping at higher values of ℓ is also present. Figure from ref. [61].

through spherical components

$$F_\gamma(\vec{x}_0, \tau_0, \hat{n}) = \sum_{\ell,m} a^I_{\ell m} Y^m_\ell(\hat{n}) \tag{14.4}$$

where $a^I_{\ell m}$ are the angular coefficients of the intensity fluctuations related to those for temperature by

$$a^T_{\ell m} = \frac{a^I_{\ell m}}{4} \tag{14.5}$$

We recall that the sum on ℓ proceeds from 0 to ∞, while m from $-\ell$ to ℓ.

Now there is a finer point. We recall during our discussion of initial conditions in section 13.15, we had specified how these must be stochastic in nature. At least, experimentally they appear to be. For adiabatic initial conditions we could reduce all the possibilities to a dependence on one single stochastic parameter. We characterized the randomness of this variable by its average (13.333), which is zero, over all possible realizations and its variance (13.334). Indeed, through the variance we defined the primordial power spectrum, for scalar modes,

$$< \psi(\vec{k}, \tau_i)\psi^*(\vec{k}', \tau_i) >= \frac{1}{(2\pi)^3} P_\psi(k)\delta^3(\vec{k} - \vec{k}') \tag{14.6}$$

Since the equations are linear, it implies that $a^I_{\ell m}$ are also stochastic in nature. When we observe the CMB we are sampling one point $\{a_{2m}, a_{3m} \dots\}$ from some distribution. A priori we *may only predict the distribution not the sampled point*. We do this by predicting

statistical propertiesof $a_{\ell m}$.

At first it may seem hopeless. How can we find the full distribution by observing only one sample? Luckily, we may invoke the cosmological principle. As we mentioned when discussing assumptions in section 5, the homogeneity and isotropy of the universe can only be true in a statistical sense. Assuming this we can take

$$a^I_{\ell m} \to a^I_\ell \tag{14.7}$$

The angular coefficients, in a statistical sense, can depend only on the quantum number ℓ and not on m due to isotropy. Indeed, changing the $\hat{z}$ axis rotates spherical harmonics Y^m_ℓ into one another at fixed ℓ. Thus, all the $a_{\ell m}$ at the same value of ℓ may be thought to be sampled from the same distribution. For each value of ℓ, we then have $2\ell+1$ samples. This gives us a fighting chance of getting a handle on the underlying distribution.

For $\ell = 0$, the monopole term, we can still say nothing. Indeed, if the photon monopole were non-zero, it would correspond to a uniform increase or decrease of the CMB temperature across the sky, which we cannot discern from the background value with a measurement at Earth. The dipole terms, $\ell = 1$, are detected because they give a clean angular dependence on the sky. We drop these as well from any CMB analysis. Indeed, a dipole term is obtained by Doppler shift due to the peculiar motion of the Earth (with our solar system and galaxy) with respect to the Hubble flow. Our rest frame is not the CMB's rest frame and the CMB appears blueshifted in the direction we are moving towards it, redshifted in the opposite. Currently our peculiar motion lies towards the direction of the constellation Centaurus. Since the dipole moment appears to be quite larger than other moments and have a consistent velocity, the peculiar motion explanation is usually accepted.

For $\ell \geq 2$, we can legitimately believe to be observing physical anisotropy in the CMB. For smaller values of ℓ especially, the number of samples is $2\ell + 1$. This can be quite small and our ability to reduce the error in observations of statistical quantities is diminished, a fact that is known as *cosmic variance.* This being said, we are now ready to give predictions on $a^{T,I}_{\ell m}$

$$< a^I_{\ell m} >=< a^I_{\ell 0} > \quad = \quad \int d\Omega \int d^3k Y^{0*}_\ell(\hat{n}) < F_\gamma(\vec{k}, \tau_0, \hat{n}) >$$

Through the line of sight integral, and the sources (13.336)-(13.338), we see that $< a^I_{\ell m} >$ is linear in all the perturbation quantities. This means they are linear in the initial stochastic value, whose average is zero. There are no homogeneous terms and therefore

$$< a^{I,T}_{\ell m} >= 0 \tag{14.8}$$

We could have guessed this due to statistical isotropy. From an experimental perspective, checking that these $a_{\ell m}$ average to zero is a test of the cosmological principle. This has been verified experimentally, within error. What we need to predict, which will be non-zero, is the *variance of the distribution* giving the $a_{\ell m}$

$$< a^T_{\ell m} a^{T*}_{\ell' m'} >\equiv \delta_{\ell\ell'}\delta_{mm'} C^{TT}_\ell \tag{14.9}$$

The variance, denoted C_ℓ, is the main quantity obtained from experiment which we wish to confront with theory. Of course, any value C_ℓ^{TT} measured by experiment is an estimate of an underlying variable. Apart from experimental error, one must consider that the estimate has an intrinsic error due to the limited number of $a_{\ell m}$ we may sample. The cosmic variance error is

$$(\frac{\Delta C_\ell}{C_\ell})_{\text{cosmic variance}} = \sqrt{\frac{2}{2\ell+1}} \tag{14.10}$$

We may calculate the C_ℓ^{TT} by

$$\begin{aligned}(2\ell+1)C_\ell^{TT} &= \sum_m < a_{\ell m}^T a_{\ell' m'}^{T*} > \\ &= \sum_m \sum_{M,L} \sum_{M',L'} \int d^3k d^3k' d\Omega d\Omega' Y_\ell^{m*}(\hat{n}) Y_\ell^m(\hat{n}') \times \\ &\quad (-i)^{L-L'} \sqrt{(4\pi)^2(2L+1)(2L'+1)} Y_L^M(\mathcal{R}\hat{n}) Y_{L'}^{M'*}(\mathcal{R}'\hat{n}') \\ &\quad < \Theta_{\gamma L}^{(M)}(k,\tau_0) \Theta_{\gamma L'}^{(M')*}(k',\tau_0) >\end{aligned}$$

We have expanded the temperature into its angular modes. There is technical point here. The angular dependence of the photon distribution was defined by taking $\hat{z} \parallel \vec{k}$. This was indeed how we were able to separate out scalar and tensor modes. On the other hand the $a_{\ell m}$ are defined with respect to a fixed z axis on the sky. In the angular expansion of Θ_γ we have $\mathcal{R}\hat{n}$, where $\mathcal{R}$ is the rotation operator which rotates $\hat{k}$ into $\hat{z}$ and $\mathcal{R}'$ rotates $\hat{k}'$ into $\hat{z}$. But this is not a problem. Indeed the spherical harmonics transform in a $2L+1$ dimensional irreducible representation of the rotation group, thus

$$Y_L^M(\mathcal{R}\hat{n}) = \sum_K D_{M,K}^{(L)}(\mathcal{R})^* Y_L^K(\hat{n}) \tag{14.11}$$

where $D_{M,K}^{(L)}$ is a $(2L+1)\times(2L+1)$ unitary matrix realizing the rotation $\mathcal{R}$, known as the Wigner D-matrix. The exact form is not needed. Then the angular part of the integral is

$$\begin{aligned}\sum_m \int d\Omega d\Omega' Y_\ell^{m*}(\hat{n}) Y_\ell^m(\hat{n}') Y_L^M(\mathcal{R}\hat{n}) Y_{L'}^{M'}(\mathcal{R}'\hat{n}') &= \sum_m \sum_{K,K'} \int d\Omega d\Omega' Y_\ell^{m*}(\hat{n}) Y_\ell^m(\hat{n}') \\ &\quad \times D_{MK}^{(L)}(\mathcal{R}) D_{M'K'}^{(L')}(\mathcal{R}') Y_L^K(\hat{n}) Y_{L'}^{K'}(\hat{n}') \\ &= \sum_m D_{Mm}^{(\ell)}(\mathcal{R})^* D_{M'm}^{(\ell)}(\mathcal{R}') \delta_{\ell,L}\delta_{\ell,L'} \\ &= D_{MM'}^{(\ell)}(\mathcal{R}' \circ \mathcal{R}^{-1}) \delta_{\ell,L}\delta_{\ell,L'}\end{aligned} \tag{14.12}$$

Where we used the orthogonality relation for spherical harmonics to get the second line. In the third line, we used the fact that $D_{Mm}^{(\ell)}$ is a representation of the rotation group: the matrix product gives the element which represents the product of the elements of the group. With this we obtain

$$(2\ell+1)C_\ell^{TT} = 4\pi(2\ell+1) \sum_{M,M'} \int d^3k d^3k' < \Theta_{\gamma\ell}^{(M)}(k,\tau_0) \Theta_{\gamma,\ell}^{(M')*}(k,\tau_0) > D_{MM'}^{(\ell)}(\mathcal{R}' \circ \mathcal{R}^{-1}) \tag{14.13}$$

Next we recall that, with adiabatic initial conditions for scalar modes, we may solve the equations with initial conditions given by (13.325)-(13.329) and $\psi(\vec{k},\tau_i) = 1$ and call that solution $\Delta_L^{(0)}(k,\tau)$. The same can be done for tensor modes, where one selects $h_{+,\times}(\vec{k},\tau_i) = 1$, assuming the initial conditions of both terms would be the same. The real solution is, due to the linearity of the Boltzmann equations,

$$\Theta_{\gamma\ell}^{(0)}(k,\tau) = \Delta_\ell^{T(0)}(k,\tau)\psi(\vec{k},\tau_i) \tag{14.14}$$

$$\Theta_{\gamma\ell}^{(\pm 2)} = \Delta_\ell^{T(\pm 2)}(k,\tau)h(\vec{k},\tau_i) \tag{14.15}$$

$\Delta_L^{T(M)}$ is the *transfer function*. We can neatly separate the initial condition from the subsequent evolution. Next, we notice that all the randomness is in the initial conditions $\psi(\vec{k},\tau_i), h_{+,\times}(\vec{k},\tau_i)$ and not in the transfer function whose form is absolutely deterministic. We separate out scalar and tensor modes

$$\begin{aligned} C_\ell^{TT} &= 4\pi\sum_{M'}\int d^3k d^3k' \left(\Delta_\ell^{T(0)}\right)^2 D_{0M'}^{(\ell)}(\mathcal{R}'\circ\mathcal{R}^{-1}) < \psi(\vec{k},\tau_i)\psi^*(\vec{k},\tau_i) > + \\ &+4\pi\sum_{M=\{-2,2\}}\sum_{M'}\int d^3k d^3k' \left(\Delta_\ell^{T(M)}\right)^2 D_{MM'}^{(\ell)}(\mathcal{R}'\circ\mathcal{R}^{-1}) < h(\vec{k},\tau_i)h^*(\vec{k},\tau_i) > \end{aligned}$$

All the stochastic part of this is in the variance of the initial condition, which we had defined in (14.6). Using the Delta function to fix $\mathcal{R} = \mathcal{R}'$, the rotation matrix becomes the identity. Writing $P_0(k) = P_\psi(k)$ and $P_{\pm 2}(k) = P_h(k) = (2\pi)^3 < h_x(\vec{k},\tau_i)h_x^*(\vec{k},\tau_i) >$ with $x = +,\times$,

$$C_\ell^{TT} = \frac{1}{2\pi^2}\sum_M \int d^3k \left(\Delta_\ell^{T(M)}(k,\tau_0)\right)^2 P_M(k) \tag{14.16}$$

The angular integral on k can be performed

$$C_\ell^{TT} = \frac{2}{\pi}\sum_M \int k^2 dk \left(\Delta_\ell^{T(M)}(k,\tau_0)\right)^2 P_M(k) \tag{14.17}$$

This is our final form for C_ℓ^{TT}. $\Delta_\ell^{(M)}$ can be of course expressed via the line of sight integral and is calculated numerically. The C_ℓ^{TT} are measured directly from experiment. The final formula contains a sum over scalar, vector and tensor modes. When measuring the temperature anisotropies today all the modes contribute. For temperature C_ℓ^{TT} the scalar modes dominate. Taking $M = 0$ is enough to get the correct values in most situations.

The C_ℓ^{TT} is also known as the temperature self-correlation. It encodes how a fluctuation in temperature at one point on the sky depends statistically on some other point.

Finally, we note that the power spectrum we have defined is dimensionless, as it is the power spectrum of *fractional* temperature fluctuations $\frac{\delta T}{T}$. In the literature results are often quoted with the power spectrum of the temperature fluctuations. The two are related as $C_\ell^{\text{Fractional}} = C_\ell^{\text{absolute}} T_{CMB}^2$. So when looking at any plot of the power spectrum it is important to take note of the dimensions, which indicate which quantity is being used.

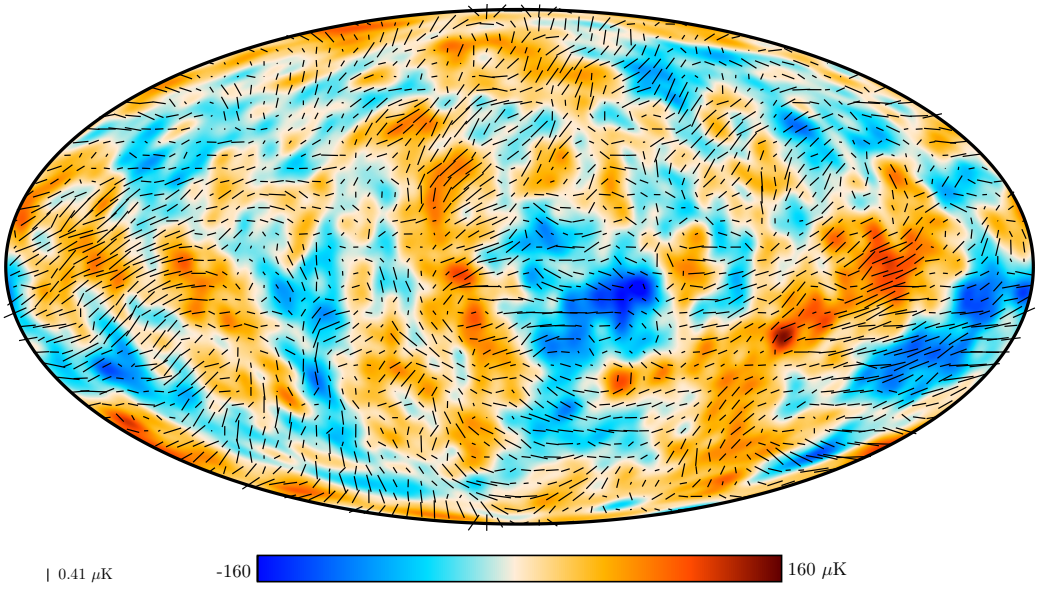

Figure 14.2: Reconstructed polarization map of the CMB from the PLANCK mission (2018). The background coloring is the temperature anisotropy, smoothed by five degrees. Polarization is indicated by rods whose direction and length represent the direction and amplitude of the polarization. Figure from ref. [57].

14.2 Polarization and Cross-Correlation

The CMB is polarized and the polarization varies in intensity and direction across the sky. We must express these fluctuations in a simple way to confront theory and experiment. The same ideas we used when discussing the temperature self-correlation C_ℓ^{TT} apply for polarization. Particularly that the fluctuations are stochastic in nature and we only observe one sample of a distribution.

Experimentally, the detected polarization is described by the Stoke parameters $Q(\hat{n})$, $U(\hat{n})$ which vary across the sky, in some arbitrary basis that is convenient for the experiment. We had discussed that the dependence on Q and U on an arbitrary basis is not the easiest way to go about. Rather, we notice that $(Q \pm iU)$ are spin-2 (-2) quantities, see equation (13.188). The angular dependence on the direction on the sky may be expanded in terms of spin-weighted spherical harmonics ${}_{\pm 2}Y_\ell^m(\hat{n})$[135, 136]

$$(Q \pm iU)(\hat{n}) = \sum_{\ell m} (a_{\ell m}^E \pm i a_{\ell m}^B)\, {}_{\pm 2}Y_\ell^m(\hat{n}) \tag{14.18}$$

The experimentally measured $(Q \pm iU)(\hat{n})$ is none other than $F_\pm(\vec{x}_0, \tau_0, \hat{n})$, the Fourier transform of the photon polarization distribution defined in 13.233 whose spherical expansion is given by (13.240) and which we can express through line of sight integrals(13.356).

It should not shock anyone that

$$< a_{\ell m}^{E,B} >= 0 \tag{14.19}$$

so we move on to calculate the variance

$$< a_{\ell m}^X a_{\ell' m'}^{Y*} > \equiv \delta_{\ell,\ell'} \delta_{m,m'} C_\ell^{XY} \tag{14.20}$$

Where $X, Y = E, B, T$. Indeed, not only way may look at the form of polarization fluctuations across the sky, but at how they correlate between types and with the temperature. Projecting onto spherical harmonics, and noting that the spin-weighted spherical harmonics transform in some representation of rotations, just as regular spherical harmonics, within the same value of ℓ, s , we arrive, in the same manner as we did to get (14.17), to the more general formula

$$C_\ell^{XY} = \frac{2}{\pi} \sum_M \int k^2 dk \Delta_\ell^{X(M)} \Delta_\ell^{Y(M)} P_M(k) \tag{14.21}$$

As we had pointed out in equations (13.199) and (13.200) the $a_{\ell m}^{E,B,T}$ don't have the same parity. Under a parity transformation

$$a_{\ell m}^{E,T} \to (-1)^\ell a_{\ell m}^{E,T} \tag{14.22}$$

$$a_{\ell m}^{B} \to (-1)^{\ell+1} a_{\ell m}^{B} \tag{14.23}$$

This implies that $< a_{\ell m}^{T,E} a_{\ell m}^{B*} >= 0$ since these parity transformations imply $< a_{\ell m}^{TE} a_{\ell m}^{B*} >= - < a_{\ell m}^{TE} a_{\ell m}^{B*} >$. Indeed this is the case experimentally. In the standard cosmology there is no cross-correlation of T and E with the B modes.

It is slightly more subtle than this. The derivation of the transformation under a parity operator $\hat{\mathcal{P}}$ assumed that the action of this operator would simply be

$$\hat{\mathcal{P}} F_{\gamma,\pm}(\hat{n}) = F_{\gamma,\pm}(-\hat{n}) \tag{14.24}$$

or a simple reversion of the spatial axis. This assumes that the physics that generates $F_{\gamma,\pm}(\hat{n})$ and $\hat{\mathcal{P}} F_{\gamma,\pm}(\hat{n})$ is the same. We know the standard model badly violates parity. Therefore there is a priori no guarantee that (14.24) holds, because a parity transformation involves not only changing the shape of the sky today, but flipping the axis throughout the history of the universe which would have generated a different distribution. This being said, in standard cosmology the only parity violations may come from neutrinos. The existence of parity violating interactions is not enough to generate a cosmological signal. There are subtleties, but at the very least there must be some breaking of equilibrium between left and right neutrinos. Nothing of the sort happens and we are left with no large scale parity violations.

14.3 The primordial power spectrum

The CMB anisotropies today are the evolution of the primordial power spectrum $P_\psi(k)$ for the scalar modes and $P_h(k)$ for the tensor modes. Scalar modes have been measured and the form of $P_\psi(k)$ is known from experiment. Tensor modes have not yet been measured. The physical origin of both power spectra is unconfirmed. The most well known hypothesis for their origin is from the theory of inflation. Certain models of inflation are known to produce primordial spectra consistent with observations. In the context of inflation, the stochastic spectrum is none-other than quantum fluctuations of the field, or fields, driving

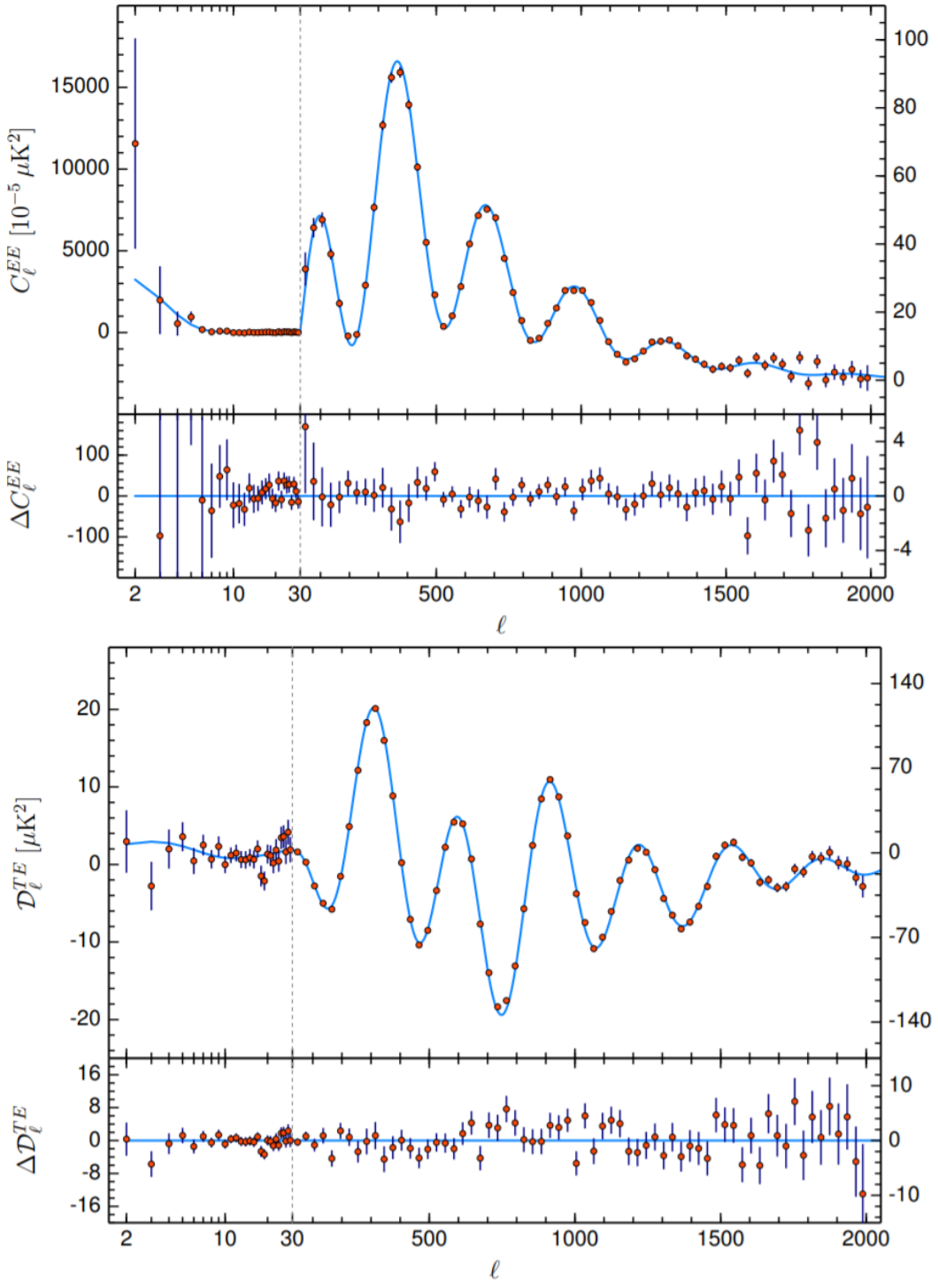

Figure 14.3: CMB polarization power spectrum and polarization-temperature cross correlation $D_\ell^{XX} = \frac{1}{2\pi}\ell(\ell+1)C_\ell^{XX}$ reported by the PLANCK experiment. The blue line is the global best fit (including polarization data) and the residuals are shown. The error bars represent $\pm 1\sigma$ Gaussian uncertainties. The figures are from ref. [61].

inflation which have become classical. This connects the stochastic nature of cosmological perturbations with the well known probabilistic interpretation of quantum mechanics.

In the literature, the spectrum for scalar modes is defined through the *curvature perturbation* $\mathcal{R}$. To define this, we gauge transform to a comoving metric, where there is no heat flow $T_i^0 = 0$ and take $\mathcal{R}$ to be the curvature of the spatial slicing.

It can be shown that the curvature perturbation is conserved when the mode is outside the horizon. Using curvature perturbations allows to connect perturbations generated by inflation with those that we use as initial conditions in the calculation of cosmological perturbations.

Let's discuss the primordial power spectrum dependence on k. Inflation predicts, and this is verified by experiment, that the power spectrum is nearly scale invariant, that is

$$P_{\mathcal{R}}(k) \propto \frac{1}{k^3} \tag{14.25}$$

There is always one term k^{-3}, which is why some cosmologists define $P_{\mathcal{R}}(k)$ with this factor explicitly. This is due to the dimensionality of k. In the absence of any physical scale, this can be the only possible form of the primordial power spectrum by dimensional analysis. We only would need to fix the normalization. We cannot say that this is a smoking gun of inflation for this reason.

What is interesting is that there actually is a scale dependence. Let's parametrize the power spectrum as

$$P_{\mathcal{R}}(k) = (2\pi)^3 k^{-3} A_s (\frac{k}{k_0})^{n(k)} \tag{14.26}$$

$$n(k) = (n_s - 1) + \frac{1}{2}\frac{dn}{d\ln k}|_{k_0} \ln\frac{k}{k_0} + o(\ln\frac{k}{k_0})^2 \tag{14.27}$$

k_0 is an arbitrary parameter usually chosen to be $k_0 = 0.05 Mpc^{-1}$, in fact changing k_0 implies a reparametrization of n_s and $\frac{dn}{d\ln k}|_{k_0}$ from the fitting of experimental data. These are the interesting experimental parameters. n_s is called the index of scalar modes and is very close but not exactly equal to 1, which would imply a scale invariant spectrum. It is a constant. In fact, we have Taylor expanded the exponent in powers of $\ln\frac{k}{k_0}$ with only the two terms being experimentally relevant. Indeed, the first order coefficient $\frac{dn}{d\ln k}|_{k_0}$ is experimentally consistent with zero. It is known as the running of the scalar index. Finally, we have the amplitude of the scalar modes A_s.

The primordial spectrum for the tensor modes is parametrized as

$$P_h(k) = (2\pi)^3 k^{n_T - 3} r A_s \tag{14.28}$$

where n_T is the index of the tensor modes. Note it is defined so that $n_T = 0$ indicates a scale invariant spectrum. r is a constant, and is the ratio of scalar to tensor modes. r is thought to be small, perhaps $r \sim 10^{-3}$. Indeed, tensor modes have not yet been measured.

14.4 ΛCDM Parametrization

We have written extensively about the ΛCDM model. Much has been understood from first principles, however we need to fix some parameters from experiment to calculate quanti-

ties in the model. These parameters, such as the Hubble factor today, cannot be calculated a priori. In practice, cosmological observations such as analysis of the CMB anisotropies allow a very good determination of these parameters.

There are a few subtleties to discuss. We will eventually want to use the measured anisotropies of the CMB, or other observables, to determine the values of the cosmological parameters of the model and their error bars. Although the physics is of course independent of any parametrization, this process will be very much so dependent on it. To understand, let's try to think of a flat universe with dark energy, dark matter and radiation. We'd like to determine the values of their density and of the Hubble factor by some measurement. The four parameters $\Omega_{\Lambda 0}, \Omega_{m0}, \Omega_{r0}$ and H_0 are related to each other by the flatness condition and the Friedmann equations, so in any analysis we must choose three parameters and take the fourth as a dependent quantity. Suppose we choose $\Omega_{\Lambda 0}, H_0$ and Ω_{m0}. When performing an analysis, we would input different values of the dark energy density to understand its effect on the universe, while keeping the others constant. However, because of the flatness condition, changing $\Omega_{\Lambda 0}$ implies changing Ω_{r0} which would have the drastic effect of changing the epoch of matter-radiation equality a_{eq}, changing the form of the CMB anisotropies in a specific way which is not really related to the dark energy density. Thus, our understanding of dark energy would be obfuscated. On the other hand if we had chosen H_0 as the dependent parameter, we would not see this effect. All this would have an effect of changing the error bars that result from the analysis.

Another issue is what parameters to consider as fixed. In our previous example the curvature parameter k was assumed to be fixed to zero. If we didn't do this, our analysis would be altered. There would be an error bar on k, but more importantly, the error bars on the other parameters would be likely to get larger. This brings us to the concept of model dependence. Any analysis and parameter estimation we perform is dependent on the validity of the model and what assumptions we use. For example, if we suppose neutrinos are massless when they actually are not, the inferred values of our parameters will be biased.

The "base" ΛCDM model is considered to have the parameters

- $\Omega_{cdm0}h^2$, $\Omega_{b0}h^2$, $\Omega_{r0}h^2$, the cold dark matter, baryon and radiation density today. $H_0 = h \cdot 100\frac{km}{s \cdot Mpc}$.
- $\ln(10^{10}A_s)$, the scalar mode primordial spectrum amplitude.
- n_s the scalar mode primordial spectrum index.

In addition, some technical parameters are used to fit the data, for example the optical depth $\tau_{\text{reioniation}}$ to re-ionization. Other parameters may be added, otherwise their baseline values are used. These are values assumed to be true from other experiments and theory. Common parameters that are fixed but may be allowed to vary in extended models are the neutrino masses, the effective number of standard model neutrinos, the curvature and the dark energy equation of state parameters w.

Part IV

Numerical methods in cosmology

15 CLASS Boltzmann Code

15.1 General philosophy

CLASS[26, 148, 8, 208, 147] is a cosmological Boltzmann code in C developed and maintained in the last decade. It is meant to be an alternative to previous codes CMBFAST[222], CMBEASY[82] and CAMB[151]. As of today, only CAMB and CLASS are regularly maintained codes. CAMB is written in Fortran. These codes calculate CMB anisotropies and other cosmological observables from first principles using the equations we have described in the previous chapter. The self-stated goal of the developers was that CLASS be user-friendly, flexible and accurate.

Accuracy is the first issue. We compare experimental data to theoretical predictions made by a highly complex numerical code. A mistake in the code, or a numerical error not under control, may affect the final result of any calculation. If this is not under control, it may bias the conclusions we make about inferred parameters. In science, redundancy is key. Independent scientists must make there own predictions, with independent codes, and, hopefully, agree on the results. Thus, CLASS filled a need to check that the results from CAMB were indeed reliable and not biased. The accuracy of these codes is claimed to be $\sim 0.1\%$ on the CMB anisotropies. Indeed, the two codes are known to agree.

User-friendliness and flexibility are more arbitrarily defined concepts. They become essential in this era of precision cosmology. We are now at a stage where the experimental data is very precise and the theory behind it is reliably understood. So much understood that we now want to use this data to study extensions to standard cosmology. This involves editing code. To stimulate research it must be easy to edit code and change the underlying equations. This is where the concept of user-friendliness and flexibility come in. By user-friendliness the developers refer to the fact that compiling and running the code is simple and straightforward, even with non-standard cosmologies. A large number of parametrizations are supported, with the software doing an internal conversion to a convenient one. On the other hand, flexibility refers to the developer side. It must be easy for newcomers to edit the code and implement new physics. The use of legacy code encourages bad programming habits and the continual use of old, uncommented and obscure, code. In this sense, CLASS has the advantage of being developed from scratch during a time when best programming practices are much more known, resulting in a cleaner code. CLASS also has the advantage, of course, of being developed after previous attempts and therefore can use the wisdom acquired by the community in the meantime.

A fundamental design decision is to avoid any hard-coding. Not only must parameters be user-supplied but assumptions must as well. In this manner, new developers don't have to worry about the validity of any piece of code, as what assumptions go into it are clearly specified throughout. In particular, indexes of matrices and other quantities are determined dynamically at run-time. This allows developers to add to a matrix or table

without breaking any other code.

15.2 Program structure

CLASS implements all the zero-order and first-order equations needed to describe the universe and its perturbations. The goal is to reliably calculate the CMB spectrum and other cosmological observables in a given model and with given parameters. The files are separated into those implementing the physics (Boltzmann, Einstein equations, recombination, thermodynamics, etc.) and the mathematical tools. These mathematical tools are stored in the folder **/tools**. They contain algorithms to solve ordinary differential equations , such as a fourth order Runge-Kutta, perform integration and interpolations. Interpolations are much used in the code, as many quantities depending on conformal time τ and Fourier component k are calculated at discrete intervals.

The physics code is stored in either the **/source** folder or the **/include** folder, which holds the needed definitions. We will focus on this code. The code is implemented through *modules* which are defined in a header file, in the include folder, and implemented in a *.c* file, in the source folder. Each module has a related *struct* containing all the persistent information it calculates.

Every module is called by the ***main*** function, implemented in **/main/class.c**. A module is called by a function *module_init*, where the particular name depends on the name of the module. The module takes as input pointers to all the *structs* related to other modules evaluated previously. Importantly, the function *module_init* initializes all necessary memory and does the physical evaluations. Thus, the name is misleading, as the function not only initializes but runs all the physics in the module. Before quitting the program, or when the module is no longer needed, a function *module_free* , the exact name depending on the module, deallocates all the memory of the relative struct.

Most modules depend on the structs of other modules, therefore they must be initialized and run in order. Care must be taken before freeing any module, as its memory may be needed by successive modules. In practice there is little reason for a developer to add another module. Rather, they would modify an existing one.

Each module performs a specific task which can be inferred from its name. The modules are *input, background, thermodynamics, perturbations, bessel, transfer, primordial, spectra, nonlinear, lensing* and *output*. We will review the specific workings of these in the following sections. It is not required to run every module. Indeed the user may specify which modules to run or not, if only a limited functionality is needed.

CLASS implements a system which reminds of C++ exceptions for error handling. Every function call in CLASS is wrapped by a macro *class_call* which takes as input the function to be called and pointers to error messages. Every function returns a code indicating error or success. The macro checks for this return value and, on error, formats an error message returning another code to terminate program. In this way, the process implements a *stack unwinding*. All this process is taken care inside the macro. Unfortunately, macros only serve to obfuscate code and can be extremely frustrating to debug if something goes wrong since, due to the pre-processor, the error message will not necessarily indicate the line where the error occurs. In addition, this system limits functions to only return error codes, leaving the

return of actual values via pointers passed to functions. This can allow further problems in development to crop up.

One issue is that of memory ownership of the pointers, which the developers do not give a consistent paradigm to deal with. Invalid pointer issues are notoriously hard to debug. The problem of ownership of pointers arises when it is not clear which part of the program must allocate a pointer and which must de-allocate them, leading to different developers possibly using an invalid pointer, or causing a memory leak. The problem is such well known that the newer versions of C++, such as *C++17*, support *smart pointers* which require the ownership to be clearly known to the developer.

CLASS also extensively uses macros in other ways, for example to format and write to file. There is no need for this. A design improvement to CLASS would be to replace most macros with actual functions.

CLASS uses the OpenMP library for multi-threading. Since the differential equations to be solved decouple for various value of the Fourier mode k , and a large number of different modes must be calculated, it makes sense to parallelize the computing to use as many cores as possible. Speed is indeed important in cosmological numerical calculations since parameter estimation, for example using a Markov-Chain MonteCarlo, requires a great many number of runs of the code with different parameters.

15.3 Input module

The first module to be called is the input module. It is different than the others as it does not have a relative *struct*. This module will initialize all parameters to default values and then set any user-specified parameters to its correct value. There are two ways to initialize this module, which must be called before any other.

The function *input_init_from_arguments* opens an initialization file **<NAME>.ini** and a precision file **<NAME>.pre** . The first contains all the physics parameters, and others related to what must be calculated, as well as the verbosity. The second contains parameters which influence the precision of the calculation. In practice, any parameter can be put in any file, as the software will read the files together. In the standard CLASS download, a template initialization file is provided with explanations to the user of the usage of most parameters. *input_init_from_arguments* parses the initialization file and generates a *file_content* struct which is passed to *input_init*. This function can be called directly, without calling *input_init_from_arguments* first, as long as a *file_content* struct is passed. This allows CLASS to be easily initialized and used from within a larger software. Default parameters are set in *input_default_params*.

Once the parameters are passed to *input_init*, CLASS will process the parameters in order to select the most convenient parametrization of the cosmological model, which may not be what the user provided. Thus, the input module already contains some notions of cosmology and does some processing on the parameters. For the most part this is trivial. For example. it transforms input parameters Ω_{cdm} and H_0 to $\Omega_{cdm}h^2$. For some inputs there are complications.

Some of the input parameters refer to quantities which are experimentally easily accessible today but whose dependence on the actual initial conditions of the evolution equations is

not straightforward. CLASS needs some initial conditions to solve the differential equations. If the user specifies a parameter evaluated today, CLASS must figure out which initial conditions are required to get the correct parameter today. For the density of matter or radiation, this is analytical, since the densities scale as a^{-3}, a^{-4}. For other parameters, an analytical link is not known. One example, whose physics we will study in detail later, is the initial value of a cosmological scalar field. A scalar field will oscillate during the history of the universe and the scaling of its density is not a simple function of a. Indeed, the density of the evolution must be solved numerically given some initial value. But what we would specify is the density today, dictated by some experimental observation or theory.

To solve this problem, CLASS implements a *shooting algorithm*. In essence, it will take an educated guess for the initial field value and then *run the background module* with this parameter. It compares the value of the density today that is obtained by solving the equations and compares it to the user specified value. If they agree within a tolerance, CLASS will proceed with the rest of the program. Otherwise, it will change the value of the initial value, stepping in some direction, and trying again. It repeats this, decreasing its step until either it converges or it does not manage to find a solution quickly. If necessary, CLASS will also run other modules, such as the thermodynamics or perturbation module.

In practice, we have found it hard to use the CLASS shooting mechanism for a scalar field, especially when changing the potential to accommodate our needs. The initial guess is not good enough and convergence does not happen. For our purposes, the scalar field density today can be found given the initial value at some very early time and solving the Friedmann equation coupled with a Klein-Gordon-like equation for the scalar field. This is very easy to implement on any software, and we did. Allowing us to always know the correct initial value needed and providing, by hand, this value directly to CLASS.

We edited the input module to read from the initialization file all the parameters relating to a cosmological scalar field which causes cosmological birefringence.

15.4 Background module

The purpose of the background module is to solve the zero order, homogeneous, equations of the universe. In a standard cosmology, this is the Friedmann equation. Extended models are solved as well, if the user needs. Indeed in this work we used extensively the equations for a cosmological scalar field. A cosmological scalar field satisfies a Klein-Gordon like equation which must be solved (equation (18.81)). The density of the scalar field (equation (18.82)) then participates in the Friedmann equations as usual.

The information generated and needed by the module, is stored in the **background** struct defined in **include/background.h**. The struct contains all the cosmological parameters that were set, either by the user or to default values, in the input module. It contains a series of variables, denoted as **index_bg_XXX** and **index_bi_XXX**. These are the dynamically allocated indexes of the background module. Indeed all the background values which are calculated in this module will be time dependent and stored in the variable **background_table** of the struct. This is a rectangular array where the rows correspond to a time step and the columns to the background variables. The column relative to quantity **XXX** is the **index_bg_XXX** column. The **index_bi_XXX** quantities are in some cases duplicates of the

index_bg_XXX quantities and they are not stored in this table. Rather, those indexes refer to the quantities that are integrated over. Most background quantities can be in fact inferred from the scale factor. Other two arrays are **tau_table** and **z_table**, which contain the time steps. We note that all times are conformal times. Time step i will have conformal time **tau_table[i]** and redshift **z_table[i]**. All these quantities are calculated by the module.

As we pointed out in section 15.2, calling *background_init* actually runs the module in full and solves the background equations, allocating all the memory needed and storing the found values in the aforementioned tables. The function *background_solve* is called, which deals with solving the actual differential equations. This function first calls for all initial conditions to be set in *background_initial_conditions*. That is a possible place to start editing if one wants to change some physics. However, note that the philosophy of CLASS is never to hardcode values or physical equations and assumptions.

After setting the initial conditions and initializing the memory of arrays to store values, the integration begins shortly after in a while loop. CLASS integrates from a very small scale factor $a \sim 10^{-14}$ deep in the radiation era and when all perturbation modes of interest can be considered to be out of the horizon. The time step is chosen dynamically based on the Hubble factor, since H^{-1} is the typical time. A fraction of this typical time, which is user specified, is used. In extended models with a scalar field which oscillates, a check should be added to ensure that the time step is shorter than the period of oscillation.

The actual integration happens inside the call to **generic_integrator**. This is actually a variable describing the integration function, in order for the user to easily change what numerical ODE solver is used. We have used a fourth-order Runge-Kutta algorithm. *generic_integrator* receives as arguments the start and end of the time step and a pointer to the function *background_derivs.*

background_derivs is where the physics happens. Here, the conformal time τ and current background quantities under integration are passed, and their derivatives calculated and returned via pointer to array.

When the integration is complete, all the necessary quantities are stored in the aforementioned tables. Indeed, using the function *background_function* allows to calculate all the background variables based on the few ones that are integrated, at any time.

There exists two functions which are useful to be called from other modules. These are *background_tau_at_z* and *background_at_tau*. The first simply returns the value of conformal time given the redshift. Indeed, our variable of integration is the conformal time but in many applications the redshift is a more natural variable. That is why we need to be able to convert quickly if needed. The second function, *background_at_tau*, returns a vector (an array) of background quantities at a time τ. The quantities returned are interpolated between those at the closest sampled timestep. For speed issues, an integer value *return_format* can be passed specifying how many background quantities are needed to be returned, if only the essential or a full set.

15.5 Thermodynamics module

The thermodynamics module solves the equations for recombination and re-ionization. It calculates several thermodynamical quantities necessary for use in the later modules, such

as the free electron density and the visibility function. For recombination, the RECFAST code is used. This is Peebles' two hydrogen levels with fudge factors which we mentioned in section 11. CLASS also allows the user to use custom recombination code, since a more precise recombination model may be needed for certain applications. For re-ionization, since there is no precise model and the CMB only depends on the integral of the free electron density on the line of sight (the optical depth), a reionization function is used, based on empirical observations. This same re-ionization function for the free electron density is used in the Boltzmann code CAMB.

Recombination requires the value of the primordial Helium fraction X_{He} (some texts write this as Y_{He}). The user may either fix this, or indicate he wishes it to be extrapolated from a BBN code. A pre-computed table from the BBN code Parthenope[67] is used for interpolation, based on other physical parameters. The primordial Helium fraction from BBN in fact depends on a few parameters, such as the effective number of neutrinos N_{eff} and the baryon density $\Omega_{b0}h^2$.

The calculated quantities are stored in the thermodynamic structs, namely in the **thermodynamics_table** which is , like the **background_table** of the background module, a rectangular array where the rows correspond to a step in redshift z and the columns to the parameters, whose dynamical indexes are indicated as **index_th_XXX**. In addition, values such **z_rec** and **tau_rec** are stored indicating the values of redshift and conformal time at which the visibility function $g(\tau)$ peaks.

Once *thermodynamics_init* is called, all the quantities are calculated. Then one can use *thermodynamics_at_z* to obtain interpolated values of the quantities at some redshift z.

In studying the dynamics of a scalar field which induces cosmic birefringence, we must calculate the uniform rotation across the sky $\bar{\alpha}$, as defined in (19.19). This uniform rotation value depends only on the background and thermodynamic quantities, so it makes sense, from a software structure point of view, to calculate it and store it as "thermodynamic" quantity. We add indexes, and therefore space, in the thermodynamics table to store the uniform angle as a function of z. We then define new methods in which the uniform angle is calculated. This is an integration from the initial time to the current conformal time $\tau(z)$. We use the numerical methods that come pre-packaged with CLASS to perform the integral. The methods are called at the end of *thermodynamics_init* when all other thermodynamical quantities have been calculated. Indeed, since other quantities don't depend on this angle it is simpler, if perhaps not most efficient, to separate out the new code from the baseline code.

15.6 Perturbation module

The perturbation module solves all the Boltzmann and Einstein equations relating to the perturbations of all the species. Baryons, photons, neutrino and metric terms are tracked. In addition, the module supports equations for non-standard cosmologies, for example massive neutrinos, non-standard perfect fluids and scalar fields. Not all the equations are solved and the variables tracked. CLASS makes a determination about what to solve for based on what observables and physics the user has asked for in the initialization file. The equations can be solved in either the synchronous or conformal-Newtonian gauge. Which

can be specified by the user in the initialization file.

The CMB power spectrum C_ℓ^{XX} are expressed through a line of sight integral (13.348). A source is convolved with some radial function related to the spherical Bessel function. The perturbation module does not do this convolution, leaving it for later. Rather, it solves all the Boltzmann equations and saves the sources that are needed for the *transfer module*.

Another aspect that goes into the computation are the various approximations which are used by the code. For example the tight coupling approximation we described in section 13.14. Other approximations are listed in **include/perturbations.h** and a full description is given in ref. [26]. Not all the approximations are for standard cosmology. During the numerical integration, the code will continually check which approximation it should be using and will switch between equations to be used. This is actually quite important for the numerical success of the software. Indeed, the tight coupling approximation is required because the derivative of the optical depth, $\dot{\tau}_c$, is quite large at early times, making the derivative of the photon and baryon velocity possibly quite large as well. If the tight-coupling approximation is not used, a very small time step would be needed to integrate properly. This would increase integration time, in a moment where we know the equations are actually simpler (in fact we had found an approximate analytical solution (13.296)).

Let's look at the perturbation struct. The first variables are boolean, named **has_**, allowing the module to decide what equations and variables to track. These have mostly been set back in the input module. The next important variables to take note of are the integers **l_scalar_max**, **l_vector_max** and **l_tensor_max**. These are specified by the user and are telling us the maximum number of ℓ we want to calculate the C_ℓ^{XX} of the CMB for. As we had remarked at the end of section 13.16 the Fourier mode k that contributes to a C_ℓ is roughly proportional to ℓ. Thus, these values are indicating the largest value of the Fourier components the code must solve for. Higher values increase calculation time.

Next we find switch variables **switch_sw, switch_eisw** and so forth. As we saw in the line of sight integral for the higher photon modes, the source is a sum of different terms. Although we physically can only measure the sum, CLASS allows to switch off any term of the sum. These switches are set by the user in the input function. The usefulness of being able to look at each term separately is manifest when we attempt to understand exotic models, as the new physics may alter only one of the terms.

The indexes for the various perturbation terms are indicated as **index_tp_XXX**.

There are arrays for the k sampling, which are allowed to be different for different modes(scalar-vector-tensor). These are in the array ****k** whose size whose first index is the actual mode and the second the value of k. So it is actually three arrays in one(one per SVT). The size of these arrays is given in **int *k_size**.

Towards the end of the struct there an array for time steps, **tau_sampling**, and the size of the array **tau_size**. These are the sampling times for the sources which CLASS saves as output of the perturbation module. Indeed, the sources for today's observables are very small much before decoupling, so it makes sense to avoid saving anything that does not contribute to the integral of the source later.

The saved sources are in the pointer to double ***sources**. Sources is a rank-5 pseudo-tensor whose indexes mean, in this order: mode (scalar, vector, tensor), initial condition type (adiabatic, isocurvature), perturbation variable, conformal time, Fourier mode k. Note

that it is very much *not a rectangular array*. The explicit dependence is indicated in the comment, and we will see that CLASS provides a simple macro to put any new source we may want to save into this array.

In order to support the calculation of the rotation spectra of the CMB, a phenomena we describe in section 19, we add the booleans **has_cl_cmb_rotation , has_source_rotation** and the **index_tp_rotation**. This first two booleans allow the user to specify if he wants to calculate these quantities. The index will serve to indecise the source of the rotation in the ***source** array.

The implementation in **source/perturbations.c** begins again from the *perturbations_init* function. Throughout this code the variables **index_md, index_ic** and **index_k** are present. The first will indicate what mode(scalar, vector or tensor) it is calculating, and the code will loop over the modes requested by the user. **index_ic** specifies which initial condition we are studied. These can be the standard adiabatic ones, or isocurvature perturbations of different kinds. Any type of perturbation is looped over by the code. **index_k** will refer to which Fourier mode the code is currently integrating over.

The table of values k to be integrated is initialized in the function *perturb_get_k_list*, which is called via *perturb_indices_of_perturbs* in *perturbs_init*. The algorithm to generate the values of k on which to integrate is quite complex but it is optimally tuned for standard cosmology and most non-standard applications. Instead of hard-coding any edits in the algorithm, it is better to change the precision parameters via the input file. The relevant parameters to be set in the initialization file can be found in **source/input.c** at and after the line

```
class_read_double("k_min_tau0", ppr->k_min_tau0);
```

The algorithm essentially fills the k values in ascending order, concentrating around the most important values that dominate around recombination and evenly spaced on a log scale after that.

The function *perturb_timesampling_for_source* instead fills the **tau_sampling** array in ascending order. The sampling size is set by the initialization variable *perturb_sampling_stepsize*. On the other hand, if one wants to control the integration step-size, the correct initialization variable to change is *perturb_integration_stepsize.*

After these initializations and some code to parallelize the process, *perturb_solve* is called, which begins the perturbation in earnest. The numerical ODE solver can be chosen by the user. We have always chosen the fourth-order Runge-Kutta algorithm. The reason for this, is that a variable time-step can be used in the algorithm which is chosen in *perturb_timescale*. This function is passed via pointer, and called by the Runge-Kutta algorithm at every step. The timescale is not the timestep but the Runge-Kutta algorithm will make that determination based on what we provide it from this function. As mentioned, this function checks if we are using some approximation, for example the tight-coupling approximation. When solving for highly oscillating scalar fields, we must make sure that the timestep is much smaller than the period of the oscillation (in this case of the perturbation to the field).

Let's discuss where the physics in earnest come into play. Which is why were here. The initial conditions to any given mode are in the function *perturb_initial_conditions*. Usually, these are set by the adiabatic, or isocurvature, conditions, if we needed to implement a new species we would need to that here. The code actually specifies the initial conditions in the

synchronous gauge and then does a gauge transformation to the conformal-Newtonian gauge.

The Boltzmann equations are implemented in the function *perturb_derivs.* To this function a current vector y of perturbations, indexed by variables **pt->index_pt_XXX,** is passed and their derivatives must be stored in the pointer ***dy** before returning the function. Note that the metric quantities are stored in the array ***pvecmetric** indexed by **ppw->index_mt_XXX.** These are defined in the struct *perturb_workspace* in **include/perturbations.h**. Although the metric values are stored differently their derivatives must be given here. This is the main function to edit the Boltzmann and dynamical Einstein equations if needed. To facilitate writing several equations the variables

$$\text{metric\_continuity} = \begin{cases} \frac{\dot{h}}{2} & \text{Synchronous} \\ -3\dot{\phi} & \text{Newtonian} \end{cases} \tag{15.1}$$

$$\text{metric\_euler} = \begin{cases} 0 & \text{Synchronous} \\ k^2\psi & \text{Newtonian} \end{cases} \tag{15.2}$$

$$\text{metric\_shear} = \begin{cases} \frac{\dot{h}+6\dot{\eta}}{2} & \text{Synchronous} \\ 0 & \text{Newtonian} \end{cases} \tag{15.3}$$

are defined and used throughout.

Some Einstein equations are not dynamical, as various quantities can be found algebraically given the integrated quantities. These relations are given in the function *perturb_einstein*. The equations here can be edited if one wants to include alternate theories of gravity.

As we solve the Boltzmann equations, we wish to save the sources of the line of sight integrals. This is done in the function *perturb_sources.* To add a new source, one simply needs to call the macro

```
_set_source_(ppt->index_tp_rotation) = SOURCEVALUE;
```

within the function, taking the other sources as examples. As we had mentioned earlier, the source for the temperature is separated in code into separate terms, to allow separate study and eventually summed later.

When we will want to calculate the rotation spectrum of the CMB due to a cosmological scalar field, we won't need to edit the Boltzmann equations as CLASS already supports them. What we want to add is a source function to the cosmological rotation. Therefore we defined the relevant indexes and booleans, for user control, in the struct. We initialize these booleans consistently as well as the index (this can be done where all the indexes are initialized in a similar fashion).

In *perturb_timescale* we estimate the period of oscillation of the scalar field and make the timescale at least twenty-times smaller than that, if this were be the smallest time scale. This precision can be relaxed if speed is more important. We didn't need the extra speed, so kept on the safe side.

The source term is added in *perturb_source* so we may later integrate the source of cosmological birefringence over a radial function.

15.7 Transfer module

The transfer module performs the line of sight integral for the photon perturbations, given the source calculated in the perturbation module. It also performs other integrals related to other experimentally accessible quantities. The transfer module is run after the *primordial and nonlinear* modules. The latter applies non-linear corrections to matter perturbations and updates the sources.

At the end of the integrations, the results of the transfer module are stored in the **double ** transfer** pointer in the transfer *struct*. This pointer is a rank five pseudo-tensor whose indexes mean, in order: mode (scalar-vector-tensor), initial condition type (adiabatic, isocurvature), type of source, value of angular moment ℓ, value of Fourier wavevector k. We note it is not rectangular.

For each type of source, the array has a number of timesteps which is given in the function *transfer_source_tau_size.* When adding a new source, for example the source of a CMB rotation spectra, one must add a line here to specify the size of the source array. In particular, it is possible to set the size of the source array equal to 1. The code will later interpret this as the presence of a Dirac delta in the integration. It is useful when one wants to compute a source quickly by replacing the visibility function in the line of sight integral with a Dirac delta centered on the time of last scattering.

The function *transfer_source* takes the source calculated in the perturbations to define the transfer sources. Indeed, CLASS makes a logical distinction between the two. In most cases, such as the photon temperature, the sources calculated by the perturbation module are copied. In other situation, the source is *redefined*. For example, we may want to only take the perturbation source for a small set of times. This is the case of the CMB rotation spectra when one uses the instantaneous decoupling approximation ($g(\tau) = \delta(\tau - \tau_{LSS})$) . Then we need to redefine the transfer source to only take the value for one time. We fill in the **sources** array of the method with one point, which is the value of the source at the time of recombination. When redefining a source, the array **tau0_minus_tau** must filled in the same way, as these represent the conformal time difference between today and the sampled point, which will be used during integration in the radial (spherical Bessel) functions.

If one has specified the size of the source array to be the maximum value allowed **tau_max** and defined the new dynamical index for the source, in the same way all the other dynamical indexes are defined and initialized, and the source to be integrated is the same as that saved from the perturbation module, then nothing needs to be done in *transfer_source*.

Most line of sight integrals are a convolution with a radial function. Most radial functions dependence on their argument are already known by CLASS and so we only have to specify which of the known radial functions can be used. The type **radial_function_type** is an *enum* defined in **include/transfer.h** . With this enumeration, the desired radial function can be set in the function *transfer_select_radial_function* which is called before integration. By default the chosen function is the usual spherical Bessel function. If a new kind of radial function is needed, that is not provided by CLASS, one can dig in the function *transfer_radial_function* which actually returns via array the values of a selected radial function. However, it may be more straightforward to edit the perturbation or transfer sources instead, since *transfer_radial_function* is not as documented.

The actual integration is done by the function *transfer_integrate* once all the above ingredients have been defined. Therefore it is usually not necessary to look at the details of this integration.

15.8 Spectra module

The transfer module has outputted, in the case of photons, the values of the modes today $F_{\gamma\ell}^{(m)}(k,\tau_0)$. It is the spectra module which integrates these over k using formulas such as (14.17). The primordial spectra is calculated in the *primordial module* and given as an input to the spectra module. The spectra module, just as the rest of CLASS, calculates the C_ℓ^{TT} separately for each pair of initial condition type and mode (Scalar-Vector-Tensor). Of course the sum is physical, but a separate calculation allows a researcher to understand each component on its own.

The result of the spectra module integrations are stored in the relative struct defined in **/include/spectra.h** . In particular the array ****cl** holds all the results. It is a rank four pseudotensor holding the computed results. Each index is, in order: mode (SVT), value of multipole ℓ, pair of initial conditions (adiabatic x adiabatic, adiabatic x isocurvature etc.), and type (temperature, polarization, etc.) . As can be seen in the immediately successive lines there is a similar array, of the same size, ****ddcl** which holds the *second derivatives of* C_ℓ with respect to ℓ . The derivatives are, of course estimated discretely. This is because CLASS, as other Boltzmann codes, does not calculate all the values of C_ℓ but samples on a few values of ℓ , interpolating in between. The second derivatives are used to interpolate. The values of ℓ to be used will be determined dynamically based on user-supplied input and stored in the array **double *l** in the same struct.

When including a new spectrum to be calculated, such as a rotation spectrum for the CMB, one defines and implements the dynamical indexing used by CLASS. There is usually not much to do, in terms of new physics, except to define the integrand to be integrated over on the Fourier mode k. This is done in the function *spectra_compute_cl*, inside the main loop.

The integration is then taken care of by CLASS.

15.9 Other modules

CLASS contains other modules which perform other computations.

The *primordial* module contains code to give a functional form of the primordial power spectrum as specified by the user. Many options are available by default for an analytical function. Alternatively, it is possible to supply a potential for a scalar mode driving inflation and equations are solved to give the primordial power spectrum generated by the quantum fluctuations of that field.

The *nonlinear and lensing* module apply second-order corrections to the spectra. By lensing we mean the tiny deflection of the the photon world-lines from last scattering to today due to the intervening matter potentials[149]. Since both the matter potentials and the photon perturbations are first order in the small perturbations, this is a second order effect. However, it is necessary to be included for next generation experiments. The non-linear module uses a modern version of the HALOFIT[207] fitting algorithm, whose underlying

idea we had sketched in section 13.17. CLASS runs the nonlinear module before the transfer module and applies non-linear corrections to the sources generated by the perturbation module. On the other hand ,the lensing module is run after the spectra module and applies a correction to the CMB spectra.

The output module handles all the output to file which is done before the application freeing the memory and quitting. There are extensive options to format the output that can be set by the user, including what normalization and dimensions are used.

16 Parameter Estimation

The point of a cosmological experiment is often to *infer* some fundamental quantity, a parameter of our theory. In general, the parameter of our theory is not what we measure directly. Rather, a cross section, resdshift or CMB anistropy is measured and the fundamental parameters is inferred. In cosmology this is not a trivial task. The relationship between the fundamental parameters of the theory (H_0, Ω_{cdm}, etc.) and the observable C_ℓ^{XX} is *not trivial*. Somehow, we'd like to invert the relationship so that, given the data, we can quote the estimated values of the parameters along with their error bars. We will review a few concepts which are widely used to perform parameter estimation of cosmological quantities.

16.1 The likelihood

We will adopt a Bayesian approach to cosmology. This is by no means the only possibility, and statistical analysis with a frequentist approach may be done as well[93]. In a Bayesian approach, we are interested in determining a probability distribution, *the posterior*, of some theoretical parameters. A common frequentist objection would be that a fixed "true" parameter is not a random variable and cannot have an actual distribution. The Bayesian would respond that the posterior is a reflection of our *subjective knowledge of the problem*, often joking by specifying what odds he/she would take on a bet of the outcome. The approach of a frequentist would be to quantify what is the probability of the data arising within a certain model (ie. the p-value) and at best rejecting the model or the parameters. This is an approach used in particle physics. The two approaches share a lot of mathematics and both are useful.

The Bayesian approach reflects the idea of probability as *missing knowledge* which gets updated with information and is therefore closer to what a layman would call *probability*. Bayesian statistic is intrinsically subjective. There is no mathematically consistent way to define a subjective Bayesian approach. Indeed, the posterior distribution found for any parameter will depend on the *prior distribution*. The prior distribution is a collection of knowledge we have before performing an experiment. Of course, a good scientist uses a prior which is actually the posterior from some previous experiment. But that also depends on yet another prior, and all the way back to some initial *Ur-prior,* as some may call it. One may attempt to choose an "ignorant" prior, which expresses "complete lack of knowledge". However, this also turns out to contain a degree of subjectivity. The case in point is that a uniform prior in one variable is not so in some combination of it. Why does that Bayesian

approach work then? It works if the knowledge gained during the experiment is much more precise than that known before and codified in the prior. The Bayesian interpretation must always be kept in mind when reading parameter estimations in the literature.

We begin with defining a fundamental quantity, the *likelihood* $\mathcal{L}$*[80].* Let's suppose we run an experiment where we are trying to estimate the parameters of some model

$$\{\lambda_j\}_{j=1}^{N} \tag{16.1}$$

Each time we run the experiment we measure some vector of data $\vec{x}_i$, a random sample which depends on the parameters λ_j. At the end of the experiment we collect $\{\vec{x}_i\}_{i=1}^{M}$ pieces of data. The definition of likelihood is the probability that the model with parameters λ_j produces the data $\vec{x}_i$.

$$\mathcal{L}(\{\vec{x}_i\},\{\lambda_j\}) \equiv P(\{\vec{x}_i\}|\{\lambda_j\}) \tag{16.2}$$

In the terminology of probability theory, it is the conditional probability of obtaining the data given fixed values of the underlying model. Using the elementary relation

$$P(A \cap B) = P(B|A)P(A) = P(A|B)P(B) \tag{16.3}$$

we may relate the posterior probability to the prior one, through the Likelihood. The posterior probability is the probability that the model parameters λ_j are the "true" ones, given the data.

$$P(\{\lambda_j\}|\{\vec{x}_i\}) = P(\{\vec{x}_i\}|\{\lambda_j\})\frac{P(\{\lambda_j\})}{P(\{\vec{x}_i\})} \tag{16.4}$$

The left hand side is the posterior. The term on the right hand side $P(\{\lambda_j\})$ is the prior probability for the underlying parameters and reflects the knowledge we have of them before conducting an experiment. This is the "subjective" term in the Bayesian approach. The denominator term $P(\{\vec{x}_i\})$ does not depend on the parameters λ_j, whose probability we are interested in. In fact, it is nothing more than a normalization constant and we will usually forget about it. Now we come to understand that the likelihood function is that which updates our knowledge of the parameters, passing from the prior probability distribution to the posterior one. If $\mathcal{L}$ is much more "peaked" in its central values of λ_j than the prior, the precise form of the prior becomes irrelevant. Therefore, dropping the prior (or taking it to be uniform), we get the probability distribution for the λ_j is proportional to the likelihood

$$P(\{\lambda_j\},\{\vec{x}_i\}) \propto \mathcal{L}(\{\vec{x}_i\},\{\lambda_j\}) \tag{16.5}$$

To get a handle on what is going on, let's consider a very simple example. Although simple, the ideas we will now show will be of use in the more complex case, when we are dealing with the cosmic microwave background (or any other cosmological experiment). Consider a random variable X which we measure many times to obtain a data set $\{x_i\}_{i=1}^{N}$. Each time we perform the measurement, we obtain a normally distributed sample with true mean and standard deviations $\lambda_1 = \mu$, $\lambda_2 = \sigma$. These represent the underlying parameters of our simple model. The likelihood (density) is the probability density of obtaining the

data set $\{x_i\}$ from a specific model.

$$\mathcal{L}(\{x_i\},\mu,\sigma) = \prod_{i=1}^{N} \frac{1}{\sqrt{2\pi\sigma^2}} e^{-\frac{x_i-\mu}{2\sigma^2}} = (2\pi\sigma^2)^{-\frac{N}{2}} \exp(-\frac{\sum_i (x_i-\mu)^2}{2\sigma^2}) \tag{16.6}$$

We assumed independence between successive measurements. We interpret the likelihood as the (un-normalized) probability density for μ and σ. Notice how, although the likelihood is Gaussian in the mean μ, it is absolutely not in the standard deviation. The posterior for the standard deviation is not normally distributed. Now we will employ a useful trick. It is usually much more algebraically manageable to work in terms of the log-likelihood $\ln \mathcal{L}$. Indeed in our case

$$\ln \mathcal{L} = -\frac{N}{2}\ln(\sigma^2) - \sum_i \frac{(x_i-\mu)^2}{2\sigma^2} - \frac{N}{2}\ln 2\pi \tag{16.7}$$

Because the log is a strictly growing function, the maxima and minima of the log-likelihoods are the same as the likelihood itself. We are interested in the maxima, as they represent the best-fit value for the underlying parameters. In general, they are good statistical estimators for them. Let's find the maximum of the likelihood as a function of μ. This answers the question: "Which value of μ gives the largest probability for the observed data?". Let's take the first derivative with respect to μ

$$\frac{d\ln\mathcal{L}}{d\mu} = \sum_i \frac{(x_i-\mu)}{2\sigma^2} \tag{16.8}$$

This is zero for

$$\mu = \bar{\mu} = \frac{1}{N}\sum_i x_i \tag{16.9}$$

exactly what we expected. $\bar{\mu}$ is our best fit estimate for the parameter μ, given the data. Next we ask ourselves what is the error we can associate to μ. We notice that for a normal distribution, the variance is linked to the second derivative of the log of the distribution as

$$\frac{1}{\sigma^2} = -\frac{d^2}{dx^2}\ln e^{-\frac{(x-\mu)^2}{2\sigma^2}}|_{x=\mu} \tag{16.10}$$

This formula is exact for a normally distributed variable, such as μ, but is good trick to remember. We often encounter quantities which are not normally distributed and estimation of the variance can be difficult. By assuming that the distribution is close to being normal, or that we can neglect terms larger than the quadratic one in the log-distribution, we can use the last relation to give a quick estimate of the variance. More generally, the value of the second distribution around the maximum indicates how flat the distribution is around the best-fit value. If the second derivative is large in absolute value, then the distribution is highly peaked (small variance), while if it is small then the distribution is not very peaked. This has implications for the size of the error bars on the estimate of the parameter.

Returning to our simple model, since μ is really normally distributed, we may immediately

infer the error we associate to μ

$$\frac{1}{\sigma_\mu^2} = \frac{N}{\sigma^2} \rightarrow \sigma_\mu^2 = \frac{\sigma^2}{N} \tag{16.11}$$

which is a well known result. Of course, we don't know the underlying value of σ so we'd replace it in the above formula with the value $\bar{\sigma}$ which maximizes the likelihood. Taking the derivative $\frac{d}{d\sigma^2} \ln \mathcal{L}$ we get another famous result

$$\bar{\sigma}^2 = \frac{\sum_i (x_i - \bar{\mu})}{N} \tag{16.12}$$

where we have replaced μ with the best fit value (which is correct since we are at the maximum). We may use these results to re-arrange the terms in the likelihood in order to write

$$\mathcal{L} = \frac{1}{\sqrt{2\pi C_N}} \exp\left(-\frac{(\mu - \bar{\mu})}{2C_N}\right) \tag{16.13}$$

where C_N is the "noise variance"

$$C_N = \frac{\sigma^2}{N} \tag{16.14}$$

compressing the information in the likelihood. Note that we dropped a few multiplicative constants in this step. We want to use this analogy to speak about the CMB.

In a measurement of the cosmic microwave background one observes some point on the sky. Even in the absence of any experimental noise or nuance, the temperature at that point is a random variable, since the primordial fluctuations are stochastic. In most theories, the distribution for the temperature fluctuation can be assumed to be Gaussian with mean zero (factoring out the monopole) and variance C_S, which is the power spectrum C_ℓ^{XX} we can compute from theory, expressed in real space. C_s contains all the dependence on the cosmological parameters we wish to study. The probability that the *signal*, the temperature on the sky, falls between s and $s + ds$ is given by

$$f(s)ds = \frac{1}{\sqrt{2\pi C_S}} e^{-\frac{s^2}{2C_S}} ds \tag{16.15}$$

The value s is *not the temperature we actually measure. Rather, it is what we wish we could measure: the signal*. The actually measured value will depend on the experimental setup. Indeed, if s is the real value we assume the quantity Δ actually measured by experiment is normally distributed around s with mean C_N, the noise variance.

$$f(\Delta|s)d\Delta = \frac{1}{\sqrt{2\pi C_N}} e^{-\frac{(\Delta - s)^2}{2C_N}} d\Delta \tag{16.16}$$

The probability of obtaining a value of Δ is given by the product of the two distributions, integrated over all possible values of s, which we cannot see. All the dependence from the cosmological parameters is only in C_S, while in C_N we have the experimental parameters.

The likelihood is

$$\mathcal{L} = \int_{-\infty}^{+\infty} \frac{ds}{\sqrt{2\pi C_S}} \frac{1}{\sqrt{2\pi C_N}} \exp\left(-\frac{s^2}{2C_S} - \frac{(\Delta - s)^2}{2C_N}\right) \tag{16.17}$$

The integral may be calculated by completing the square

$$\begin{aligned} -\frac{s^2}{2C_S} - \frac{(\Delta - s)^2}{2C_N} &= -\frac{1}{2}\left(\frac{C_N + C_S}{C_N C_S}\right) \times \\ &\left[\left(s - \frac{C_S}{C_N + C_S}\right)^2 + \Delta^2 \frac{C_S}{C_S + C_N}\left(1 - \frac{C_S}{C_N + C_S}\right)\right] \end{aligned}$$

Defining the *total covariance*

$$C = C_N + C_S \tag{16.18}$$

the likelihood becomes

$$\mathcal{L}(\Delta, C) = \frac{1}{\sqrt{2\pi C}} e^{-\frac{\Delta^2}{2C}} \tag{16.19}$$

In a more realistic experiment, measurements are taken in N_p distinct pixels across the sky, each with a finite size. In that case the above procedure can be generalized, with some lengthy but straightforward algebra.

$$\mathcal{L} = \frac{1}{(2\pi)^{\frac{N_p}{2}} (\det C)^{\frac{1}{2}}} \exp\left(-\frac{1}{2}\Delta C^{-1} \Delta\right) \tag{16.20}$$

where Δ is now an N_p sized column vector and C is a $N_p \times N_p$ covariance matrix. In principle, the measurements in different pixels may not be uncorrelated, and so the covariance matrix is not diagonal. However, it is still the sum of a signal and noise term $C = C_N + C_S$. Usually, the noise covariance C_N can be thought to be proportional to the identity. On the other hand, the theoretical covariance C_S is not. We know that the power spectrum of the CMB is non-trivial and strong correlations exist across the sky. Furthermore, we note that although the distribution is normally multivariate in the Δ , it is not in the matrix C which contains all the theoretical parameters, of which we want to know the distribution.

16.2 CMB Window functions

Understanding the sensitivity of a CMB anisotropy measurement to the underlying parameters can be reduced to understanding the form of the covariance matrix. In principle all the information is contained in (16.20). In practice, the form can become quite complicated. To proceed we shall study a typical CMB experiment. Although we are idealizing a little, the formulas and concepts used here are fundamental to any experiment.

In a realistic experiment, CMB temperatures are measured within a number of pixels N_p. Within the i-th pixel, the temperature is sampled with a *beam pattern* $B_i(\hat{n})$. The temperature fluctuation reported in the pixel is given by the integral of the underlying, "real",

temperature weighed by the beam[80].

$$s_i = \int d\Omega \Theta(\hat{n}) B_i(\hat{n}) \tag{16.21}$$

The beam pattern codifies the capacity of an experiment to distinguish points on the sky from one another and how the measurement of the temperature is smudged by the inability to distinguish points closer than the beam. Here $\Theta(\hat{n})$ represents the actual theoretical temperature fluctuation, and s_i the *signal* in pixel i . Ideally, we'd like to measure this signal, but in practice we will measure a value Δ which is normally distributed around it due to the experimental noise as (16.17).

s_i is a random variable due to the stochastic nature of the primordial fluctuations, so we can calculate the covariance of the signal $C_{S,ij}$ between two pixels i, j.

$$C_{S,ij} =< s_i s_j > \tag{16.22}$$

the average being due to the *theoretical fluctuations only*. Expanding into spherical harmonics

$$C_{S,ij} = \int d\Omega d\Omega^{'} B_i(\hat{n}) B_j(\hat{n}^{'}) \sum_{\ell m} \sum_{\ell' m'} Y_\ell^m(\hat{n}) Y_{\ell'}^{m^{'}}(\hat{n}^{'}) < a_{\ell m}^T a_{\ell' m'}^{T*} > \tag{16.23}$$

Using the definition of the C_ℓ^{TT} given in (14.9)

$$C_{S,ij} = \int d\Omega d\Omega^{'} B_i(\hat{n}) B_j(\hat{n}^{'}) \sum_{\ell} C_\ell^{TT} \sum_{m} Y_\ell^m(\hat{n}) Y_\ell^m(\hat{n}^{'}) \tag{16.24}$$

We use the addition theorem for spherical harmonics (B.22) and rewrite in terms of the Legendre polynomial P_ℓ.

$$C_{S,ij} = \sum_{\ell} \frac{2\ell + 1}{4\pi} C_\ell^{TT} W_{\ell,ij} \tag{16.25}$$

where we defined the *window function*

$$W_{\ell,ij} = \int d\Omega d\Omega B_i(\hat{n}) B_j(\hat{n}) P_\ell(\hat{n} \cdot \hat{n}^{'}) \tag{16.26}$$

The window function contains all the experimental smudging of nearby points. If $B_i(\hat{n}) = \delta(\hat{n} - \hat{n}_i)$, $\hat{n}_i$ being the position of a pixel then, $P_\ell(1) = 1$ and the window function would reduce to the identity. By observing (16.25) in this perfect experiment limit, it is obvious that the covariance matrix for the $a_{\ell m}^T$ is given by $C_\ell^{TT} W_{\ell,ii}$ That is, if we write the likelihood (16.20) using measured $a_{\ell m}$ as the vector Δ of data, instead of spatial pixels, the signal covariance matrix is simple enough.

Let's find explicit forms for the window functions. Assuming the beam is much smaller than the curvature of the sky, it is possible to work in the tangent plane to the celestial sphere. Thus instead of $\hat{n}$, we use the two-dimensional vectors $\vec{x}$, $\vec{x}^{'}$, and

$$\hat{n} \cdot \hat{n}^{'} = \cos |\vec{x} - \vec{x}^{'}| \tag{16.27}$$

With this change of variables

$$W_{\ell,ij} = \int d^2x d^2x^{'} B_i(\vec{x}) B_j(\vec{x}) P_\ell(\cos|\vec{x}\cdot\vec{x}^{'}|) \tag{16.28}$$

Now we use property (B.28) of the Legendre polynomial which states that asymptotically, with $\ell \to +\infty$, the Legendre polynomial tends to the Bessel function of order zero. Since we are working in the small angle limit this use of large ℓ is justified. We also suppose that the beams are narrow enough that there is no significant overlap between different pixels. This reduces the window function to being diagonal, $i = j$ and allows us to take $\vec{x} - \vec{x}^{'}$ to be small. So that we may write

$$W_{\ell,ii} \simeq \int d^2x d^2x^{'} B_i(\vec{x}) B_j(\vec{x}) J_0(\ell|\vec{x} - \vec{x}^{'}|) \tag{16.29}$$

with $J_0(x)$ the order zero Bessel function. Of course, in any serious experiment the window function can be evaluated numerically from the definition without any approximations needed. These approximations are very useful and the final results we obtain are used extensively in the literature. Now we write the Bessel function as the integral

$$J_0(\ell|\vec{x} - \vec{x}^{'}|) = \frac{1}{2\pi}\int_0^{2\pi} d\phi e^{-i\ell|\vec{x}-\vec{x}^{'}|\cos\phi} \tag{16.30}$$

And call $\vec{\ell}$ the vector with magnitude ℓ and direction such that it forms an angle ϕ with $\vec{x} - \vec{x}^{'}$. Then, in the window function two Fourier transforms appear as

$$\tilde{B}_i(\vec{\ell}) = \int d^2x B_i(\vec{x}) e^{-i\vec{\ell}\cdot\vec{x}} \tag{16.31}$$

So that the window function may be written as

$$W_{\ell,ij} \simeq \delta_{ij}\frac{1}{2\pi}\int_0^{2\pi} d\phi |\tilde{B}_i(\vec{\ell})|^2 \tag{16.32}$$

We consider now a very common beam pattern, known as the Gaussian beam. The beam for pixel i is given by

$$B_i(\vec{x}) = \frac{1}{2\pi\sigma^2}\exp\left(-\frac{|\vec{x} - \vec{x}_i|^2}{2\sigma^2}\right) \tag{16.33}$$

Since we are getting a diagonal window function we may set $\vec{x}_i = 0$ for simplicity. The Fourier transform is Gaussian and independent of the angle

$$\tilde{B}_i(\vec{\ell}) = e^{-\frac{\ell^2\sigma^2}{2}} \tag{16.34}$$

Which gives us the Gaussian window function

$$W_{\ell,ii} = e^{-\ell^2\sigma^2} \tag{16.35}$$

Let's discuss this a moment. σ is the experimental beam width in radians on the sky. An

angular fluctuation θ is represented by modes $\ell \sim \theta^{-1}$. We see that if $\ell < \sigma^{-1}$ then the window function is order ~ 1, which means the modes are being well discriminated. When $\ell > \sigma^{-1}$, or $\theta < \sigma$ the window function falls off exponentially, signaling the experimental limitation to the discrimination of high enough modes. Often in the literature the Gaussian beam width is quoted by it full width at half maximum (FWHM) value and the relation is

$$\sigma = \frac{FWHM}{\sqrt{8 \ln 2}} \tag{16.36}$$

16.3 Fisher matrix and Newton-Raphson method

Once we have calculated covariance matrices, we need to somehow make estimates for the underlying parameters of the models. We will illustrate the Newton-Raphson method to iteratively find the best fit values for a theoretical parameter, given the likelihood function, and estimate the error using the Fisher matrix. These methods are useful since they do not require complete knowledge of the likelihood function to gain valuable knowledge.

For simplicity, we study a likelihood $\mathcal{L}$ with only one underlying theoretical parameter λ. Generalization of the following to more parameters is straightforward, but the algebra becomes heavy and it is not more illuminating. We will give the general results at the end.

To get an estimator for λ we will look for the value $\bar{\lambda}$ which maximizes $\mathcal{L}$. The likelihood is a complicated function of λ and solving the derivatives analytically is often impossible. The trick for the Newton-Raphson method is to take an initial guess for the maximum $\lambda^{(0)}$. With this guess we will proceed in an iterative fashion until a sequence of guesses converges. In practice, this usually turns out to be very efficient. However, as with any maximization problem, there is a chance of falling into a local, but not absolue maxima. Now, with the guess $\lambda^{(0)}$ which is hoped to be somewhat close to $\bar{\lambda}$ we Taylor expand the log-likelihood around $\lambda^{(0)}$. The derivative, of which we want to find the roots, will be given by[80]

$$\frac{d \ln \mathcal{L}}{d\lambda}|_{\lambda} = \frac{d \ln \mathcal{L}}{d\lambda}|_{\lambda^{(0)}} + \frac{d^2 \ln \mathcal{L}}{d\lambda^2}|_{\lambda^{(0)}} (\lambda - \lambda^{(0)}) + o(\lambda - \lambda^{(0)})^2 \tag{16.37}$$

At the maximum the derivative is zero so the solution is

$$\bar{\lambda} \simeq \lambda^{(0)} - \frac{d \ln \mathcal{L}}{d\lambda}|_{\lambda^{(0)}} \left(\frac{d^2 \ln \mathcal{L}}{d\lambda^2}|_{\lambda^{(0)}} \right)^{-1} \tag{16.38}$$

The solution is not exact, since we have left out higher order terms. However, this solution of $\bar{\lambda}$ can be used as the input for a new guess. We may repeat with $\lambda^{(0)} = \bar{\lambda}$ and proceed iteratively a few times until we believe to be reasonably close to the maximum that we expect $\ln \mathcal{L}$ to be nearly quadratic in $(\lambda - \bar{\lambda})$ for small deviations from the maximum. Once we have arrived close, we derive more explicit expressions.

Let's now take a likelihood of the form (16.20). The dependence on the theoretical parameter is in the covariance matrix. The first derivative of the log likelihood is

$$\frac{d \ln \mathcal{L}}{d\lambda} = -\frac{1}{2} \ln(\det C') - \frac{1}{2} \Delta C^{-1'} \Delta \tag{16.39}$$

where a prime indicates derivatives with respect to λ. The derivative of the inverse matrix

can be written as

$$C^{-1'} = (C^{-1}CC^{-1})' = -C^{-1}C'C^{-1} \tag{16.40}$$

using the Leibnitz product rule. Thus

$$\frac{d\ln\mathcal{L}}{d\lambda} = \frac{1}{2}\Delta C^{-1}C'C^{-1}\Delta - \frac{1}{2}\mathrm{Tr}(C^{-1}C') \tag{16.41}$$

where we used the identity $\ln\det C = \mathrm{Tr}\ln C$. The second derivative is straightforward to obtain as well

$$\frac{d^2\ln\mathcal{L}}{d\lambda^2} = -\Delta C^{-1}C'C^{-1}C'\Delta + \frac{1}{2}\mathrm{Tr}(C^{-1}C'C^{-1}C' - C^{-1}C'') + \frac{1}{2}\Delta C^{-1}C''C^{-1}\Delta \tag{16.42}$$

Now we may estimate $\bar{\lambda}$ using (16.38). However we will use a trick. To remove the dependence on the data in the estimate, instead of taking the explicit value of the data vectors Δ we take their averages, $\Delta_i\Delta_j \to <\Delta_i\Delta_j> = C_{ij}(\bar{\lambda}) \simeq C_{ij}(\lambda^{(0)})$. This equality is pretty much the definition of covariance matrix. We have made the indexes of the the vector and matrix explicit, so there is no confusion in the algebra. Let's see how one term of the second derivative works out, for example the last. Making all the indexes explicit (and summing over all repeated indexes)

$$\frac{1}{2}\Delta_i(C^{-1})_{ij}(C'')_{jk}(C^{-1})_{kl}\Delta_l \tag{16.43}$$

Writing out the indexes allows us to commute the variables. We move the Δ close to one another and take $\Delta_i\Delta_l \to C_{il}$. Now all the indexes are contracted, so we get the trace of the product

$$\frac{1}{2}\Delta C^{-1}C''C^{-1}\Delta \to \frac{1}{2}\mathrm{Tr}(CC^{-1}C''C^{-1}) = \frac{1}{2}\mathrm{Tr}(C''C^{-1}) \tag{16.44}$$

The other term works similarly. The *average curvature* is therefore defined as

$$F_{\lambda\lambda} \equiv -<\frac{d^2\ln\mathcal{L}}{d\lambda^2}> \tag{16.45}$$

Indeed, we recall how the second derivative of a Gaussian is the inverse of the variance. So this curvature should also roughly be the inverse of a variance if we are close enough to the maximum of the log-likelihood. Explicitly

$$F_{\lambda\lambda} = \frac{1}{2}\mathrm{Tr}(C'C^{-1}C'C^{-1}) \tag{16.46}$$

The approximate solution is given by

$$\hat{\lambda} = \lambda^{(0)} + F^{-1}_{\lambda\lambda}\frac{\Delta C^{-1}C'C^{-1}\Delta - \mathrm{Tr}(C^{-1}C')}{2} \tag{16.47}$$

We take care to repeat that all the C are evaluated at $\lambda^{(0)}$. Notice how we are now using a hat $\hat{\lambda}$ instead of a bar, since $\hat{\lambda} \neq \bar{\lambda}$. Indeed the new quantity $\hat{\lambda}$ is different from the solution we were looking for. However, it has other good properties. We may take $\hat{\lambda}$ to be *the estimator* for the best fit solution $\bar{\lambda}$. It has the nice property of being quadratic in the data vector Δ. There is no mystery in using this particular combination as an estimator. Any

function may be used as an estimator, that does not make it a good estimator. We will now show that this is a good estimator of $\bar{\lambda}$, in the sense that it is an unbiased one.

The reason we defined the curvature with two similar indexes $F_{\lambda\lambda}$ is because we now generalize to the existence of more than one underlying theoretical parameter $\lambda_\alpha = \lambda_1, \lambda_2 \ldots$. Repeating the calculations in the this extended case it is straightforward to find that the same estimator is given by

$$\hat{\lambda}_\alpha = \lambda_\alpha^{(0)} + (F^{-1})_{\alpha\beta} \frac{\Delta C^{-1} C_{,\beta} C^{-1} \Delta - \mathrm{Tr}(C^{-1} C_{,\beta})}{2} \tag{16.48}$$

where we denote $C_{,\beta}$ derivatives $\frac{\partial C}{\partial \lambda_\beta}$ and the repeated indexes are summed over. The *Fisher matrix is defined as*

$$F_{\alpha\beta} \equiv - < \frac{\partial^2 \ln \mathcal{L}}{\partial \lambda_\alpha \partial \lambda_\beta} >= \frac{1}{2} Tr(C_{,\alpha} C^{-1} C_{,\beta} C^{-1}) \tag{16.49}$$

By the properties of the trace it is a symmetric matrix. We have already seen how this quantity is related to the curvature of the likelihood around the maximum, and therefore to the error bars on the estimation for the parameters λ_α.

Now we shall prove that $\hat{\lambda}_\alpha$ is indeed an unbiased estimator of $\bar{\lambda}$, the best fit value, ie that $< \hat{\lambda}_\alpha >= \bar{\lambda}_\alpha$. To prove this we note that the expected value

$$< \Delta\Delta >= C(\bar{\lambda}_\alpha) = C + C_{,\gamma}(\bar{\lambda}_\gamma - \lambda_\gamma^{(0)}) + o(\bar{\lambda}_\gamma - \lambda^{(0)}) \tag{16.50}$$

Indeed, all the covariance matrices are calculated at $\lambda^{(0)}$ while the average value is equal to the true, unknown to the analyst, covariance matrix $C(\bar{\lambda}_\alpha)$. When using the same expected value in the second derivative we had neglected the first order difference since that term was already multiplying a term of order 1 in the difference, from the Taylor expansion. With this expansion it is some straightforward algebra to show that

$$< \hat{\lambda}_\alpha >= \bar{\lambda}_\alpha \tag{16.51}$$

Which is actually a great result. We can iteratively find the best fit value, evaluating the likelihood function a limited number of times. The variance of the estimator can also be found

$$< (\hat{\lambda}_\alpha - \bar{\lambda}_\alpha)(\hat{\lambda}_\beta - \bar{\lambda}_\beta) >= (F^{-1})_{\alpha\beta} \tag{16.52}$$

which is what we expect when we are very close the maximum. The right hand side is the inverse Fisher matrix evaluated at the peak. The Fisher matrix does not depend on the data. Indeed, given any reasonable parameter point λ_α, which may not be $\bar{\lambda}_\alpha$, the Fisher matrix may be calculated and used as a good estimate for the error bars on the parameters. It turns out this estimate very well agrees with those made with more involved procedures involving the full likelihood function.

16.4 CMB Fisher matrix and forecasting

The Fisher matrix, defined through the covariance matrix can be used to determine the ability of an experiment to constrain parameters. Let's calculate the Fisher matrix beginning with a likelihood of the form (16.20). Previously, we thought of our data as being the temperature measured in some pixel on the sky, so that schematically $\Delta = (\Theta(\hat{n}_1), \Theta(\hat{n}_2)\dots)$. However none of the math really depended on this, except, of course, the explicit covariance matrix (16.25) on the sky, defined through the window function (16.26). We may choose to describe our experimental data by a vector of angular expansion coefficients $a^T_{\ell m}$. So $\Delta = (a^T_{2,-2}, a^T_{2,-1}, \dots a^T_{3,-3}, a^T_{3,-2}, \dots)$. The likelihood is assumed to be Gaussian in Δ with a total covariance matrix which, just as before, is the sum of an underlying theoretical (cosmic variance) one and an experimental noise: $C = C_S + C_N$.

Before proceeding, let's comment this assumption of Gaussianity. In recent years the non-Gaussianity of the CMB has been studied and is an interesting new avenue for discovery. By Gaussianity, we mean that the fluctuations on the sky follow a normal distribution, an assumption built in deriving our likelihood. Indeed, since the primordial fluctuations appear to be Gaussian and the subsequent evolution is linear, the perturbations today remain Gaussian in form. Non-Gaussianity therefore holds information either on primordial non Gaussianity or non-linear evolution of the perturbations[15]. Inflation theory predicts a Gaussian primordial spectrum. Therefore, tests of primordial non-Gaussianity may constrain inflationary models. Non-gaussianity in subsequent evolution arises mostly through the *lensing of the CMB*, expecially lensing of polarization which produces non-primordial B-modes[149]. It is conceptually important to understand that these non-Gaussian contributions are, in principle, detectable and therefore neatly separable from the Gaussian, linear, part of the fluctuations on the CMB. For example, it is possible in some limits to *delense* the sky, obtaining the unlensed CMB from the lensed one[198]. We will now proceed neglecting non-Gaussian contribution[211, 140].

The covariance matrix C can be written explicitly with its indexes $C_{\ell m;\ell' m'}$, explicitly

$$C = \begin{pmatrix} C_{\ell=2,m=-2;\ell'=2;m'=-2} & C_{2,-2;2,-1} & \dots & C_{2-2;2,2} & C_{2,-2;3,-3} & \dots \\ C_{2,-1;2,-2} & C_{2,-1;2,-1} & \dots & C_{2,-1;2,2} & C_{2,-1;3,-3} & \dots \\ \vdots & & & & & \\ C_{3,-3;2,-2} & C_{3,-3;2,-1} & \dots & C_{3,-3;2,2} & C_{3,-3;3,-3} & \dots \\ \vdots & & & & & \end{pmatrix} \tag{16.53}$$

Any matrix multiplication of $C_{\ell m;\ell' m'}$ must sum over both indexes (ℓ, m), since any combination of (ℓ, m) represents a row, or a column, of the matrix. We wish to find an explicit form for the covariance matrix. For the signal part, we recall how, following equation (16.25), we noticed that the signal covariance for the $a^T_{\ell m}$ must simply be C^{TT}_ℓ, the pure theoretical covariance of the $a^T_{\ell m}$, times the window function. This works in the limit where the window function $W_{\ell,ij}$ is proportional to the identity. We will now explicitly use the result for a Gaussian beam pattern. The function reduces to (16.35), $W_{\ell,ii} = e^{-\ell^2\sigma^2}$.

$$C_{S;\ell m;\ell' m'} = C^{TT}_\ell e^{-\ell^2\sigma^2}\delta_{\ell\ell'}\delta_{mm'} \tag{16.54}$$

The noise covariance matrix is also diagonal, as long as there is no substantial correlation between the pixels. We calculate the noise of the fractional temperature fluctuation $C_{N,\ell}$

$$< a^{N,T}_{\ell m} a^{N,T*}_{\ell' m'} >= C_{N,\ell}\delta_{\ell\ell'}\delta_{mm'} = \int d\Omega d\Omega^{'} Y^{m*}_{\ell}(\hat{n}) Y^{m}_{\ell}(\hat{n}^{'}) < \Theta(\hat{n})\Theta(\hat{n}^{'}) > \tag{16.55}$$

The noise is characterized by the variance of fluctuation at any point in the sky. Assuming it is uncorrelated across the sky, then $< \Theta(\hat{n})\Theta(\hat{n}^{'}) > \propto \delta(\hat{n} - \hat{n}^{'})$ and the spherical harmonics give the proportionality factor. To fix the constant we note that the noise in a single pixel must proportional to the size $\Delta\Omega$ of the pixel in steradians multiplied by the variance of the fractional temperature fluctuations σ^2_N/T^2_{CMB}. Where σ^2_N is the variance of the Gaussian noise of the temperature fluctuations $\delta T = T\Theta$. Therefore

$$C_{N;\ell m,\ell' m'} = \frac{1}{T^2}\Delta\Omega\sigma^2_N\delta_{\ell\ell'}\delta_{mm'} \equiv w^{-1}\delta_{\ell\ell'}\delta_{mm'} \tag{16.56}$$

Note that many authors use the absolute temperature fluctuations of the CMB, instead of the fractional fluctuations, as we discussed at the end of section (14.1). If that is the case, covariance of the signal C_S^{absolute} has the same dependence on the absolute $C_\ell^{TT,\text{absolute}}$ as in the case of the fractional fluctuations. The noise covariance matrix has the same form, but with a weight $w^{-1} = \Delta\Omega\sigma^2_N$ as the temperature of the CMB gets simplified.

Putting it all together, the covariance matrix of a typical CMB experiment may be given as

$$C_{\ell m;\ell' m'} = \delta_{\ell\ell'}\delta_{mm'}\left[C^{TT}_\ell e^{-\ell^2\sigma^2} + w^{-1}\right] \tag{16.57}$$

With this, we may calculate the Fisher matrix (16.49). Before we do, let's ask ourselves what we want to take as the underlying theory parameters λ_α. It is simplest if we choose our parameters to be the C^{TT}_ℓ themselves. Studying the Fisher matrix will give us insight on how precisely we can measure a single parameter C^{TT}_ℓ of the power spectrum. The inverse covariance matrix is given by

$$(C^{-1})_{\ell m;\ell' m'} = \delta_{\ell\ell'}\delta_{mm'}\left[C^{TT}_\ell e^{-\ell^2\sigma^2} + w^{-1}\right] \tag{16.58}$$

and the derivative with respect to a parameter λ_α is

$$C_{\ell m;\ell' m',\alpha} = \delta_{\ell\ell'}\delta_{mm'}\delta_{\ell\alpha}e^{-\ell^2\sigma^2} \tag{16.59}$$

The Fisher matrix is straightforward, one has to only take care to sum the many free indexes properly.

$$F_{\ell\ell'} = \frac{2\ell+1}{2}\delta_{\ell\ell'}e^{-2\ell^2\sigma^2}\left[C^{TT}_\ell e^{-\ell^2\sigma^2} + w^{-1}\right]^{-2} \tag{16.60}$$

The Fisher matrix for a CMB experiment is diagonal in ℓ, implying that there is no correlation between the posterior distribution for the C^{TT}_ℓ at different ℓ. Error bars on the C^{TT}_ℓ generated from an experiment can be quoted indipendently. Using the estimator (16.52) for the variance we see that the error on the parameters is

$$\delta C^{TT}_\ell = \sqrt{\frac{2}{2\ell+1}}(C^{TT}_\ell + w^{-1}e^{\ell^2\sigma^2}) \tag{16.61}$$

As we had predicted back when first encountering the power spectrum C_ℓ^{TT}, there is a fundamental lower limit in the determination of the theoretical parameters due to *cosmic variance*. We only sample the $a_{\ell m}$ $(2\ell+1)$ times at a given ℓ. This is why the factor $(2\ell+1)$ appears in the denominator. The cosmic variance is expecially important at low ℓ where we have fewer samples.

There are two corrections which we add to the above formula. They can be derived rigorously but this is out of the scope of the present text. First of all, when deriving the window function we had taken a large ℓ, small angle limit. If ℓ is not too large then it turns out that the factors that appear above are $e^{\ell(\ell+1)\sigma^2}$ instead of $e^{l^2\sigma^2}$ in the final determination of the error δC_ℓ^{TT}. This correction is mostly unimportant since, in any typical CMB analysis a large chunk of the data is at $\ell \gtrsim 10$. The other, more important, correction relates to the fact that CMB experiments never observe the full sky, which we have implicitly been assuming. An experiment will quote the fraction of the sky

$$f_{sky} \tag{16.62}$$

it is observing. In practice, this means we have less than $(2\ell+1)$ samples of $a_{\ell m}^T$ at each ℓ and the counting term $(2\ell+1)^{-1/2}$ gets cut accordingly. Thus, a more precise determination of the error is

$$\delta C_\ell^{TT} = \sqrt{\frac{2}{(2\ell+1)f_{sky}}}\left(C_\ell^{TT} + w^{-1}e^{\ell(\ell+1)\sigma^2}\right) \tag{16.63}$$

With this, let's see how we may do a forecasting of the ability of some future experiment to constrain cosmological parameters η_α (we won't use λ again as we've previously said they were the C_ℓ's themselves). We want to determine how large the *error bars* will be. The cosmological parameters are the usual ones $H_0, \Omega_{cdm}h^2$,etc. In a forecast, we assume some values η_α close to the true underlying ones. Usually, the best fit values of previous experiments are used. With these, we calculate, with a Boltzmann code what the *observed* $C_{\ell,obs}^{TT}$ *may be.* To do this, one can add to the theoretically calculated C_ℓ's an instrumental noise, $N_\ell = w^{-1}e^{\ell(\ell+1)\sigma^2}$, to simulate noisy data as in (16.63). This mock data is known as the *fiducial model.* Then, the chi-square

$$\chi^2(\eta_\alpha) = \sum_\ell \frac{(C_\ell^{TT}(\eta_\alpha) - C_{\ell,\text{obs}}^{TT})^2}{2\delta C_\ell^2} \tag{16.64}$$

relates the C_ℓ^{TT} calculated at some point in parameter space with the observed values. We take the errors δC_ℓ to be those related to $C_{\ell,\text{obs}}^{TT}$, as they would those the experimenters provide.

We expect the χ^2 to reach a global minimum at the true value of the parameters $\bar{\eta}_\alpha$. However, we are not interested in the actual values of $\bar{\eta}_\alpha$ but on the error bars we may put on them. This, of course, is related to the curvature of the χ^2 at the minimum. The second derivative at the minimum is

$$\frac{1}{2}\frac{\partial^2\chi}{\partial\eta_\alpha\partial\eta_\beta} = \sum_\ell \frac{1}{\delta C_\ell^2}\left[\frac{\partial C_\ell^{TT}}{\partial\eta_\alpha}\frac{\partial C_\ell^{TT}}{\partial\eta_\beta} - (C_\ell^{TT} - C_{\ell,\text{obs}}^{TT})\frac{\partial^2 C_\ell^{TT}}{\partial\eta_\alpha\partial\eta_\beta}\right] \tag{16.65}$$

The second term on the right hand side, containing the second derivatives, is usually neglected. The rationale is that the difference $C_\ell^{TT} - C_{\ell,\rm obs}^{TT}$ is sometimes negative and sometimes positive. The sum over many terms will average zero. Indeed if we are close to the minimum, we expect we can replace this with the average of the differences from the mean, which is zero. Now, as we remarked several times we can define the curvature matrix

$$F_{\alpha\beta} = \sum_\ell \frac{1}{\delta C_\ell^2} \frac{\partial C_\ell^{TT}}{\partial \eta_\alpha} \frac{\partial C_\ell^{TT}}{\partial \eta_\beta} \tag{16.66}$$

This looks like a Fisher matrix, but it technically isn't, since the Fisher matrix is the expectation value of the curvature, which is the second derivative only if the distribution is normal in the parameters, which it is not. However, we can use an iterative algorithm such as the Newton-Raphson described in section 16.3 to get close to the minimum. Even if then the distribution is still not gaussian, it is then good to estimate the error on the parameter η_α as

$$\sigma_\alpha^2 = (F^{-1})_{\alpha\alpha} \tag{16.67}$$

With no sum implied. This error, is the variance of η_α assuming all the other parameters are unknown.

The Fisher matrix method is a very good and easy way to forecast the reach of future experiments. It agrees with more detailed, numerical, calculations. We will describe a more precise forecasting method in section 16.5 , where we use a Markov Chain Monte Carlo method. This method is much more computationally expensive.

To conclude this section, wesee what happens if we include polarization in our experimental signal. The vector Δ to be used in the likelihood now is $\Delta = \{\dots, a_{\ell m}^T, a_{\ell m}^E, a_{\ell m}^B \dots\}$, and so on for every (ℓ, m). The covariance matrix then has 9 times more terms. Repeating the same arguments, it is still diagonal with respect to ℓ, m , but due to the intrinsic cross-correlations between temperature and polarization, it is not at fixed (ℓ, m) . Indeed the covariance matrix for a specific value of ℓ is

$$C_{(\ell)} = \begin{pmatrix} C_\ell^{TT} & C_\ell^{TE} & 0 \\ C_\ell^{TE} & C_\ell^{EE} & 0 \\ 0 & 0 & C_\ell^{BB} \end{pmatrix} \tag{16.68}$$

Here we assume there is no theoretical cross-correlation between the T, E and B modes. The total covariance matrix is, again, the sum of the signal and noise. The noise can be assumed to be uncorrelated between different modes. $C_N^{TE} = C_N^{TB} = C_N^{EB} = 0$. The noise can be given by the same form as before (16.56), but with different values of the weight w, linked to different values of the Gaussian noise variance $\sigma_{N,T,E,B}^2$. In general, it can be assumed that the experimental noise in E and B modes is the same

$$\sigma_{N,E}^2 = \sigma_{N,B}^2 \equiv \sigma_{N,P}^2 \tag{16.69}$$

The polarization noise needs to be determined by the experimenters, however a useful estimate is that the variance of polarization is double that of the temperature, since polarization involves a difference between two temperature measurements in practice. $\sigma_{N,P}^2 = 2\sigma_{N,T}^2$.

This implies that $w_P^{-1} = 2w_N^{-1}$. With these adjustments, the same steps as above can be repeated to make a Fisher forecast. There is a slight complication due to the non-diagonal terms, but since we can reduce to working to a 2×2 matrix, these are algebraically tractable. Qualitatively, the same results are achieved[169].

16.5 Markov Chain Monte Carlo

We will now describe a concrete numerical method to sample points directly from the Likelihood of a CMB experiment. The previous methods for getting the error bars of cosmological parameters are analytic but approximate. They neglect the full distribution and suppose it is nearly Gaussian around the peak. This is good for many applications, but is not without its pitfalls. The method we describe here is known a Markov Chain Monte Carlo (MCMC) method. In particular, we will describe a specific algorithm, known as the *Metropolis-Hastings* algorithm[121, 150].

A Markov Chain is a stochastic sequence of points in a parameter space where the probability distribution of the next point of the sequence depends only on the last one[107]. In this sense a Markov process is *memoryless.* One can make predictions on the next point, or event, based only on the present state, disregarding past history. A well known example of a Markov chain is the random walk. The position of a particle is measured at discrete times, and the probability distribution for the position at a discrete time t_{n+1} is only a function of the position at t_n.

The fundamental problem in the statistics of cosmology is obtaining useful information about the likelihood function $\mathcal{L}$[211]. A likelihood is given by an expression of the type (16.20). The experimental vector of data can be encoded, in measured coefficients of the angular expansion on the sky $a_{\ell m}^{T,E,B}$nstead of spatial positions. Thus $\Delta = \{\Delta_{2,-2}, \Delta_{2,-1}\dots\}$ where a single $\Delta_{\ell m} = (a_{\ell m}^T, a_{\ell m}^E, a_{\ell m}^B)$. We had seen that the covariance matrix of the likelihood is the sum of a theoretical covariance C_S and a noise C_N. Both quantities are taken to be approximately diagonal in ℓ, m giving the expression (16.57). At a fixed ℓ we need to include the polarization-temperature cross-corrleations as in (16.68). Now, suppose that the experiment has measured $a_{\ell m}^{T,E,B}$. The experimental covariance of these quantities will be the *experimental* C_ℓ . We denote this matrix as $\Delta_i\Delta_j = \hat{C}_{ij}$ where the indexes i, j are a collection $i = \{\ell, m, X\}$ with $X = T, E, B$. On the other hand, we will denote the covariance that appears in the likelihood, $\bar{C}$. This is a theoretical covariance, which contains all the dependence on the cosmological parameters λ_α, plus the noise due to the experiment. The likelihood depends on the cosmological parameters through the matrix $\bar{C}(\lambda_\alpha)$. $\hat{C}$, which encodes the data, is an experimental quantity and is a constant with respect to the theoretical variables. With this definition, the exponent of the Likelihood function becomes

$$\Delta_i(\bar{C}^{-1})_{ij}\Delta_j = \mathrm{Tr}(\hat{C}\bar{C}^{-1}) \tag{16.70}$$

where the repeated indexes are summed over on the left hand side. Recall that the likelihood function is proportional to the probability density for the cosmological parameters λ_α, assuming the priors are uniform in the parameters. The overall proportionality con-

stant of the distribution is intractable, but largely useless. Then

$$\mathcal{L} \propto |\bar{C}|^{-\frac{1}{2}} \exp \mathrm{Tr}(\frac{1}{2}\hat{C}\bar{C}^{-1}) \tag{16.71}$$

where we denote $|\bar{C}| = \det C$. It is more useful to work with the log-likelihood. Conventionally the quantity $-2\ln\mathcal{L}$ is used to remove the prefactors and make the quantity convex at the absolute minimum (ie, the second derivative positive).

$$-2\ln\mathcal{L} = \sum_\ell (2\ell+1)\left[\mathrm{Tr}(\hat{C}_\ell \bar{C}_\ell^{-1}) + \ln|\bar{C}_\ell|\right] + \text{const} \tag{16.72}$$

A sum over ℓ, m comes from the fact that the total covariance matrix is block diagonal, with each block denoted by $\hat{C}_\ell, \bar{C}_\ell$ being of the form (16.68). The sum on m is trivial, which provides the factor $(2\ell+1)$. Note that if we choose $\bar{C}_\ell = \hat{C}_\ell$, we obtain $\sum_\ell (2\ell+1)[3 + \ln|\hat{C}|]$. Adding a multiplicative factor in the likelihood means adding a constant to the log-Likelihood. Therefore we will use a common (arbitrary!) normalization for the log-Likelihood, such that it is zero when $\bar{C} = \hat{C}$.

$$-2\ln\mathcal{L} = \sum_\ell (2\ell+1)\left[\mathrm{Tr}(\hat{C}_\ell \bar{C}_\ell^{-1}) + \ln\frac{|\bar{C}_\ell|}{|\hat{C}_\ell|} - 3\right] \tag{16.73}$$

We can do this since $\hat{C}_\ell$ is a constant with respect to the theoretical parameters. We point out that if one has more observables, for example a rotation CMB spectra then it is possible extend the covariance matrices to include these extra terms. The above form remains the same, except for a possible change of the desired normalization, which is uninfluential. In our case, by working out the trace term

$$-2\ln\mathcal{L} = \sum_\ell (2\ell+1) f_{sky}\left[\frac{D_\ell}{|\bar{C}_\ell|} + \ln\frac{|\bar{C}_\ell|}{|\hat{C}_\ell|} - 3\right] \tag{16.74}$$

with D_ℓ given by[169, 42]

$$D_\ell = \hat{C}_\ell^{TT}\bar{C}_\ell^{EE}\bar{C}_\ell^{BB} + \hat{C}_\ell^{EE}\bar{C}_\ell^{TT}\bar{C}_\ell^{BB} + \hat{C}_\ell^{BB}\bar{C}_\ell^{TT}\bar{C}_\ell^{EE} - \hat{C}_\ell^{BB}(\bar{C}_\ell^{TE})^2 - 2\hat{C}_\ell^{TE}\bar{C}_\ell^{TE}\bar{C}_\ell^{BB} \tag{16.75}$$

In the above, we added by hand the sky fraction of the experiment f_{sky} , using the same plausibility argument as when we discussed the experimental error on C_ℓ^{TT} (16.63). At fixed ℓ we can only observe $(2\ell+1)f_{sky}$ independent samples $a_{\ell m}^{T,E,B}$.

With this likelihood, we apply our MCMC method. Indeed, all the $\bar{C}_\ell$ can be calculated at a given point in the cosmological parameter space. By varying these parameters along all possible direction we would obtain the shape of the distribution exactly, in theory. In practice, due to the dimensionality of the problem, this is not feasible. This is where the MCMC helps us.

A Markov Chain Monte Carlo (MCMC) is an algorithm that produces a sequence of points which, after many iterations, converges to the desired distribution. It is a way to generate a large number N of points in the parameter space distributed as the likelihood. Then, taking as a uniform random variable the position i of the sequence, the point $\vec{\lambda}_i$ in the sequence is

a random variable distributed as the likelihood. Once we have a large enough set of points, we can study the distribution as if they were the real likelihood.

The particular algorithm we describe is the Metropolis-Hastings algorithm. In the following, x_i is a point in the parameter space we are searching. For example in cosmology $x_i = \{\lambda_1, \lambda_2 \dots\}_i$. The index i indicates the position in the sequence (the Markov chain). The Metropolis-Hastings algorithm requires an arbitrary *proposal density* $g(x^{'}|x)$, which need not be symmetric. $p(x)$ is the probability density we wish to sample. In cosmology it is the likelihood, and we will see that the overall normalization constant is irrelevant. Then

1. Initialize the algorithm by setting $i = 0$, $x_i = 0$.
2. Draw a proposal value y randomly from the proposal density $g(y|x_i)$.
3. Calculate the acceptance ratio

$$A(x_i, y) = \min\left(1, \frac{p(y)g(x_i|y)}{p(x)g(y|x_i)}\right) \tag{16.76}$$

4. Accept the proposal y with probability A, otherwise reject
   (a) Generate v uniformly in $[0, 1]$
   (b) If $v < A$ accept y and set $x_{i+1} = y$
   (c) If $v > A$ reject y and set $x_{i+1} = x_i$
5. Set $i \to i + 1$ and go to step 2. Conditions for breaking will be discussed.

Two points should be made in this argument regarding the implementation. First, it must be easy to sample from the proposal density $g(x^{'}|x)$. In theory one could sample from any density, but in practice this needs to be computationally inexpensive. The most common choice is to use a normal distribution centered on the old value: $g(x^{'}|x) \sim \mathcal{N}(x^{'}; x, \sigma)$ where the proposal density σ is arbitrary. However σ, as other choices for the algorithm, must be chosen according to the problem. The second note we make is that the functional form of the probability we wish to sample is needed. Any normalization constant is un-necessary, since only the ratio appears.

Why does this work? For a given x_i , the transition probability to go to x_{i+1} is given by $g(x_{i+1}|x_i)A(x_i, x_{i+1})$. The probability of this transition actually happening, is the transition probability multiplied by that of finding the chain in the point x_i in the first place, which we denote $\pi(x_i)$. The $\pi(x_i)$ distribution is the one which the algorithm is sampling. Eventually, this distribution converges to one where the transition from $x_i \to x_{i+1}$ is just as likely to happen as the reversed one $x_{i+1} \to x_i$. This is known as the *principle of detailed balance*. According to this, for i large enough the following equality must hold

$$\pi(x_i)g(x_{i+1}|x_i)\min(1, \frac{p(x_{i+1})g(x_i|x_{i+1})}{p(x_i)g(x_{i+1}|x_i)}) = \pi(x_{i+1})g(x_i|x_{i+1})\min(1, \frac{p(x_i)g(x_{i+1}|x_i)}{p(x_{i+1})g(x_{i+1}|x_i)}) \tag{16.77}$$

Since the second argument of the two min functions are one the inverse of the other, one the two min functions will return 1 and the other the second argument. Then, regardless of which min returns 1,

$$\pi(x_i)p(x_{i+1}) = \pi(x_{i+1})p(x_i) \tag{16.78}$$

or

$$\frac{\pi(x_i)}{\pi(x_{i+1})} = \frac{p(x_i)}{p(x_{i+1})} \tag{16.79}$$

This must hold for any x, thus it's clear that $\pi(x) = p(x)$. It can be shown that for well behaved distributions and proposal densities, the Metropolis-Hastings converges after an infinite number of iterations. In practice, we can't wait forever, so we need to make some choices to speed up conversion.

Let's consider the proposal density. When using a normal distribution, its width is denoted by the standard deviation σ. If we don't use a normal distribution we can still think of some typical width σ. Ideally, we want the point x_i to move around the likelihood function quickly enough so that the region where it has the most support is well covered by points on the chain. σ must not be too small. Indeed, if it is much smaller than the scale on which the likelihood changes, it will take a long time for the chain to move to different regions. In practice this means more *wall time* for the algorithm. Making σ too large however will cause the point x_i to jump wildly across the likelihood. This is likely to miss the important features of the likelihood.

An analogy for physicists. If you want measure a structure with linear scale $\sim s$ you need a probe with a typical "wavelength" $\sigma \sim s$. In cosmology the point x_i is the set of cosmological parameters λ_α and one can use a different width for the proposal density of each parameter σ_α. A good practice is to set the σ_α to some not too small fraction $\sim 1/2 \div 1/10$ of the expected standard deviation of the parameter (which is the "length" scale of the likelihood function). There is no prescription that works best every time, finding the optimal value for the width of the proposal density takes some intuition and understanding of the problem.

Our goal is to obtain N values x_i which are distributed as the likelihood. There are a few issues we must mention. Points on a chain x_i can depend on the *initial value* x_0 that was chosen to initialize the algorithm. The initial value x_0 is arbitrary. A best practice is to set it at some central value which we expect to be close to the "true" value. For example, we may set it to the best value of a previous experiment. A common saying is to use as x_0 any point you don't mind having in your sample. In any case, the first points on the chain will be heavily correlated with whatever initial choice is made and are not expected to represent the density $p(x)$ we wish to sample. To get around this, we set a *burn in* parameter N_b. N_b is the number of initial samples from the chain which we simply throw away at the end of the algorithm. Of course we don't want to throw away too much, as doing so costs computational time. There isn't a best prescription again. If the initial value x_0 is already in a high probability region, there is little sense in discarding too many initial values.

An issue related to this, is that successive points in the chain are correlated with one another. This also may introduce a statistical bias in the sampling. To get around this, one performs *thinning*. This is the practice of only selecting every n-th element of the sequence, with this n being the thinning factor. In cosmology n may also be of the order $\sim 10^2$ due to the high dimensionality of the problem. The best practice to thin the chain is to calculate the correlation of an element x_i with x_{i+n} $:< x_i x_{i+n} >$ as a function of n. Ideally we would like this to be zero. In practice the correlation function drops off with large n and we may select a large enough value for the tinning.

To reduce correlation among successive values, it is common to launch more than one chain $x_i^{(m)}$ and then combine them at the end. There is value to this, since separate chains can be processed in a parallel fashion, speeding up the computation.

The final question is the issue of *stopping the algorithm*. Indeed, in our description we did not specify when to stop the iterations. One condition is that N is large enough. How large depends on the problem. In general, one includes this as a stopping criteria to terminate the algorithm after a sufficiently long time.

Most importantly, we wish to be sure that the distribution has converged. We know from general theory that the algorithm will eventually converge (at infinite iterations) to $p(x)$. How can we determine how close or far we are from convergence? A common measure of statistical convergence is the *Gelman-Rubin diagnostic*[109]. This gives a quantitative indication of the convergence of the algorithm. Suppose we have run M chains, each of which had a *different* initial value x_0, with $2N$ sampled points per chain. We denote $x_i^{(j)}$ the i-th point of chain j. For the diagnostic, we do not consider the first N elements of every chain. We may calculate the sample variance of a single chain

$$s_j^2 = \frac{1}{N-1}\sum_{i=1}^{N}(x_i^{(j)} - \bar{x}^{(j)})^2 \tag{16.80}$$

where $\bar{x}^{(j)}$ is the sample mean of the chain j. The *mean of the variances* of the chains is

$$W \equiv \frac{1}{M}\sum_{j=1}^{M} s_j^2 \tag{16.81}$$

On the other hand we can define the quantity

$$\frac{B}{N} = \frac{1}{M-1}\sum_{j=1}^{M}(\bar{x}^{(j)} - \bar{\bar{x}})^2 \tag{16.82}$$

which is the variance of the means. Then an estimator of the variance is

$$\hat{\sigma}^2 = (1 - \frac{1}{N})W + \frac{B}{N} \tag{16.83}$$

Both W and $\hat{\sigma}^2$ are unbiased estimators of the variance of the distribution *if the chains have converged*. Otherwise, W, the mean variance of the chains, will underestimate the actual variance, because the chains have not explored all of parameter space. On the other hand, $\hat{\sigma}^2$, which is the variance of the chains combined, overestimates the real variance. This because the chains are assumed to have started from different points in parameter space and are exploring different areas of the space, thus the dispersion is higher than the convergence one. One defines the *potential scale reduction factor*

$$\hat{R} = \sqrt{\frac{\hat{\sigma}^2}{W}} \tag{16.84}$$

$\hat{R} > 1$ for the duration of the algorithm and will tend to 1 as the sampling converges.

Thus, a stopping criteria for the Metropolis-Hastings algorithm is when the value of $\hat{R}$ is within some tolerance factor, $\hat{R} < 1 + \epsilon$. How small does ϵ need to be? There is no magic prescription, but $\epsilon \sim 0.1$ is usually a decent choice, with smaller values being better.

Once the MCMC is over, and we apply thinning, burn in, and combine the chains, we have a sequence of points distributed as the likelihood. At this point getting error bars on the individual parameters, as well as cross-correlations, is a simple matter of counting a sample. Often, we *marginalize* on every parameter of the distribution to obtain the probability distribution for only one parameter. Formally this is

$$p_\alpha(\lambda_\alpha)d\lambda_\alpha = \int \prod_{\beta\neq\alpha} d\lambda_\beta p(\lambda_\gamma) \tag{16.85}$$

with no sum implied. For discrete samples, this is just a count of the samples that fall in some bin $(\lambda_\alpha, \lambda_\alpha + \Delta\lambda_\alpha)$. In the same way, one can marginalize on all but two variables to obtain the bivariate distribution of the two variables. Once we marginalize, we may want to quote a 68% or 95% confidence interval. We simply do this by finding an interval which contains that percentage of samples generated from the sequence.

The MCMC method can be used to forecast on future experiments, as well as determining the posterior distributions for the cosmological parameters after an experiment. To do this one needs a set of measured variances which are encoded in the observed covariance matrix $\hat{C}$. If the experiment is already done, these are the covariances found by the experimentally measured $a^X_{\ell m}$. Otherwise, we need to generate them using a fiducial model, as we did in the case of a Fisher analysis. In that case, as per equation (16.63), one generates theoretical values C_ℓ from a Boltzmann code, to which one adds noise[42]

$$N_\ell^{T,P} = w_{T,P}^{-1} e^{\ell(\ell+1)\sigma^2} \tag{16.86}$$

Then the MCMC is run. The complication is that at every value of the parameter space that we traverse with the algorithm, the theoretical variance plus noise $\bar{C}$ must be calculated with a Boltzmann code in order to sample the likelihood (16.74), which is the $p(x)$ of the Metropolis-Hastings algorithm. In practice, this process is very computationally expensive. The MCMC code is usually run on a supercomputer. The most used program to perform this analysis is CosmoMC[150], which implements the MCMC and gets its theoretical values by running the Boltzmann code CAMB[151] at each step.

An example of a posterior distribution, represented as a contour plot, using the MCMC algorithm together with the likelihood (16.74) is given by figure 17.2.

Part V

Constraints on light particles

17 Light particles in the primordial universe

17.1 Decoupling of a light relic

Let's consider a species of particle X in the primordial universe. This particle interacts with the regular standard model particles through some unspecified interaction. If it is relativistic at the decoupling temperature T_D , it will have a relic density comparable to that of photons and neutrino. In the radiation dominated era, it may well play an important part, being subdominant to photons and neutrinos but participating in the expansion much more than the rest of the standard model particles. We had discussed hot relics back in section 10.2. A recap of the result is that, if the particle is light enough, so that it is still relativistic today, it has a Bose-Einstein or Fermi-Dirac distribution with a temperature[143]

$$T_X = \left(\frac{g^*(T_\gamma)}{g^*(T_D)}\right)^{\frac{1}{3}} T_\gamma \tag{17.1}$$

In the same way, we derive its relation to the temperature of the neutrinos. These decouple at a temperature $T_{\nu D}$. We obtain simply

$$T_X = \left(\frac{g^*(T_{\nu D})}{g^*(T_D)}\right)^{\frac{1}{3}} T_\nu \tag{17.2}$$

The relativistic degrees of freedom are defined in (8.11) and we report them in table (17.1) for specific values. The total density in ultra-relativistic particles, excluding photons, is then

$$\rho_{ur} = (N_{eff} + \Delta N_{eff})\rho_\nu \tag{17.3}$$

where ρ_ν is the equilibrium density of one species of neutrino assuming perfect decoupling, that is with $T_\nu = (\frac{4}{11})^{1/3} T_\gamma$, and $g = 2$ in it. $N_{eff} = 3.046$ is the standard model value. It is slightly higher than the expected value of 3 due to the fact that the neutrinos have not completely decoupled when the electrons become non-relativistic and heat up the plasma. See the discussion in section 8.4. ΔN_{eff} encodes the density of the extra relativistic species

T_γ	g^*	Species (d.o.f)
$T_\gamma \gtrsim 200MeV$	106.75	All Standard Model (90 fermionic, 28 bosonic)
$T_\gamma \sim 1GeV$	61.75	$s, \bar{s}$(12), $d, \bar{d}$(12), $u, \bar{u}$(12),$\mu^\pm$(4), $e^\pm$(4), ν(6),γ(2)
$1MeV \lesssim T_\gamma \lesssim \Lambda_{QCD} \simeq 200MeV$	10.75	$e^\pm$(4), ν(6), γ(2)
$T \ll 0.5MeV$	3.36	ν(6, $T_\nu = (\frac{4}{11})^{\frac{1}{3}} T_\gamma$), γ(2)

Table 17.1: Relativistic degrees of freedom as a function of temperature, at specific values

X . Plugging in the numbers

$$\Delta N_{eff} = \left(\frac{g^*(T_{\nu D})}{g^*(T_D)}\right)^{\frac{4}{3}} \frac{g_X}{2} \begin{cases} \frac{8}{7} & \text{Boson} \\ 1 & \text{Fermion} \end{cases} \tag{17.4}$$

where g_X is the number of degrees of freedom of the particle X . There is an overall factor $8/7$ if X is a boson. As we discussed for light relics, we don't expect this value to be modified by the specific details of the decoupling. We may take it at face value.

We observe a few details. First of all, if the particle decouples after the QCD phase transition, which happens at $T_{QCD} \sim \Lambda_{QCD} \simeq 200MeV$, the degrees of freedom are about six times less, which means $\Delta N_{eff} \sim 10$ times higher for a species which decouples after the phase transition rather than before. We won't discuss what happens to a species decouples during the transition, since the details in that regime are very messy. The QCD phase transition happens when the running coupling constant of the strong interactions becomes extremely large and the perturbative regime breaks down. Indeed, Λ_{QCD} is a Landau pole of the theory, in the infrared regime. Before the phase transition, the relevant degrees of freedom are the quarks, which form a quark-gluon plasma. The quarks may be thought as nearly free and in the usual thermal equilibrium. During the transition, the theory is highly non-perturbative and the details are not clear. The fundamental degrees of freedom of the theory become the hadrons. At the end of the transition the temperature is much lower than any hadron (the lightest is the pion $m_{\pi^0} = 135MeV$) and the interactions drive the plasma back to equilibrium. Thus, from a cosmological point of view, the QCD phase transition is relatively painless, the only significant effect being the change in degrees of freedom.

Returning to our relic, the importance of the QCD phase transition is that any ultra-relativistic particle freezing out after that time is excluded by the latest experimental data by PLANCK. We will see that N_{eff} has a measurable effect on the small scale structure of the CMB. The 95% limits on the parameter by PLANCK are $N_{eff} = 3.00 \pm 0.5$, which excludes any value of ΔN_{eff} larger than 0.5. So any plausible light relic must decouple before the QCD phase transition.

The smallest possible value of N_{eff} is obtained if the decoupling happens when all the standard model particles, up and including the top quark, are relativistic. In that case we have that

$$\Delta N_{eff} = \begin{cases} 0.027 & \text{Scalar} \\ 0.047 & \text{Spin } \frac{1}{2} \text{ fermion} \\ 0.053 & \text{Massless vector} \end{cases} \tag{17.5}$$

As a scientific goal, it is crucial that the next generation of experiments be able to constrain N_{eff} with an error of ~ 0.03[18, 53]. This may indeed prove crucial to either discover or exclude a large class of models. Indeed, if this goal is achieved, we could effectively be studying physics at temperatures much larger than the electroweak scale $T_{EW} \sim 200GeV$. Even a species which decouples at temperature order of magnitudes larger would have an effect. In the same manner, a non detection may put many constraints on models of new physics.

The constraints on N_{eff} from cosmology arise from three observables: abundances from BBN, CMB power spectrum and matter power spectrum[19]. The next generation of CMB experiments, such as CMB-S4[53] have the ambition of reaching a level of sensitivity to probe these lightest relics.

17.2 Effect of light relics on the CMB through diffusion damping

A light relic contributing to a change in the effective number of neutrinos will change the power spectrum of the CMB. The effect of the light particles, neutrinos plus a new species X, is to drive the expansion of the universe faster during the radiation dominated era. They affect the primordial plasma of baryons and photons through their gravitational interactions alone. For this reason, we expect X and ν to produce the same effects. An increase in the density of radiation before recombination has several effects on the early universe. The most obvious is that it changes the scale factor a_{eq} of matter-radiation equality; it affects the damping of the acoustic oscillations of the baryon-photon fluid; and changes the sound horizon $r_s(\tau_{LSS})$. The scale factor of matter radiation equality, given by (7.95), is very important in the evolution of matter perturbations, as the metric potentials evolve very differently in the two eras. The sound horizon is well measured by the position of the CMB power spectrum peaks.

At first, all modes are outside the horizon, $k\tau \ll 1$ and, in the conformal Newtonian gauge, approximate solutions are given by equations (13.325)-(13.329), where the matter perturbations and metric potential are constant. Outside the horizon they are frozen. This is of course gauge dependent, in the synchronous gauge the out of horizon modes are growing, but once they re-enter the horizon we expect the modes to evolve in the same way. Any mode that enters the horizon $k\tau \sim 1$ during radiation dominated era will have its metric and matter potentials damped. Physically, the free-streaming radiation washes out any possible growing concentrations of matter and the matter potentials ϕ, ψ are diluted away by the expansion of the universe[80]. On the other hand, a mode that enters the horizon during the matter dominated era will see its matter perturbations grow. This is the origin of large scale structures today.

This being said, we understand that the effect of a_{eq} is very important on the CMB spectrum and matter power spectrum. For this reason, the value of a_{eq} is well constrained by experiment. Does this mean we have fixed ΔN_{eff} to a very precise value? It does not. Indeed if we change N_{eff} *and the matter density* (dark matter+baryons) by an equivalent amount, we can keep a_{eq} at the same value and remain consistent with experiment. We say there is a degeneracy between the parameters Ω_r and Ω_m from the experiment, since changing *both together* can give the same experimental results. We could measure only some combination. Fortunately, there exists an effect due to the change of Ω_r, equivalently N_{eff}, that is not replicable with a change of the matter density. This is said to *break the degeneracy*. The effect in question is the diffusion damping[123, 29].

We had discussed the tight coupling approximation in section 13.14. Assuming the baryons and photons formed a single fluid, due to the high value of $\dot{\tau} = -an_e\sigma_T$, we had shown that an oscillatory solution (13.296) can be found for the common fluid. The solutions are the acoustic oscillations, and are responsible for the oscillatory nature of the CMB spectrum. We had done a first order calculation in $\dot{\tau}^{-1}$ which neglected the *slip* between the

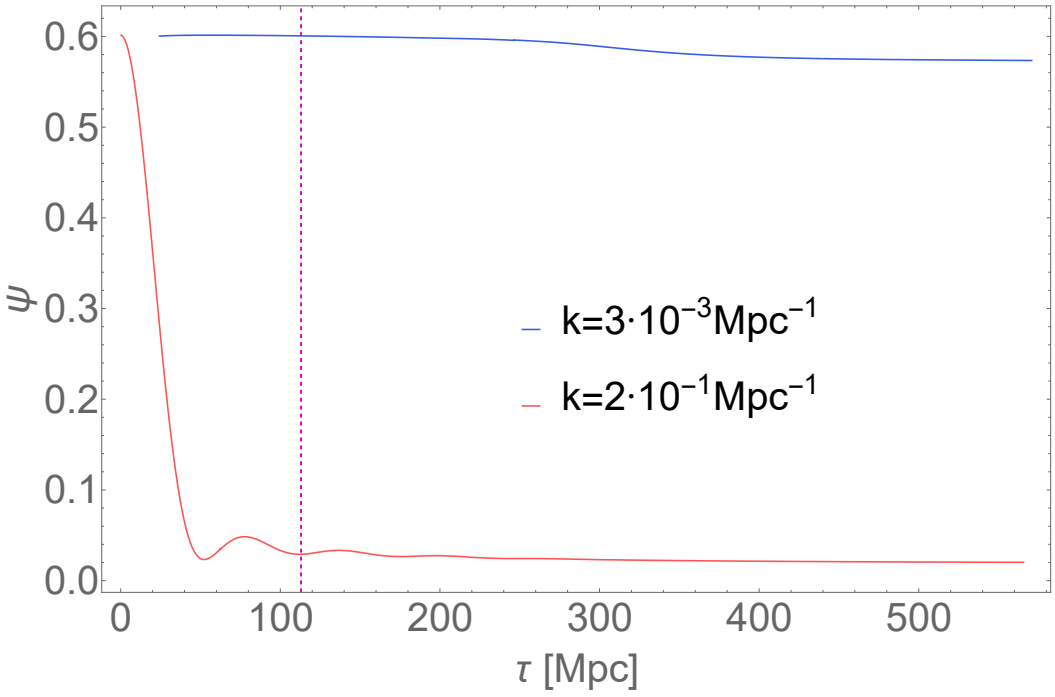

Figure 17.1: A numerical calculation of the evolution of the metric potential ψ for two different values of k. The dashed purple line is the moment of matter radiation equality $\tau_{eq} \simeq 110 Mpc$. The mode $k = 2 \cdot 10^{-1} Mpc^{-1}$ (red line) enters the horizon well before τ_{eq} and the solution is oscillatory and damped. On the other hand, a mode which enters the horizon well in the matter dominated era (blue line) has a roughly constant amplitude before and after horizon crossing. During horizon crossing the amplitude reduces by a factor $\sim \frac{9}{10}$. The calculations are performed with the CLASS Boltzmann code, assuming the best fit ΛCDM parameters.

photon and baryon fluid, the difference in the velocities $\theta_{\gamma,b}$. We will now consider this, and find that the acoustic oscillations are damped with a characteristic *damping scale* k_D^{-1}. This scale,will turn out to depend on the expansion rate of the universe during the radiation dominated era, and so on N_{eff}.

We will work with modes that enter the horizon during the radiation dominated era. Which means at small scales k^{-1}, or at large values of k (which implies the effects will be seen at large ℓ in the CMB power spectrum). First, we want to show that the metric potentials decay in a radiation dominated era. Working in the conformal-Newtonian gauge we take the Time-Time Einstein equation (13.117) and the time-space equation (13.123). These can be linearly combined to get an algebraic equation for ϕ

$$-k^2\phi = \frac{3}{2}\mathcal{H}^2(\delta_r + 4\mathcal{H}\frac{\theta_r}{k^2}) \tag{17.6}$$

where δ_r and θ_r are the total radiation perturbations, $\delta_r = \delta_\gamma + \delta_\nu$ and $\theta_r = \theta_\gamma = \theta_\nu$. In deriving this equation the first Friedmann equation (7.25) is used $\mathcal{H}^2 = \frac{8\pi G}{3}\rho_r$, $\rho_r = \rho_\gamma + \rho_\nu$ being the total radiation density at zero order. Since we are in the radiation dominated era we neglect the density of baryons and dark matter. At this stage we don't care yet whether ρ_ν counts any light particles or not. Turning around the equations

$$\delta_r = -\frac{2}{3}\frac{k^2\phi}{\mathcal{H}^2} - \frac{4\mathcal{H}\theta_r}{k^2} \tag{17.7}$$

$$\dot{\delta}_r = -\frac{2k^2}{3\mathcal{H}^2} - \frac{4}{3}\frac{k^2\phi}{\mathcal{H}} + 4\frac{\mathcal{H}^2}{k^2}\theta_r - 4\frac{\mathcal{H}}{k^2}\dot{\theta}_r \tag{17.8}$$

To obtain the second equation we have used the second Friedmann equation. In the radiation dominated era we have $(\rho_r + 3P_r) = 2\rho_r$. Using this, the we obtain $\dot{\mathcal{H}} = -\mathcal{H}^2$.

We use these quantities in the photon and neutrino Boltzmann equations. The photon Boltzmann equations are given in (13.253)-(13.256) and the neutrino equations in (13.169)-(13.175). The equations are of the same form, except for the Compton collision term. Now we will argue that equations for the moments $\ell = 0, 1$ of the two species take the same form. Indeed, the scattering term $\dot{\tau}$ is very large which implies that $\theta_b \simeq \theta_\gamma$. We can neglect the collision term $\dot{\tau}(\theta_\gamma - \theta_b)$ in the photon Boltzmann equations. In terms of the tight coupling approximation, this means we are working at order zero in $\frac{1}{\dot{\tau}}$. Of course, this approximation is not good enough for estimating photon perturbations, but it simplifies the analysis of the evolution of the metric potential. For $\ell > 2$, the photon perturbations are driven to zero by the tight coupling with the baryons. The neutrino higher moments can instead be neglected as they are much smaller than the others outside the horizon. With this argument, we can treat the photon-neutrino as one fluid with only one over-density and one velocity term. The evolution equations are given by

$$\dot{\delta}_r = -\frac{4}{3}\theta_r + 4\dot{\phi} \tag{17.9}$$

$$\dot{\theta}_r = \frac{k^2}{4}\delta_r + k^2\psi \tag{17.10}$$

Another consequence of ignoring higher moments of the neutrino distribution is that, through the longitudinal traceless space-space Einstein equation (13.118), we get

$$\phi = \psi \tag{17.11}$$

Indeed, as a general rule of thumb the metric potentials in the Newtonian gauge can be taken to be equal if anisotropic neutrino stresses can be neglected. Next, we plug in the expressions of δ_r in terms of the metric (17.7)-(17.8) into the Boltzmann equations (17.9)-(17.10) and obtain

$$-\frac{\mathcal{H}}{k^2}\dot{\theta}_r + \theta_r\left(\frac{1}{3} + \frac{\mathcal{H}^2}{k^2}\right) = \dot{\phi}\left(1 + \frac{k^2}{6\mathcal{H}^2}\right) + \frac{k^2\phi}{3\mathcal{H}^2} \tag{17.12}$$

$$\dot{\theta}_r + \mathcal{H}\theta_r = k^2\phi\left(1 - \frac{k^2}{6\mathcal{H}^2}\right) \tag{17.13}$$

The two equations can be combined by multiplying the second by $\frac{\mathcal{H}}{k^2}$ and eliminating the derivative $\dot{\theta}_r$

$$\theta_r = \frac{k^2}{2\mathcal{H}^2}(\dot{\phi} + \mathcal{H}\phi) \tag{17.14}$$

which can be differentiated to obtain

$$\dot{\theta}_r = \frac{k^2}{2\mathcal{H}^2}(\ddot{\phi} + 3\mathcal{H}\dot{\phi} + \mathcal{H}^2\phi) \tag{17.15}$$

Plugging these last two into (17.13), we obtain a second order differential equation

$$\ddot{\phi} + 4\mathcal{H}\dot{\phi} + \frac{k^2}{3}\phi = 0 \tag{17.16}$$

In the radiation dominated era $\mathcal{H} = \tau^{-1}$. Note that this equation already supposed we are in a radiation dominated by the use of the Friedmann equations. With the explicit expression of $\mathcal{H}$, this equation actually has an analytical solution. Before giving it, let's see what we can guess by just looking at it. Since $\mathcal{H} = \tau^{-1}$ and we can suppose the derivative $\dot{\phi} \sim \frac{\phi}{\tau}$. The ratio of the Hubble friction term $\mathcal{H}\dot{\phi}$ to the oscillating term $k^2\phi/3$ is $\sim (k\tau)^{-2}$. Therefore, while the mode is well out of the horizon $k\tau \ll 1$, the friction term is very large. The Hubble friction keeps the mode frozen to its initial value. On the other hand, when the mode is well inside the horizon $k\tau \gg 1$, the oscillating term dominates and we can expect the solution to be damped oscillations.

The analytical solution can be found by passing to the auxiliary variable $u = \phi\tau$ and finding that the equation reduces to the Spherical Bessel equation of order 1 . u can be written as the spherical Bessel function of order 1 (plus a spherical Neumann function which blows up at early times and we may discard with proper initial conditions). The algebra is straightforward and is an application of the properties of the spherical Bessel functions. The result is[80]

$$\phi = 3\phi_i \left(\frac{\sin\frac{k\tau}{\sqrt{3}} - \frac{k\tau}{\sqrt{3}}\cos\frac{k\tau}{\sqrt{3}}}{(\frac{k\tau}{\sqrt{3}})^3} \right) \tag{17.17}$$

One can verify that the solution satisfies the above equation with initial condition $\phi(\tau = 0) = \phi_i$ and $\dot{\phi}(\tau = 0) = 0$. It has the properties we mentioned above, namely that for $k\tau \to 0$ it is constant and for large values of $k\tau$ it is a damped oscillatory solution as $\sim \cos(\frac{k\tau}{\sqrt{3}})/(k\tau)^2$. A numerical calculation of the matter potential is in figure 17.1.

This is an interesting result by itself. We want to use this fact, for modes well inside the horizon during the radiation dominated era. We are therefore considering large enough values of k, or small scales, such that $k\tau > 1$ for $\tau < \tau_{eq}$: $k \gtrsim \frac{1}{\tau_{eq}}$. We are going to examine the tight coupling limit for these modes, when the matter potential may be neglected. We already know that the solutions should have oscillatory form thus it makes sense to suppose the time dependence of the variables δ_γ, θ_γ, $F^{(0)}_{\gamma 2}$, θ_b is of the form

$$\sim \exp(i\int_0^\tau \omega(\tau')d\tau') \tag{17.18}$$

Indeed, in the absence of metric perturbations this is the time dependence of the explicit solution (13.296) we had obtained in the tight-coupling limit. In that case $\omega = kc_s$, where c_s is the speed of sound given by (13.291)

$$c_s^2 = \frac{1}{3}\frac{R}{1+R} \tag{17.19}$$

R being the ratio of densities (13.284) $R = \frac{4}{3}\frac{\rho_\gamma}{\rho_b}$. Now, we will be doing a higher order approximation with $\omega = kc_s + \delta\omega$. It will turn out that the extra term is pure imaginary and corresponds to an extinction factor of the wave, also known as the damping term.

Let's take the Boltzmann equations for photons for the moments $\ell \leq 2$ (13.253)-(13.255) and the baryon velocity equation (13.273). Neglecting the metric potentials and replacing the conformal time derivatives $\frac{d}{d\tau} \to i\omega$

$$i\omega\delta_\gamma = -\frac{4}{3}\theta_\gamma \tag{17.20}$$

$$i\omega\theta_\gamma = \frac{k^2}{4}\delta_\gamma - \frac{k^2}{2}F_{\gamma 2}^{(0)} + \dot{\tau}(\theta_\gamma - \theta_b) \tag{17.21}$$

$$i\omega F_{\gamma 2}^{(0)} = \frac{8}{15}\theta_\gamma + \frac{9}{10}\dot{\tau}F_{\gamma 2}^{(0)} \tag{17.22}$$

$$i\omega\theta_b = -\mathcal{H}\theta_b - R\dot{\tau}(\theta_\gamma - \theta_b) \tag{17.23}$$

We have neglected the polarization of the photons. This is a set of four *homogeneous* algebraic equations in four variables. The solution is the null vector unless they are not all independent. This means that the parameter ω will have to take on specific values to make the determinant of the matrix of coefficients zero. Then, there will be one free parameter: the amplitude of the oscillations.

There is no conceptual problem to solving the system exactly, but we derive a solution up to order $(\frac{1}{\dot{\tau}})^2$. In the equation for $F_{\gamma 2}^{(0)}$ the right hand side term $i\omega F_{\gamma 2}^{(0)}$ is negligible compared to $\frac{9}{10}\dot{\tau}F_{\gamma 2}^{(0)}$, due to the large value of $\dot{\tau}$ and the fact that we already expect $F_{\gamma 2}^{(0)}$ to be $o(\theta_\gamma/\dot{\tau})$. Furthermore, in the last equation, for θ_b, the $\mathcal{H} \ll \omega$ and so the term $\mathcal{H}\theta_b$ can be neglected. This can be inverted to

$$\theta_b = \frac{1}{1 - \frac{i\omega}{R\dot{\tau}}}\theta_\gamma = (1 + \frac{i\omega}{R\dot{\tau}} - \frac{\omega^2}{R^2\dot{\tau}^2})\theta_\gamma + o(\frac{1}{\dot{\tau}})^3 \tag{17.24}$$

Now we plug expressions of δ_γ, $F_{\gamma 2}^{(0)}$ and θ_b written in terms of θ_γ into equation (17.21). With trivial algebraic manipulations we get the form

$$\omega^2\theta_\gamma(1 + \frac{1}{R}) - \frac{k^2}{3}\theta_\gamma = -\frac{i\omega}{\dot{\tau}}(\frac{8}{27}k^2 + \frac{\omega^2}{R^2})\theta_\gamma \tag{17.25}$$

As expected, the value of the photon velocity is not constrained by the equations so long that ω takes on the correct value. We simplify $\theta_\gamma \neq 0$. At lowest order in $\frac{1}{\dot{\tau}}$, dropping the right hand side, $\omega = kc_s$, which is the solution we already knew. We can plug the first order solution in the right hand side to obtain the second order equation

$$\delta\omega = -\frac{ik^2}{2\dot{\tau}}\frac{R}{1+R}(\frac{8}{27} + \frac{c_s^2}{R^2}) \tag{17.26}$$

As we had anticipated, the correction is pure imaginary. It represents an extinction factor for the wave. The time dependence of the solution is expressed through

$$\delta_\gamma \sim \exp\left(i\int kc_s(\tau)d\tau\right)\exp\left(-\frac{k^2}{k_D^2}\right) \tag{17.27}$$

where we defined the damping scale k_D^{-1}[127]

$$\frac{1}{k_D^2} = \int_0^\tau d\tau^{'} \frac{1}{2an_e\sigma_T} \frac{R}{1+R}(\frac{8}{27} + \frac{c_s^2}{R^2}) \tag{17.28}$$

By changing variables $d\tau^{'} = \frac{da}{\mathcal{H}a} = \frac{da}{Ha^2}$, where H is the Hubble factor in coordinate time.

$$\frac{1}{k_D^2} = \int_0^\tau \frac{da}{2a^3Hn_e\sigma_T} \frac{R}{1+R}(\frac{8}{27} + \frac{c_s^2}{R^2}) \tag{17.29}$$

This last form makes explicit the dependence of the damping scale to the expansion rate during the radiation dominated era since $H \propto \sqrt{\rho_r}$.

We have proven an effect exists on the CMB spectrum which is due to a change in N_{eff} that can't be replicated by changing the matter content, thus *breaking the degeneracy*. Naively one may think that increasing N_{eff} , increases H and therefore decreases the damping scale k_D^{-1}, resulting in less damping in the CMB power spectrum at high ℓ (ie *more power at large* ℓ). This is not wrong per se. However, an analysis will usually consider a situation where N_{eff} changes at fixed values of the first acoustic peak of the CMB power spectrum, which is very well constrained by experiment. When one changes N_{eff} purely, disregarding other parameters, the change in damping scale is subdominant to the change in the sound horizon $r_s = \int d\tau\omega = \int \frac{da}{aH}\omega$, since this has a dependence of $\sim H^{-1}$ instead of $\sim H^{-1/2}$. If we change other parameters to keep the sound horizon fixed (which determines the locations of the peaks through equation (13.296)), then it turns out that increasing N_{eff} in *that particular combination increases the damping at high* ℓ (ie less power at large ℓ)[123, 29].

Another way to see this, is that increasing N_{eff} on its own moves the spectrum $\ell \sim 200$ more than it does at $\ell \sim 1000$. When other parameters are changed to compensate for the shift at $\ell \sim 200$, which would violate experimental data, this brings the spectrum at $\ell \sim 1000$ down by a larger amount then it was raised. The result, is a power spectrum which is lower than it was before the change in N_{eff}, at large ℓ. This is an examples of the subtleties in the interpretation of physical effects due to the choice of parametrization which we had discussed in section (14.4).

There is an important issue to note in (17.28) which has relevance to the sensitivity to N_{eff} in experiments. It is a degeneracy with the number of electrons n_e[53, 42, 29]. By degeneracy we mean that a change in N_{eff} could be mimicked by a change in n_e , which results again in a problem of figuring out how to separate the effects. Since we are well before recombination, we might expect the number of free electrons to be the number of baryons n_b. This is not true, because Big Bang Nucleosynthesis has converted some protons to Helium. Indeed, the mass fraction of Helium (12.12) is $X_{^4H_e} \simeq 0.23$. A ^4He nucleus is produced by a total of 2 protons, 2 neutrons. For each Helium nuclei produced the total baryon number decreases by 3, while the electron number does not change. Before BBN $n_e = n_p$, by charge neutrality, and, of the n_p, neutrons $2n_{He}$ become Helium. The total number of free electrons long after BBN is

$$n_e = n_b + n_{H_e} = n_b(1 + \frac{X_{^4H_e}}{4}) \tag{17.30}$$

Experiment	θ	$T^{-1}w^{-\frac{1}{2}}[\mu K \cdot \text{arcmin}]$	f_{sky}	ℓ_{min}	ℓ_{max}
CMB-S4[53]	$3'$	1	0.4	5	3000
CMB-S4+	$1'$	0.5	0.4	2	5000

Table 17.2: Experimental configurations considered for the determination of N_{eff} and n_{run}. θ, the full width at half maximum of the beam, is defined in (16.36) . The power noise is as defined in (16.56) .

This implies there is a degeneracy between N_{eff} and the primordial Helium fraction from BBN. If one can be sure of the value of the fraction, from theory or measurement, than this can be used as a parameter. However, if we want to measure the mass fraction from the CMB we must be careful with this degeneracy. It must also be noted that the fraction of Helium is uncertain, both in theory and experiment.

17.3 Impact of theoretical assumptions in the determination of the neutrino effective number

The experimental goal of the next generation of CMB mapping experiments is to reach a sensitivity on N_{eff} of the order ~ 0.03[53]. With this sensitivity, there is the possibility to exclude a large class of models or, hopefully, to discover new physics[18]. Planned experiments, such as CMB-S4, have forecasted their error on N_{eff} to be of this order of magnitude. We will now point out that this level of sensitivity is affected by two important assumptions: the value of the running of the spectral index (14.27) and the lifetime of the neutron. We analyze the bias introduced by these assumptions by running a MCMC forecast, as explained in section 16.5.

17.3.1 Impact of the running of the scalar index

The primordial power spectrum P_ψ is characterized by a spectral index which breaks the scale invariance. The best fit to PLANCK data is $n_s = 0.963 \pm 0.011$ at 95% confidence level[60]. In addition there may be a yet unmeasured value of the running of the spectral index $n_{run} = \frac{dn}{d\ln k}$[37, 141, 46, 70, 86]. The running of the spectral index is present in any inflation model. In a Starobinsky model of inflation a typical value of n_{run} is[202]

$$n_{run} \simeq -\frac{1}{2}(1 - n_s)^2 \tag{17.31}$$

Consistently with the value reported by PLANCK[61], the running of the spectral index may be of the order $n_{run} \sim 0.001$. Another goal of CMB-S4 is indeed to measure n_{run} with this sensitivity[53]. Since its relationship to n_s depends on the particular inflation model, this might probe a vast number of models of inflation. Crucially, an analysis of CMB-S4, using the same MCMC methods as we do, was performed by *either fixing* n_{run} *or fixing* N_{eff}. The possible degeneracies between the two values were not studied. Indeed, one can use n_{run}, or even n_s for that matter, to change the form of the spectra at large ℓs, all else being equal. The experimental signature of a small n_{run} and a small N_{eff} in the CMB power spectrums may be similar. This degeneracy can decrease the sensitivity to both parameters.

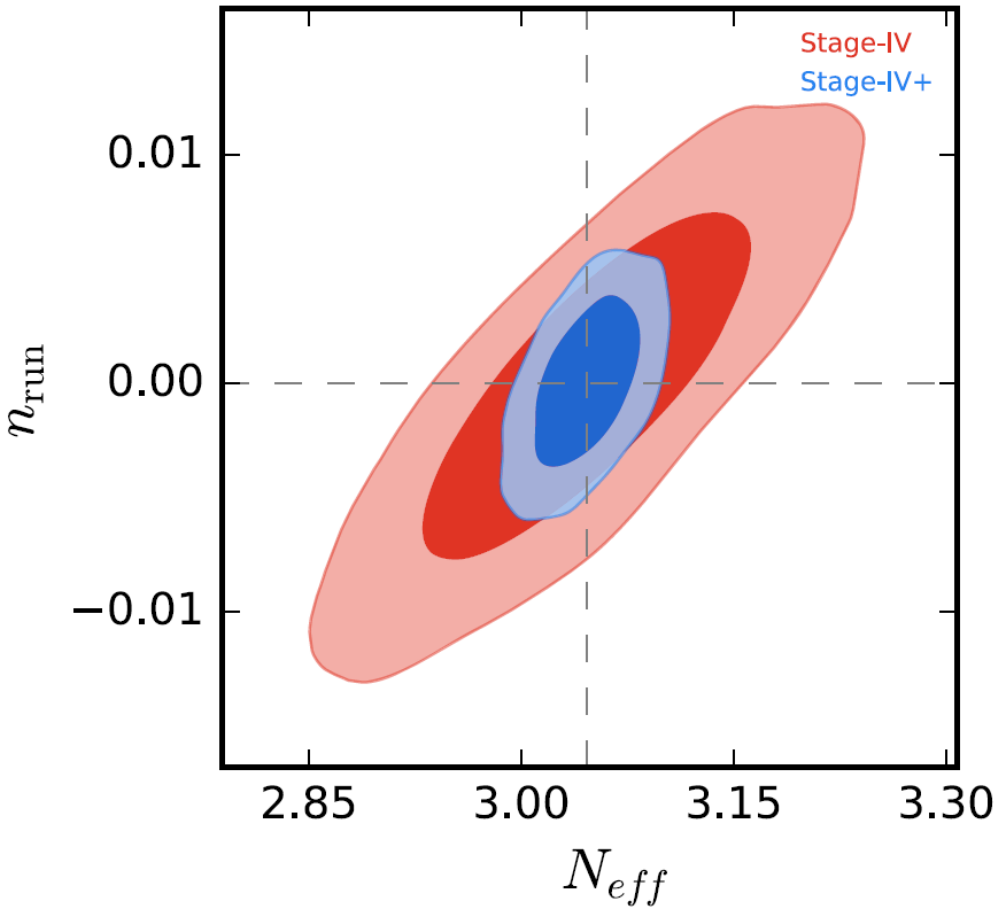

Figure 17.2: Contour plots forecasted for the CMB-S4 (Stage IV)[53] and CMB-S4+ (Stage IV+) experiments at 68% and 95% C.L. in the $n_{run} - N_{eff}$ plane. A degeneracy between the two parameters is present and is more pronounced for CMB-S4.

Case	N_{eff} - CMB-S4	N_{eff} - CMB-S4+
Varying n_{run}	3.049 ± 0.076	3.048 ± 0.024
$n_{run} = 0$	3.048 ± 0.043	3.047 ± 0.021
$n_{run} = 0.002$	3.019 ± 0.043	3.035 ± 0.021
$n_{run} = 0.004$	2.996 ± 0.044	3.024 ± 0.021
$n_{run} = -0.002$	3.074 ± 0.044	3.056 ± 0.021
$n_{run} = -0.004$	3.098 ± 0.044	3.071 ± 0.019

Table 17.3: Constraints at 68% C.L. for N_{eff} assuming different values for the running. If we include n_{run} in the analysis (first row) then the forecasted error on N_{eff} increases by $\sim 75\%$ for the CMB-S4 experiment ($\sim 17\%$ for the CMB-S4+ experiment) with respect to the no-running case (second row onwards). On the other hand, if there is a running present in the mock data, which is not accounted for in the analysis, the mean value of N_{eff} shifts by approximately $\Delta N_{eff} \sim -12 n_{run}$ for CMB-S4 and $\Delta N_{eff} \sim -5 n_{run}$ for CMB-S4+ experiment. In all mock data $N_{eff} = 3.046$ is used.

We run a MCMC on mock data generated with the experimental configuration of CMB-S4. In addition, we consider a possible futuristic experimental configuration, in the interest of long term forecasting. The experimental parameters are reported in table 17.2. The results are tabulated in 17.3.

We find that if n_{run} is allowed to vary in the analysis as a fundamental parameter, the 68% C.L. becomes about $\sim 75\%$ larger than the case where it is not[42]. On the other hand, if n_{run} is not included in the analysis, but the mock data contains a non-zero value, the reconstructed mean of N_{eff} is shifted, which could give a false positive detection. We also find that the degeneracy is improved for the futuristic CMB-S4+ experiment, indicating that if we probe values of ℓ up to 5000 the physical effects of the two parameters may be discerned better.

To make the degeneracy more evident, we plot the $2D$ contour plots, in figure 17.2, of the posterior at 68% and 95% C.L. in the $N_{eff} - n_{run}$ plane. The degeneracy is apparent in the CMB-S4 experiment. However it is ameliorated in the case of CMB-S4+. We also plot the posterior distributions obtained for N_{eff} when n_{run} is not considered in the analysis but included in the fiducial model. These are plotted in figure 17.3. Notice the different x-axis scales, which again indicates the problem is ameliorated for a futuristic configuration.

The degeneracy between the posterior distributions of N_{eff} and n_{run} is therefore an experimental challenge to keep in mind.

17.3.2 Impact of the lifetime of the neutron

We had noticed in section 17.2, that the effect of N_{eff} on the CMB anisotropies is degenerate with a change in the number of free electrons, through the primordial Helium density, as given by (17.30). Even the most optimistic experimental configuration cannot hope to measure N_{eff} with a sensitivity of about ~ 0.03 without assuming the validity of Big Bang Nucleosynthesis[53]. In a typical Boltzmann code, such as CLASS, the Helium mass fraction is interpolated from a pre-computed table which contains the dependency $X_{^4H_e}(N_{eff})$. In CLASS the table used is generated from the Parthenope code[67]. In the generation of this code a neutron lifetime (quoted by the Particle Data Group (PDG)[99])

$$\tau_n = 880.3s \tag{17.32}$$

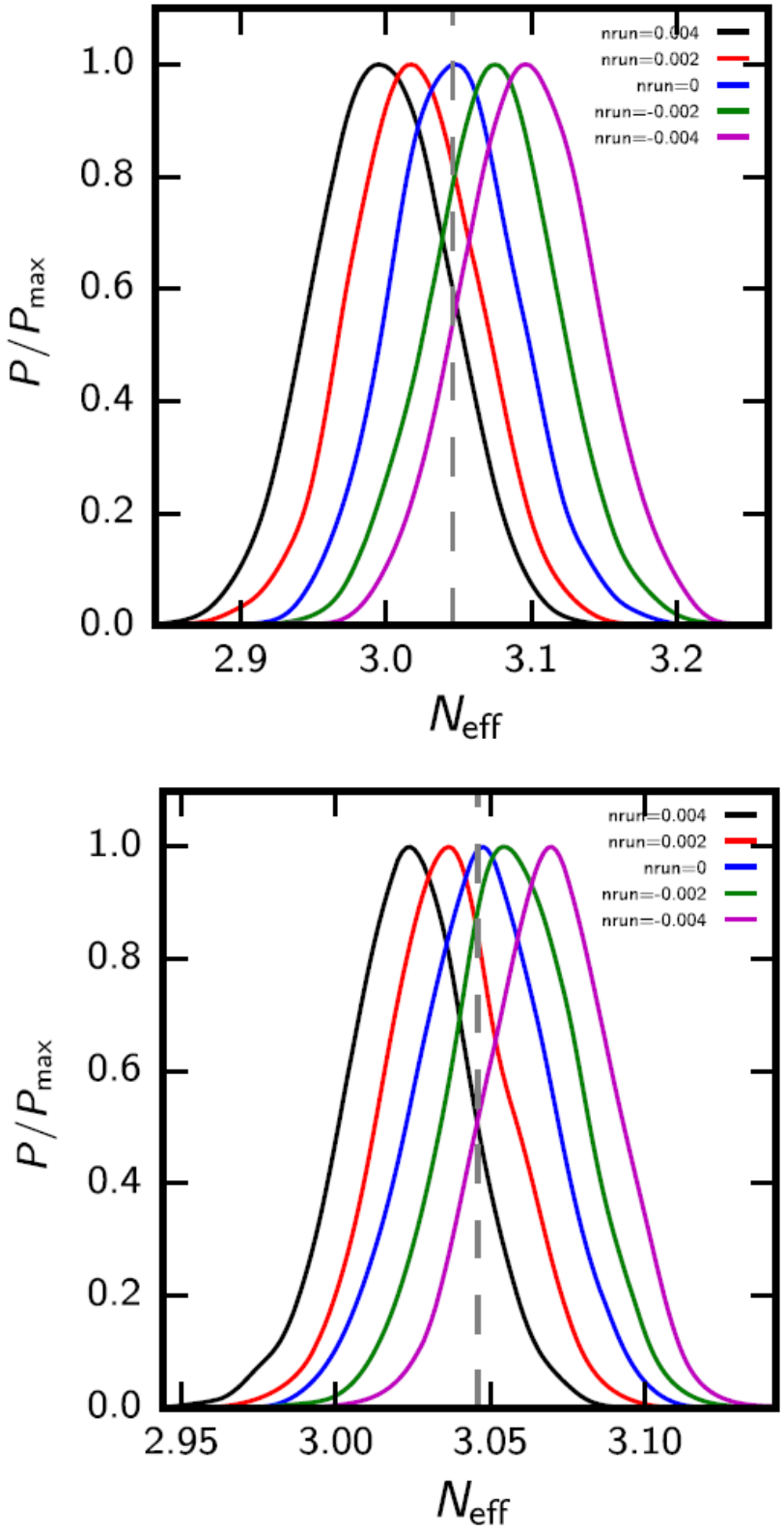

Figure 17.3: Posterior distributions on N_{eff} assuming $N_{eff} = 3.046$ and different values for n_{run} for the fiducial model but performing an analysis including n_{run}. In the top panel are the posteriors from CMB-S4, while in the bottom we have the posteriors for a CMB-S4+ experiment. As we can see, not accounting for a negative running could produce a significant shift in the recovered values of N_{eff}.

Case	N_{eff} (CMB-S4)	N_{eff} (CMB-S4)
$\tau_n = 880.3s$	3.048 ± 0.043	3.047 ± 0.021
$\tau_n = 888.0s$	3.062 ± 0.040	3.064 ± 0.021
$\tau_n = 877.0s$	3.039 ± 0.041	3.037 ± 0.020

Table 17.4: Constraints at 68% C.L. for N_{eff} on fiducial data generated with different neutron lifetimes, if the analysis uses the PDG value.

is assumed.

Neutron lifetime affects the final Helium abundance, as explained in section 12. A shorter lifetime implies a smaller number of neutrons at the beginning of BBN which implies a lower amount of Helium, since nearly all available neutrons are very quickly bound in the Helium nuclei. Constraining the neutron lifetime has been difficult and the measurement may be affected from an unknown systematic[186, 215].

There is, in fact, a long standing tension between values of the neutron lifetime obtained through two different techniques. In one technique, known as the "beam" method, the number of β electrons in a passing beam of cold neutrons is counted[220]. In the second technique, the "bottle" method, a number count of surviving neutrons in a container is measured[193, 95]. The two techniques provide different values

$$\tau_n^{\text{bottle}} = (878.5 \pm 0.7(\text{Stat}) \pm 0.3(\text{Sys}))\, s \tag{17.33}$$

$$\tau_n^{\text{beam}} = (887.7 \pm 1.2(\text{Stat}) \pm 1.9(\text{Sys}))\, s \tag{17.34}$$

These two measurements are discrepant to about ~ 3.9 standard deviations. It makes sense to study what bias the assumption of a known value of the neutron lifetime has on the determination of N_{eff}. A numerical fit of Helium mass fraction as a function of the neutron lifetime is given by

$$X_{^4He}(\tau_n) = \left(\frac{\tau_n}{880.3s}\right)^{0.73} X_{^4He}(\tau_n = 880.3s) \tag{17.35}$$

We therefore produce mock data with the two extremal values related to the "bottle" and "beam" measured values, $\tau_{\text{low}} = 877.0s$ and $\tau_{\text{high}} = 888.0s$, as well the PDG value $\tau_n = 880.3s$. We also use $N_{eff} = 3.046$, the standard model value, in the production of the mock data. Then, we proceed to analyze this data assuming the ΛCDM model with a variable N_{eff} and the neutron lifetime quoted by the PDG. In this manner we can estimate the bias introduced by assuming a fixed value of the neutron lifetime. The results are tabulated in 17.4.

In both the CMB-S4 and the futuristic CMB-S4+ experiment, changing the value of the neutron lifetime has the effect of shifting the reconstructed mean value of N_{eff}. In particular, using the extremal high value $\tau_n = 888.0s$ causes a shift, with respect to the base value, of $\Delta N_{eff} = 0.014$, while the lower value $\tau_n = 877.0s$ shifts $\Delta N_{eff} = -0.009$. In practice, neither of the values would constitute a false positive in the determination of the presence of the light relic. However, including the error on τ_n may result in a larger statistical fluctuation of the reconstructed N_{eff}. In practice, the uncertainty in the measured neutron lifetime is introducing an additional uncertainty in the determination of N_{eff} of ~ 0.02.

18 Cosmic Scalar Fields

18.1 The Strong CP Problem

Theoretical arguments show that in a gauge theory such as QCD the Lagrangian should contain the CP violating term[166, 77, 188]

$$\frac{\theta}{32\pi}F^a_{\mu\nu}\tilde{F}^{\mu\nu}_a \tag{18.1}$$

where $F^a_{\mu\nu}$ is the gauge field strength, such as the Gluon fields, and $\tilde{F}^a_{\mu\nu} = \frac{1}{2}\varepsilon_{\mu\nu\rho\sigma}F^{\rho\sigma}_a$ its dual. The parameter θ is to be determined by experiment and, in principle, may take any value in the interval $[0, 2\pi)$. Experimentally, θ is constrained by the electric dipole moment of the neutron and has a value $\theta \ll 10^{-9}$[11, 99]. The question is therefore, why is this parameter very close to zero? This is known as the *strong CP problem.* Let's illustrate how the problem arises in a gauge theory. In fact, we will investigate an intriguing part of the structure of gauge theory.

Let's work with a non-Abelian gauge theory with a gauge group $SU(N)$. The generators of the group are T^a which form the Lie algebra

$$[T^a, T^b] = if^{abc}T^c \tag{18.2}$$

f^{abc} are the group's structure constants and the sum is implied on repeated indexes. The gauge field is $A^\mu(x) = A^\mu_a(x)T^a$ and the field strength tensor is given by

$$F_{\mu\nu} = \partial_\nu A_\mu - \partial_\mu A_\nu + i[A_\nu, A_\mu] \tag{18.3}$$

The dual is given explicitly by

$$\tilde{F}_{\mu\nu} = \varepsilon_{\mu\nu\rho\sigma}(\partial_\sigma A_\rho - iA_\rho A_\sigma) \tag{18.4}$$

Under a gauge transformation with $U(x) \in SU(N)$ the field transforms as

$$A^{'}_\mu(x) = U(x)A_\mu U^{-1}(x) + i(\partial_\mu U(x))U^{-1}(x) \tag{18.5}$$

In this section we will simply use x to denote the four vector x^μ, to simplify the notation. The field strength transforms as

$$F^{'}_{\mu\nu} = UF_{\mu\nu}U^{-1} \tag{18.6}$$

A field ψ^a would transform as $\psi^{'} = U_\psi(x)\psi(x)$ where U_ψ is the element of the gauge transformation in the representation of ψ. The generators of its representation are T^a_ψ. The covariant derivative is defined[17]

$$D^\psi_\mu\psi(x) = (\partial_\mu + iA^a_\mu T^a_\psi)\psi(x) \tag{18.7}$$

[17] In Euclidean space it is $D_\mu = \partial^E_\mu + A^a_\mu T^a$, where ∂^E_μ is the derivative with respect to the Euclidean variables, which are the usual spatial ones and imaginary time.

The *Euclidean action is*

$$S_E = \frac{1}{4g^2}\int d^4x F^a_{\mu\nu}F^{\mu\nu}_a \tag{18.8}$$

g being the coupling constant. We work with the Euclidean action directly and want to find classical solutions to the field equation, which are simply configurations of the fields A^μ_a in space time which extremize the action. In the path integral formulation, the integral is over all classical paths, whether they extremize the action or not. The aspect we are interested in will be the extension to quantum field theory of the idea of ordinary mechanics that the transition probability between position eigenvectors $|x_i>$ to $|x_f>$ during a time T is given by[188]

$$< x_f|e^{-\frac{HT}{\hbar}}|x_i >= N\int D[x]e^{-S[x]/\hbar} \tag{18.9}$$

where $D[x]$ is the integration measure over all functions $x(t)$ such that $x(-T/2) = x_i$ and $x(T/2) = x_f$. This is done in Euclidean space. In ordinary quantum mechanics, the above integral gives the *tunneling amplitude* between x_i and x_f .

18.1.1 Winding number and Chern-Simons current

Let's consider possible configurations for the fields A^μ_a that do not give an infinite contributions to the action (18.8). Indeed, any configuration which gives a positive infinite value would contribute zero to the path integral due the exponential suppression. It cannot be $S = -\infty$, as by hypothesis the action is bounded from below. At first, the configurations we consider need not be a classical solution. Quantum mechanics allows us to sum over all possible paths, or configurations. We only want to get a feeling for the structure of these configurations.

In order for the integral to be convergent, F^2 in the integral must fall to zero at least as fast as r^{-4}, r being the Euclidean radius $r = \sqrt{\sum_i x_i^2}$. Apparantly, A^μ must fall at least as r^{-1} for the integral to be convergent. However, A^μ may also contain a pure gauge term which can be put to zero with a transformation (18.5). We know that the action is Gauge invariant and this term does not contribute to $F_{\mu\nu}$. So at very large values of $r \to \infty$

$$A^\mu = i(\partial_\mu U)U^{-1} + o(\frac{1}{r}) \tag{18.10}$$

where $U(x)$ is some element of the gauge group. It's easy to show that under a gauge transformation with a group element $V(x)$, through (18.5), $A^\mu \to A^{\mu'}$ parametrized in the same way with the changed matrix $W(x) = V(x)U(x)$.

At first, it would seem that every solution at $r \to \infty$ is gauge equivalent. This is, surprisingly, not true. Naively, one would think that, by the above equation, for any configuration given by $U(x)$, it is possible to choose $V(x) = U^{-1}$ and eliminate the pure gauge part at infinity. This is not possible, due to the fact that gauge transformations must be continuous and differentiable functions in $\mathbb{R}^4$. Although $U(x)$ can be set to zero *almost everywhere*, it cannot be set to zero *everywhere*.

Let's look at the global structure of the gauge transformation $V(x) = V(r, \vec{\theta})$ used to make the transformation. We separate out the angular parts $\vec{\theta}$ on the 3-sphere. Instead of thinking

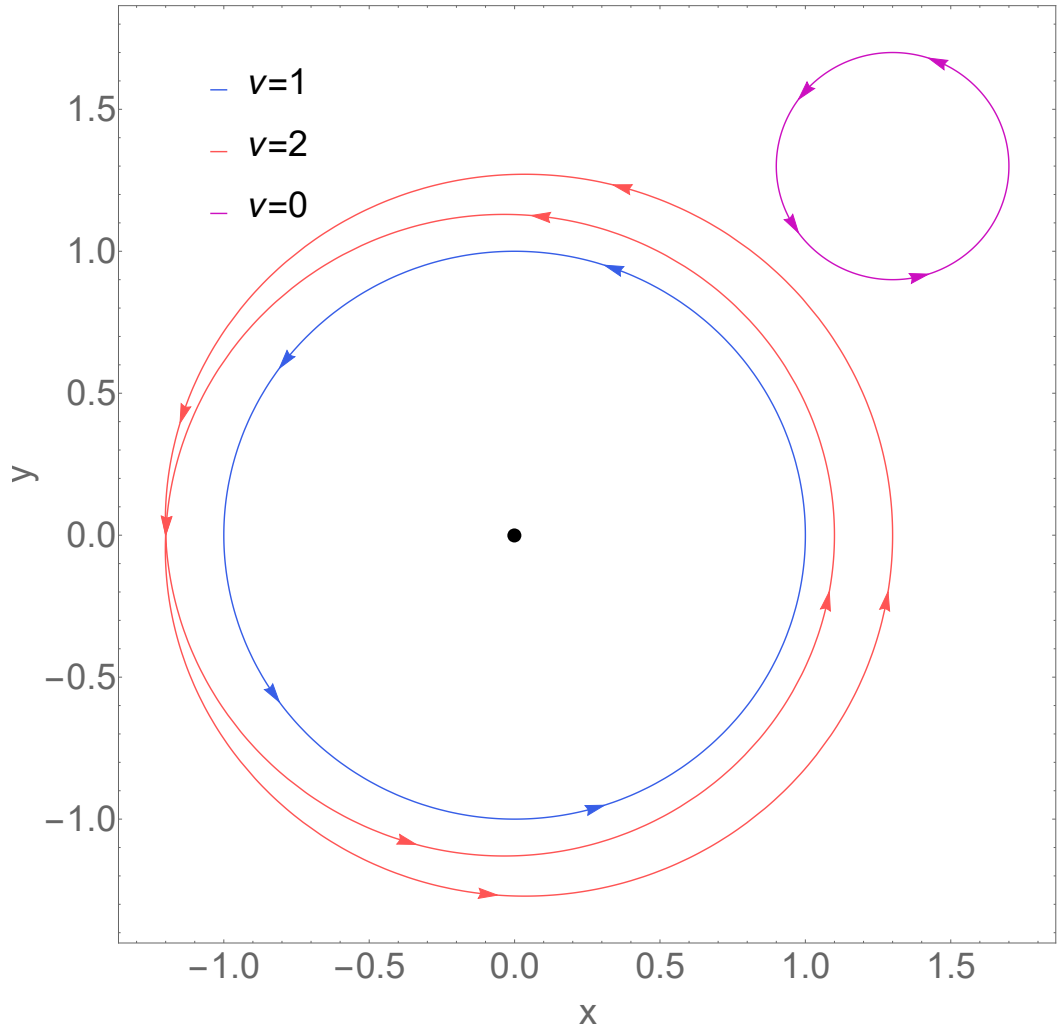

Figure 18.1: Example of closed curves, parametrized by $t \in [0, 2\pi)$, with different winding numbers on the plane $\mathbb{R}^2 - \{0\}$. The curves cannot be continuously deformed into one another without passing through the origin. The winding number around the origin is conserved through any continuous deformation and is a topological invariant.

of $V(x)$ as a function on $\mathbb{R}^4$, we think of it as as set of functions $S^3 \mapsto SU(N)$, from the 3-sphere, at fixed r, into the gauge group, parametrized by r. The functions $V(r, \vec{\theta}) \equiv V_r(\vec{\theta})$ are continuous and differentiable both in $\vec{\theta}$ and r. Now consider what happens as $r \to 0$. Clearly at $r = 0$ the function $V_{r=0}(\vec{\theta}) = V(0) = \text{const}$. This implies that whatever map $V_r(\vec{\theta})$ is at infinity, it must be *continuously deformable to the identity function.* On the other hand, $U(x)$ which appears in (18.10) need not be so, since as we go to $r \to 0$ the configuration A^μ can assume any form and cancel out the U. Therefore, the pure gauge term of A^μ at infinity *can only be eliminated if $U(x)$ is continuously deformable to the identity.* Indeed, if two maps are continuously deformable into the identity they are into one another. A set of maps that can be continuously deformed into one another is known as a *homotopy* class. It remains to be seen if there exists more than one homotopy class of gauge functions $U(\vec{\theta}) : S^3 \mapsto SU(N)$. Indeed there are.

To understand homotopy classes in our situation we note that $SU(2)$ has the topology of a 3-sphere; the maps under consideration are actually $S^3 \mapsto S^3$. $SU(N)$ with $N > 2$ has $SU(2)$ as a subgroup and would be a more general case. As an analogy we may consider maps on circles $S^1 \mapsto S^1$ or, equivalently, close curves on a plane to which we remove the origin $\mathbb{R}^2 - \{0\}$. This latter case has the same feature of maps between two circles if we consider the angular variable of the curve and take the parameter $t \in [0, 2\pi)$. Of course, the curve must be closed since $t = 0$ and $t = 2\pi$ are the same point on the circle. In that case, all the closed curves are characterized by a *topological winding number,* which is how many times the closed curve goes around the origin. As is well known, we cannot continuously deform two curves which have a different winding number into one another, without passing through the origin. Curves with specific winding numbers form a homotopy class. It can be shown that in the examples under consideration here, the winding number is the only parameter needed to specify the homotopy class.

Going back to the situation under consideration, let's take the specific example of a $SU(2)$ gauge group. Any element of $SU(2)$ in the 2 representation may be written as

$$U(x) = \alpha(x) I - i\vec{\beta}(x) \cdot \vec{\sigma} \tag{18.11}$$

where I is the identity and $\vec{\sigma} = \{\sigma_1, \sigma_2, \sigma_3\}$ are the Pauli matrices and α, β real functions satisfying $\alpha^2 + |\vec{\beta}|^2 = 1$. In particular let's consider gauge functions

$$U^{(0)}(x) = I \tag{18.12}$$

$$U^{(1)}(x) = \frac{x_4 - i\vec{x} \cdot \vec{\sigma}}{r} \tag{18.13}$$

$$U^{(\nu)} = (U^{(1)})^\nu \tag{18.14}$$

we will call ν the winding number or *Pontryagin index*[188]. It can easily be seen that $U^{(-1)} = \frac{x_4 + i\vec{x} \cdot \vec{\sigma}}{r}$. The Pontryagin index is a *topological quantity,* it cannot be removed through a continuous transformation. The homotopy classes are identified by a given Pontryagin index. We will not prove it, but it can be shown that this is enough to identify all regular homotopy classes. Any $U(x)$ can be transformed to some $U^{(\nu)}$ with a continuous

transformation, and no continuous transformation brings a $U^{(\nu)}$ into $U^{(\nu')}$ with $\nu \neq \nu'$. We will instead give an analytic expression of the Pontryagin index and show that it is indeed invariant under continuous deformations. We propose that

$$\nu = \frac{1}{24\pi^2} \int\limits_{S^3_\infty} d^3S \frac{x_\mu}{r} \varepsilon^{\mu\nu\rho\sigma} \text{Tr}\left(U^{-1}(\partial_\nu U)U^{-1}(\partial_\rho U)U^{-1}(\partial_\sigma U)\right) \tag{18.15}$$

gives the Pontryagin index of a field $U^{(\nu)}$ as defined above. The integral is extended over the 3-sphere at infinity (we are working in Euclidean space), so the definition is to be intended as the limit of the integral as $r \to \infty$. First, let's show that if we perform a continuous change of the map $U(x)$ the integrand does not change. We work at first order for a small continuous change

$$U \to U^{'} = U + \delta T U \tag{18.16}$$

with δT is another map from $\mathbb{R}^4 \mapsto SU(2)$. It follows that

$$(U^{'})^{-1} = U - U^{-1}\delta T \tag{18.17}$$

Note that δT and U, U^{-1} do not commute. For example, one of the terms that appears in the above trace is

$$U^{-1}\partial_\mu U \to U^{-1}\partial_\mu U + -U^{-1}\delta T \partial_\mu U + U^{-1}\partial_\mu(\delta T U) = U^{-1}\partial_\mu U + U^{-1}(\partial_\mu \delta T)U \tag{18.18}$$

where we have used the Leibnitz product rule in the last equality. Combining the elements into the trace in (18.15) the term at zero order is unchanged. The trace contains first order terms of the form

$$\text{Tr}\left(U^{-1}(\partial_\nu \delta T)UU^{-1}(\partial_\rho U)U^{-1}(\partial_\sigma U)\right) \tag{18.19}$$

and others with $\nu \to \rho \to \sigma \to \nu$. Using the identity $(\partial_\sigma U)U^{-1} = -U\partial_\sigma U^{-1}$, together with the ciclicity of the trace, this term is

$$-\text{Tr}\left((\partial_\nu \delta T)\partial_\rho U \partial_\sigma U^{-1}\right) \tag{18.20}$$

It appears under an integral, so we may integrate by parts in $x^\nu (= x_\nu$ in Euclidean space). There is no boundary term, since the integral is over a closed hypersurface. It's easy to see that every reamining term is symmetric in a pair of indexes. Since it is contracted with the Levi-Civita tensor, the contribution of these terms is zero. So ν is indeed invariant under a continuous transformation of the field $U(x)$.

Now we want to check that the definition (18.15) coincides with what we defined in (18.14). It is possible to take the derivatives explicitly and work through the well known Pauli algebra, this would be relatively straightforward. However, we can skip the tedious calculations by noticing that the trace term

$$\text{Tr}\left(U^{-1}(\partial_\nu U)U^{-1}(\partial_\rho U)U^{-1}(\partial_\sigma U)\right) \tag{18.21}$$

must be a tensor built out of the x^ν, the Kronecker Delta and the Levi-Civita tensor. Since

it will be contracted with the Levi-Civita tensor we must only take the completely anti-symmetric part:

$$\mathrm{Tr}\left(U^{-1}(\partial_\nu U)U^{-1}(\partial_\rho U)U^{-1}(\partial_\sigma U)\right) \to C(r)\varepsilon_{\nu\rho\sigma\lambda}x^\lambda \tag{18.22}$$

where the overall constant can only be function of the radius. To determine this function we can calculate the integrand at any point on a 3-sphere. Let's choose the simplest: $x^4 = r$, $x^{1,2,3} = 0$. Due to the Levi-Civita tensor then $\nu, \rho, \sigma = 1, 2, 3$ and at this point $U = 1$, $\partial_\nu = -\frac{i}{r}\sigma_\nu$. So the trace reduces to

$$(-i)^3\frac{1}{r^3}\mathrm{Tr}(\sigma_\nu\sigma_\rho\sigma_\sigma) = -\frac{2}{r^3}\varepsilon_{\nu\rho\sigma 4}(\frac{x^4}{r}) = -\frac{2}{r^4}\varepsilon_{\nu\rho\sigma\lambda}x^\lambda \tag{18.23}$$

having used the Pauli matrices trace relation $\mathrm{Tr}(\sigma_i\sigma_j\sigma_k) = 2i\varepsilon_{ijk} = 2i\varepsilon_{ijk4}$. So we conclude that $C(r) = -\frac{2}{r^4}$. So ν for $U^{(1)}$ is

$$\nu(U^{(1)}) = \frac{1}{24\pi^2}\int r^3 d\Omega_3 \frac{x^\mu}{r}(-\frac{2}{r^4})\varepsilon_{\mu\nu\rho\sigma}\varepsilon_{\nu\rho\sigma\lambda}x^\lambda \tag{18.24}$$

We use the relation of the Levi-Civita tensor

$$\varepsilon_{\mu\nu\rho\sigma}\varepsilon_{\nu\rho\sigma\lambda} = -\varepsilon_{\nu\rho\sigma\mu}\varepsilon_{\nu\rho\sigma\lambda} = -6\delta_{\mu,\lambda} \tag{18.25}$$

To obtain

$$\nu(U^{(1)}) = \frac{1}{2\pi^2}\int d\Omega_3 = 1 \tag{18.26}$$

As the surface of the 3-sphere is $2\pi^2 r^3$. Next, we should prove that the relation holds even for $U^{(\nu)}$. To do so, we will prove the more general fact that if two maps U, V have winding numbers ν_1, ν_2, the product UV has winding number $\nu_1 + \nu_2$. One could work out trace algebra, but there is a simpler deeper way to understand this. Through a continuous transformation, we may always deform the map U so that it is the identity on one hemisphere of the 3-sphere. Of course, it will not be the identity on the other hemisphere. This deformation is possible, it amounts to taking the naive deformation to eliminate the pure gauge field we were discussing at the start. There is no problem while we work on one hemisphere. All the "windiness" gets confined to the other hemisphere. If we take the Pontryagin integral now, the integrand on one hemisphere does not contribute, but the integrand over the other hemisphere still gives the complete winding number. Thus when considering the map UV we simply deform it (which keeps its winding number constant) so that it is equal to U on one hemisphere and to V on the other hemisphere. It is now obvious that the total Pontryagin index $\nu = \nu_1 + \nu_2$.

We are not done with the mathematics yet, but we can start connect all this winding number business with some physics. The CP violating term is peculiar (18.1). In QED, with the gauge group being $U(1)$, it is straightforward to show that this is a total derivative and therefore cannot contribute to any process. This is true even for a more complex gauge group $SU(N)$! Indeed, since θ is a constant, we compute the value of the trace explicitly,

using the antisymmetry of $\varepsilon_{\mu\nu\rho\sigma}$ and the ciclicity of the trace,

$$\text{Tr}(F_{\mu\nu}\tilde{F}^{\mu\nu}) = -2\varepsilon_{\mu\nu\rho\sigma}\text{Tr}\left(\partial_\mu A_\nu \partial_\sigma A_\rho + iA_\mu A_\nu \partial_\sigma A_\rho - iA_\rho A_\sigma \partial_\mu A_\nu + A_\mu A_\nu A_\rho A_\sigma\right) \quad (18.27)$$

The last term does not contribute since, by the ciclicity of the trace, it is equal to itself with the indexes permuted cyclically. This permutation is odd in four dimension and so the sum over all indexes vanishes. On the other hand, the indexes of the second term can be rearranged with the permutation $(\mu\nu\rho\sigma) \to (\rho\sigma\nu\mu)$ which is odd. So the second and third term are equal

$$\text{Tr}(F_{\mu\nu}\tilde{F}^{\mu\nu}) = -2\varepsilon_{\mu\nu\rho\sigma}\text{Tr}\left(\partial_\mu A_\nu \partial_\sigma A_\rho - 2i\partial_\mu A_\nu A_\rho A_\sigma\right) \quad (18.28)$$

That this is a four divergence is not so obvious. The solution is given in terms of the *Chern-Simons* current

$$G_\mu \equiv \varepsilon_{\mu\nu\rho\sigma}\text{Tr}(2iA_\nu F_{\rho\sigma} - \frac{4}{3}A_\nu A_\rho A_\sigma) \quad (18.29)$$

It is not too complicated to take the divergence and show indeed that

$$\text{Tr}(F_{\mu\nu}\tilde{F}^{\mu\nu}) = \frac{i}{2}\partial_\mu G^\mu \quad (18.30)$$

Now everything comes together in a quite spectacular way. The CP violating term is a total divergence and so by Gauss' theorem, only its boundary term is important. Usually boundary terms are glossed over in physics when calculating the action. We integrate by parts and drop the boundary term, assuming that the integral goes to zero sufficiently fast at infinity. However, G^μ is not negligible at the boundary, due to the fact that fields may be pure gauge, (18.10). Indeed if A^μ is pure gauge

$$\begin{aligned} G_\mu &= -\frac{4}{3}\varepsilon_{\mu\nu\rho\sigma}\text{Tr}(A_\nu A_\rho A_\sigma) \quad (18.31)\\ &= \frac{4i}{3}\varepsilon_{\mu\nu\rho\sigma}\text{Tr}\left(U^{-1}(\partial_\nu U)U^{-1}(\partial_\rho U)U^{-1}(\partial_\sigma U)\right) \end{aligned}$$

But this quantity is the integrand of the Pontryagin index (18.15)!

$$\nu = -\frac{i}{32\pi^2}\int d^3S\frac{x^\mu}{r}G^\mu \quad (18.32)$$

x^μ/r is no other than the normal vector to the 3-sphere. The Pontryagin index is the integral of the flux of G^μ across the surface at infinity. By Gauss' theorem, we can turn this into an integral over all the volume and so

$$\nu = -\frac{1}{16\pi^2}\int d^4x\text{Tr}(F_{\mu\nu}\tilde{F}^{\mu\nu}) \quad (18.33)$$

We have ran through many interesting concepts in a short time, so let's pause to reflect on what we have discovered. We have found that the field configurations A^μ, which we must integrate over in the path integral, can be separated into many homotopy classes based on the winding number ν they acquire in the limit $r \to \infty$. Next we considered the CP violating term, found that it is a total divergence, and with this fact shown that the

winding number of a field configuration A^μ can be calculated by the integral (18.33) on all space.

This is curious. The result would indicates that the CP violating term does not indeed contribute at any order in perturbation theory, because it is a total divergence. It represent the winding number, which is an inherently non-local quantity, depending on the global shape of the field configuration, rather than on any local detail per se. We may ask, if it does not contribute perturbatively, *is it physical?* The answer is that it affects *non-perturbative physics.* In QCD at low energies, we know that a perturbative approach is not possible. The coupling constant explodes and this means that the complete structure of the path integral must be considered. This structure inherently contains information about the winding number. This is the reason the CP violating term affects the electric dipole moment of the neutron. One last thing before we proceed. Although we have shown what the CP violating term is related to, we haven't given a reason why it should exist, what is θ and why there is a "Strong CP problem".

Before we proceed, let's point out that for electromagnetism, with $U(1)$, there is no winding number. Or, better, the winding number is always zero. This is because the gauge group $U(1)$ is too "simple" and a continuous transformation can always be found that deformes a map $S^3 \mapsto U(1)$ into the identity map. Specifically, for a pure $U(1)$ gauge field the Chern-Simons current (18.31) is zero, since the fields A^μ commute with one another.

18.1.2 Instantons and the θ vacua

Let's take a trip to memory lane, back to ordinary quantum mechanics of one particle in one dimension. In the Schrödinger picture we are given a set of time independent states $|x_i>$, eigenstates of the position operator, which form a complete basis of the Hilbert space. Then, any transition from an initial state $|\psi(-\frac{T}{2})>$ at the initial time $-\frac{T}{2}$ to the final state $|\psi(\frac{T}{2})>$ at $\frac{T}{2}$ can be described through a sum of the transition matrix elements[194]

$$< x_f|e^{-iHt}|x_i >= \int D[x]e^{iS} \tag{18.34}$$

where the integration is done upon all paths with $x(-\frac{T}{2}) = x_i$ and $x(\frac{T}{2}) = x_f$. When we pass from ordinary quantum mechanics to quantum field theory, we let the role played by the path $x(t)$ to be played by the value of a field $\phi(\vec{x}, t)$. Of course there are many points $\vec{x}$ which means that for each point in space there is a quantum mechanical degree of freedom. The evolution operator is unchanged in form, it is still the complex exponential of the Hamiltonian. The transition element is still in the same spirit as in (18.34): we get the S matrix. Of course, being in the same spirit, the path integral must evaluated on all paths which have definite initial and final conditions, in terms of the field. In the path integral

$$\underbrace{\int D[\phi]e^{iS[\phi]}}_{\begin{cases}\phi(\vec{x}, -\frac{T}{2}) = \phi_i(\vec{x}) \\ \phi(\vec{x}, +\frac{T}{2}) = \phi_f(\vec{x})\end{cases}} \tag{18.35}$$

the sum must be taken between *classical* initial and final configurations[189]. By classical,

we simply mean that a function $\phi_{i,f}(\vec{x})$ is given which is not an operator. Usually, this initial or final configuration is thought to be a collection of free fields. Now, with this in mind, let's take a look at vacuum to vacuum transition probabilities in gauge theories. We will Wick-rotate immediately, working with imaginary time. A vacuum transition amplitude would be some matrix element of the form

$$< \Omega|e^{-HT}|\Omega > \tag{18.36}$$

where $|\Omega >$ is the full interacting theory vacuum state and we are considering a very long transition time $T \to \infty$. Now, usually one doesn't consider these matrix elements in field theory. The reason is actually quite simple: $|\Omega >$ must be an eigenstate of the Hamiltonian H and the energy value, $H|\Omega >= E_0|\Omega >$, can be arbitrarily set to zero, $E_0 = 0$, without affecting any non-gravitational observable. Therefore we are left with

$$< \Omega|\Omega >= 1 \tag{18.37}$$

The reason we are talking about this, is because in an $SU(N)$ gauge theory it is not as simple. The conceptual problem begins when we want to write the vacuum transition as a path integral. For a transition between field configurations at $-\frac{T}{2}$ and $+\frac{T}{2}$, where we will eventually send $T \to \infty$, what are the correct initial and final configurations to use as boundary terms in the path integral? Naively, we would like to set these configurations to zero. This is very reasonable in a scalar field theory without any gauge invariance, for example. However we have seen how there exist inequivalent classes of pure gauge solutions (18.10) at infinity, characterized by their winding number. So let's see what happens if we suppose the vacuum is defined by a fixed winding number ν for the gauge field at $x^4 \to \pm\infty$. We denote such a state

$$|\nu > \tag{18.38}$$

In the Schrödinger picture it is time independent. Consider the transition amplitude between two vacua with differing winding numbers

$$< \nu_1|e^{-HT}|\nu_2 >= N \int D[A]e^{-S} \tag{18.39}$$

where $D[A]$ is the integration over all possible fields A^μ and N is some unimportant normalization constant. The path integral is done over all configurations $A^\mu(\vec{x}, t)$ which reduce to a pure gauge configuration with winding number ν_2 at $x^4 = -\infty$ and $-\nu_1$ at $x^4 = +\infty$[18]. This does *not imply that the winding number of the path A^μ is one or the other*. Rather, A^μ must interpolate between the two and will have a different winding number ν_A. To find the winding number we must evaluate the Pontryagin integral (18.15). As we discussed following equation (18.26), we may deform the initial and final configurations so that they reduce to the identity in one hemisphere or the other. If we make this choice, the winding number of the initial and final configurations is obtained simply by the integral on the boundary at $-\frac{T}{2}$ and $\frac{T}{2}$ respectively, while the integral at the opposing side is zero. Since the field configuration (which we integrate over) must reduce to these on *both hemispheres*

[18]The minus sign arises because if $|\nu_1 >$ is related to a gauge configuration $U^{(\nu_1)}$ then the Hermitian conjugate $< \nu_1|$ must be related to the adjoint of $U^{(\nu_1)\dagger} = U^{(-\nu_1)}$

then its winding number is

$$\nu_A = \nu_2 - \nu_1 \tag{18.40}$$

The configurations which interpolate between two $|\nu>$ states with different winding numbers are known as *instantons*[188, 21]. This means we can write explicitly in the path integral (18.39) *without loss of generality*

$$< \nu_1|e^{-HT}|\nu_2 >= N \int D[A]e^{-S}\delta_{\nu_A,\nu_2-\nu_1} \tag{18.41}$$

A priori, such interpolating paths do not necessarily exist. In that case, this is zero for $\nu_1 \neq \nu_2$. Any possible instanton cannot be pure gauge everywhere, as a continuous gauge transformation cannot change winding number. A known $SU(2)$ instanton solution is the BPST instanton. This changes the winding number by 1 between an initial and final configuration. An *anti-instanton* changes the winding number by -1. A characteristic of this solution is that the "jump" between winding numbers happens suddenly and quickly, hence the name instanton. This also means that one can construct interpolating solutions which change winding number by more than one by consider a sequence of instantons widely separated in time. This is fine since $T \to \infty$ while $\Delta t_{\text{instanton}}$, the "jump" time, remains finite.

The most important thing about the instanton *is that it is a solution of the classical equations of motion with imaginary time.* In other words, the instanton solution is a minimum for the action $S[A^{\mu}_{\text{instanton}}] = S_0$. We know from regular quantum mechanics that classical solutions of the equations of motion with imaginary time are related to tunneling amplitude. Thus, what we have found is that (18.39) represents a tunneling amplitude between states of different winding numbers. This means that the $|\nu>$ *cannot be the vacuum states.*

We had expected the vacuum state to be an eigenstate of the Hamiltonian. So now the question is, what eigenstate can be built out the $|\nu>$ states? It is likely that some linear combination could be the correct vacuum state. Indeed, *these are the θ vacua*

$$|\theta> \sum_{\nu=-\infty}^{+\infty} e^{i\nu\theta}|\nu> \tag{18.42}$$

This is a set of continuous states which are linear combination of the states with definite winding number. The transition between two different values of $|\theta>$ is given by

$$< \theta^{'}|e^{-HT}|\theta >= \sum_{\nu,\nu^{'}} < \nu^{'}|e^{-HT}|\nu > e^{i(\nu\theta-\nu^{'}\theta^{'})} \tag{18.43}$$

The sum over $\nu^{'}$ can be recast into a sum over $n = \nu^{'} - \nu$ so that

$$< \theta^{'}|e^{-HT}|\theta >= \sum_{\nu} e^{i\nu(\theta-\theta^{'})} \sum_{n} < \nu+n|E^{-HT}|\nu > e^{-in\theta^{'}} \tag{18.44}$$

The transition matrix $< \nu+n|e^{-HT}|\nu >$ can only depend on the change in winding number and the transition time as (18.41). The sum over n is some constant with depends only on

T and $\theta^{'}$. Therefore we may write

$$< \theta^{'}|e^{-HT}|\theta >= C(T,\theta^{'})\sum_{\nu} e^{i\nu(\theta-\theta^{'})} \tag{18.45}$$

for some function C of the transition time and the vacua. The sum over ν is equal to $2\pi\delta(\theta-\theta^{'})$ and so we find

$$< \theta^{'}|e^{-HT}|\theta >= 2\pi C(T,\theta^{'})\delta(\theta-\theta^{'}) \tag{18.46}$$

This is what we expect of eigenstates of the Hamiltonian, *no transition out of it*, and we conclude that *a $|\theta>$ state must be the vacuum of the theory*. Being in the θ vacuum affects the action. Indeed, any expectation value of the theory is written through

$$< \theta|\hat{T}|\theta >= N\int\limits_{|\theta>} D[A]e^{-S}T(A) \tag{18.47}$$

and must be calculated with the correct boundary conditions. We indicate this by the subscript $|\theta>$. N is some normalization constant. Since the classical configuration of a theta vacuum corresponds to a sum of the classical configurations of fixed winding numbers then

$$< \theta|\hat{T}|\theta^{'} >= N\sum_{\nu} e^{i\nu(\theta-\theta^{'})}\sum_{n} e^{-in\theta^{'}}\int d[A]e^{-S}T(A)\delta_{\nu_A,n} \tag{18.48}$$

where the last path integral is evaluated with the boundary conditions of the $|\nu>$ vacuums, with any initial and final winding numbers (the Kronecker delta will see to select the right transition). The sum over ν gives the $\delta(\theta-\theta^{'})$ as before, while the term $e^{-in\theta}$ can be brought in the integral and the sum evaluated using the Kronecker-delta.

$$< \theta|\hat{T}|\theta^{'} >= N\delta(\theta-\theta^{'})\int D[A]e^{-S}e^{-i\nu_A\theta}T(A) \tag{18.49}$$

When $\theta=\theta^{'}$ the prefactor $N\times\infty$ is what we would expect from the usual field theory, its interpretation is the usual one. On the other hand, we find that the theta vacuum is registered in the path integral by the adding of a term to the action, so

$$S \to S + i\theta\nu \tag{18.50}$$

So by the formula (18.33) for the Pontryagin index in terms of the field strength and its dual

$$\mathcal{L}_M \to \mathcal{L}_M + \frac{\theta}{16\pi^2}\mathrm{Tr}(F_{\mu\nu}\tilde{F}^{\mu\nu}) \tag{18.51}$$

for the Lagrangian in Minkoswski space (real time). We may account for being in a theta vacuum by adding the CP violating term to the Lagrangian. The theory cannot forego this term and the expectation value of observables depends on the θ. In particular, one can calculate the expectation value for $< \theta|\mathrm{Tr}(F_{\mu\nu}(x)\tilde{F}^{\mu\nu}(x)|\theta >$ and find it depends explicitly on the value of θ. This expectation value is important in the non-perturbative regime, which for QCD is all the low energy physics. θ is physical.

Now the CP problem becomes apparent. Why is the θ parameter so small? Why is it comparable to zero? Indeed, θ is simply some parameter in the interval $[0, 2\pi)$, we cannot expect it to be zero a-priori. The fact that it is zero seems to suggest there is some symmetry at play, but which is symmetry is a Nobel-prize problem.

States with different values of $|\theta>$ are characterized by different vacuum energies. The transition element $< \theta|e^{-HT}|\theta^{'} > \propto e^{-E(\theta)T}\delta(\theta - \theta^{'})$, so we can calculate the energy of these states. The calculation involves expanding the solution about the instanton solution, which are minima of the action with $S_0 = \frac{8\pi^2}{g^2}$. There are some subtleties and the full calculation is out of the scope of this text. The energy density of a theta vacuum is

$$\frac{E(\theta)}{V} = -2K\cos\theta e^{-\frac{8\pi^2}{g^2}} \tag{18.52}$$

where K is known as the determinant prefactor and can be estimated numerically. It turns out that $K = 0$ when there is a massless fermion present which couples to the gauge field. We will come back to this, which is known as the fermion zero mode. There are two things to note. First, the exponential factor contains a factor $-g^{-2}$ and so one can see that $e^{-\frac{8\pi^2}{g^2}}/g^n \to 0$ as $g \to 0$ for any integer n. This is a purely non-perturbative effect, it would not appear at any order in perturbation theory (small g). The second is that the energy depends with the cosine of θ, so every $|\theta>$ has a different value of energy. This does not mean that θ can change. We have obsessed about that. It is not a dynamical variable and we are still allowed to add an arbitrary constant to make the vacuum energy zero.

There is a deep connection between the θ term and chiral transformations of fermionic fields and, in turn, with quantum anomalies. In the presence of a massless fermion field ψ, one may perform a rotation of the left and right chiral components $\psi_{L,R} = \frac{1}{2}(1 \mp \gamma^5)\psi$ as

$$\psi_L \to e^{i(\alpha-\beta)}\psi_L \tag{18.53}$$

$$\psi_R \to e^{i(\alpha+\beta)}\psi_R \tag{18.54}$$

where α is vector rotation angle and β the axial rotation angle. Compactly, we may also write $\psi \to e^{i\alpha}\psi$ and $\psi \to e^{i\beta\gamma^5}\psi$ for the two possible transformations. Usually, if these are symmetries of the Lagrangian, they are denoted as $U_V(1)$ and $U_A(1)$. If the transformations are local, ie $\alpha, \beta \to \alpha(x), \beta(x)$, these are promoted to the gauge group of the theory. Quantum mechanically, this fails for the chiral symmetry $U(1)_A$. The issue is deep but it depends on the fact that loop diagrams, which are purely quantum mechanical as they vanish for $\hbar \to 0$, contain divergences which must be regularized. Any regularization scheme must respect the gauge symmetries of theory, which in the standard model is $U(1) \otimes SU(2) \otimes SU(3)$. It turns out that when a choice of regulator is made to respect these symmetries, it cannot respect the chiral symmetry at the same time. In terms of the path integral, a chiral symmetry cannot be a symmetry of the full quantum mechanical theory because the *measure of integration of the path integral is not invariant*. Indeed, when fermions are present the path integral integrates over all fermionic fields with a measure

$$D[\psi]D[\bar{\psi}] \tag{18.55}$$

where $\bar{\psi}$ is the conjugate field. It can be shown that, under a chiral transformation with constant angle β, the measure changes so that the Lagrangian acquires a term[189]

$$\delta\mathcal{L} = \frac{\beta}{16\pi^2}\text{Tr}(F_{\mu\nu}\tilde{F}^{\mu\nu}) \tag{18.56}$$

Which receives no corrections at any order in perturbation theory[14]. A chiral transformation is not a symmetry of the quantum theory *even if the fermions are massless*. Because of this change in the measure the proof for the Ward-Takahashi identities, which enforce the symmetry at all perturbation orders, fails. One can obtain a Schwinger-Dyson equation for the axial current $j_5^\mu = \bar{\psi}\gamma^\mu\gamma^5\psi$

$$\partial_\mu < j_5^\mu(x)O(x_1,\dots x_n) >= -\frac{1}{16\pi^2} < \text{Tr}(F_{\mu\nu}(x)\tilde{F}^{\mu\nu}(x))O(x_1,\dots x_n) > \tag{18.57}$$

for a generic operator $O(x_1,\dots x_n)$[189]. This is condensed into the anomaly equation

$$\partial_\mu j_5^\mu = -\frac{1}{16\pi^2}\text{Tr}(F_{\mu\nu}\tilde{F}^{\mu\nu}) \tag{18.58}$$

If we integrate over all space then we find that axial charges are not conserved throughout a transition and this non conservation is equal to the winding number change in the gauge field. Indeed this is a physical effect. Now we should note that, as the Lagrangian changes by (18.56), the θ parameter can be changed by a chiral rotation $\theta \to \theta + \beta$. If there are massless fermions coupled with the gauge field, a chiral rotation would leave the Lagrangian invariant and we are free to choose the chiral phase to eliminate the θ parameter. Thus, a solution to the strong CP problem would be having a massless fermion coupling to the gluons. In the case of QCD, with a symmetry $SU(3)$, the least massive quarks are the up u and down d. However they are not massless. We can still take a chiral rotation to eliminate θ but, since $U_A(1)$ is not a symmetry of the Lagrangian, we will be hiding the θ in the mass matrices (Yukawa couplings with the Higgs boson) of the quarks. This suggests that the physical value of the theta vacuum is combination of the actual θ and a complex phase present in the mass matrices[99]. In any case, quarks are not massless, and we may not remove the strong CP problem in this manner. We will see more ideas on how to solve the CP problem in the next section.

Let's conclude this very intense, but interesting, section with a few remarks. The instantons, which represent quantum tunneling between $|\nu >$ vacua, play a strong role in QCD and the strong CP problem is typically spoken about in this context. But the standard model also has an electroweak gauge group $SU(2)$ and through the same arguments, there should be a θ_{EW} term with the electroweak gauge fields as well. The electroweak sector has the same vacuum structure as the QCD sector. We live in a $|\theta_{EW} > |\theta_{QCD} >$ vacuum. In the electroweak case, unlike QCD, θ_{EW} can be set to zero[189]. This is due to the fact that right handed fields do not couple to the gauge fields. By a chiral rotation one eliminates the $\theta_{EW}F\tilde{F}$ term and puts all the dependence on θ_{EW} in the Yukawa couplings. Then we are free to perform a rotation on the right-handed fields to eliminate the phases, and this does not give other anomalous terms, since the right-chirality fields don't couple electroweakly. $\theta_{EW} = 0$.

This does not mean that there is not an issue about winding numbers in the electroweak

sector, which brings us to our next point. In the winding number discussion, we have remarked that a classical solution in imaginary time (as the instanton) represents a tunneling under a barrier. In the electroweak sector, since there is no infrared non-perturbative regime, the tunneling rate turns out to be extremely small and negligible in any known scenario. However, classical solutions in real time have been found. These are the *sphalerons*[19]. A sphaleron is an extrema of the action, in this sense it is a solution of the classical equations of motion, which interpolates vacuum states with different winding numbers in real time t[139, 5]. The solution can be given explicitly in terms of the Higgs field and the electroweak gauge bosons, or at least it can be found numerically. In a normal scattering experiment, the cross section to create a sphaleron and get into the next vacua is, again, extremely small[88]. However, in the early universe, when the temperature is at about the electroweak scale $T \sim 200GeV$, sphalerons are actually created in thermal equilibrium[83, 111]. The process of changing winding numbers is efficient. Indeed the barrier between the vacua is estimated at $T \sim 1TeV$, so thermal energies are about the right scale to climb over. This is extremely interesting because winding number change is related to the *non-conservation of lepton and baryon numbers*. In the standard model baryon numbers and lepton numbers are represented by a $U_{B,L}(1)$ global symmetry which is an accidental symmetry: it is not put in by hand and arises from the particular matter content and quantum numbers of the standard model. It turns[189] out the baryon and lepton current are anomalous

$$\partial_\mu j^\mu_{B,L} = -\frac{N_{B,L}}{16\pi^2}\mathrm{Tr}(F_{\mu\nu}\tilde{F}^{\mu\nu}) \tag{18.59}$$

Where $N_{B,L} = 2 \cdot N_g$ is the number of baryons/leptons in the standard model. If we integrate this on all space time, we get a change of baryon and lepton number before and after the sphaleron (which changes the winding number by one). The change in $B + L$ is

$$(\Delta_{B+L})_{1\text{ sphaleron}} = 4N_g = 12 \tag{18.60}$$

The combination $B + L$ is used since $B - L$ is not anomalous. This would seem to solve the problem of matter-antimatter asymmetry. It does not on its own due to the fact that the process does not prefer increasing winding number to decreasing it, nor does it prefer matter and antimatter. Therefore under thermal equilibrium conditions it gives no net asymmetry. But it is interesting, it shows that there are definite processes *in the standard model* which violate baryon and lepton number, one condition for baryogenesis. More importantly, it allows the possibility of an initial lepton asymmetry to be turned into a baryon asymmetry, which open the door to baryogenesis through leptogenesis[12, 72].

18.2 The Axion

The most common solution to the strong CP problem involves introducing the *axion*[165, 164, 217, 213]. This is a pseudoscalar field $\alpha(x)$ which has a coupling to the $SU(3)_{QCD}$ gauge fields as

$$\mathcal{L}_{int} \ni \frac{\alpha(x)}{16\pi^2 M}\mathrm{Tr}(G_{\mu\nu}\tilde{G}^{\mu\nu}) \tag{18.61}$$

[19]The word sphaleron comes from the greek $\sigma\varphi\alpha\lambda\varepsilon\rho o\varsigma$, meaning slippery. This is a reference to the nature of a sphaleron which is inherently unstable. It must climb to the top of a potential between states with different winding numbers and then becomes unstable and falls down the potential.

where we use $G^{\mu\nu}$ to describe the gluon fields, since here we will use $F_{\mu\nu}$ to describe the photon. M is some high energy scale. With a coupling of this form the total lagriangian, including the CP violating term (18.1) contains

$$\mathcal{L}_{int} \ni \frac{\frac{\alpha}{M}+\theta}{16\pi^2}\text{Tr}(G_{\mu\nu}\tilde{G}^{\mu\nu}) \tag{18.62}$$

In this manner, the shifted field $a(x) = M\theta + \alpha(x)$ has a potential given by (18.52). The vacuum expectation value is then the minimum of the potential

$$< a >= 0 \tag{18.63}$$

which is obtained for the classical value $a(x) = 0$, or $\alpha(x) = -\theta M$. With the introduction of this field, we have essentially promoted the variable θ to be dynamical. It relaxes to the minimum of a potential, as is due for a dynamical field and solves the strong CP problem, eliminating the spurious term in the Lagrangian.

In order to have a coupling of this form, the axion must be coupled to an anomalous current, as the Lagrangian is of the form $\alpha(x)\partial_\mu j^\mu$. The simplest way to make this happen is to introduce a global $U(1)$ symmetry, of which standard model particles, and possibly new particles to be introduced, carry a quantum number. This $U(1)$ is known as $U(1)_{PQ}$ where PQ stand for R. Peccei and H. Quinn, who first proposed it[165, 164]. The symmetry must be chiral: left and right fields must carry different quantum numbers. In addition, at least some of the fields that carry PQ charges must carry color. They must be in some representation of $SU(3)$. Different models can be built using this idea, the most common being some variation of the KSVZ[138, 196] or DFSZ[201, 78] models. We will illustrate the main concept behind them with a simple toy model.

Let's study a theory of one fermion field ψ which carries a charge q and transforms in the 3 representation of $SU(3)$. In addition, there is a complex scalar field ϕ which is uncharged and a color singlet. We shall say that the right component ψ_R carries a quantum number X_R under $U(1)_{PQ}$ and ψ_L carries X_L. To allow interaction terms between ϕ and ψ, ϕ must also carry a quantum number $X_\phi = X_R - X_L$. By this, we mean that under a $U(1)_{PQ}$ transformation the fields transform as

$$\psi_{R,L} \to e^{iX_{R,L}\eta}\psi_{R,L} \tag{18.64}$$

$$\phi \to e^{iX_\phi\eta}\phi \tag{18.65}$$

for some real η. The Lagrangian of interest is

$$\mathcal{L} = \mathcal{L}_{\text{gauge}} + \mathcal{L}_{\text{Dirac}} - \frac{1}{2}\partial_\mu\phi^\dagger\partial^\mu\phi - \frac{m^2}{2}\phi^2 + \lambda\phi\bar{\psi}_R\psi_L + h.c. \tag{18.66}$$

We note that the field ψ, carrying a PQ charge may be a regular quark, while the field ϕ is a new field which is typically referred to as a Higgs field, as we will see shortly. A $U(1)_{PQ}$ transformation actually contains a chiral rotation by the angle

$$\beta = \frac{1}{2}(X_R - X_L)\eta \tag{18.67}$$

which means that the symmetry is broken at the quantum level as the Lagrangian will gain an effective term (18.56) from the measure of the path integral, so long as $X_R \neq X_L$. This means the PQ current

$$j_\mu^{PQ} = -iX_\phi \phi^\dagger \overleftrightarrow{\delta^\mu} \phi + X_R \bar{\psi}_R \gamma^\mu \psi_R + X_L \bar{\psi}_L \gamma^\mu \psi_L \tag{18.68}$$

is anomalous. Using (18.58)

$$\partial_\mu j_{PQ}^\mu = -\frac{N}{16\pi^2} \mathrm{Tr}(G^{\mu\nu}\tilde{G}_{\mu\nu}) - \frac{E}{16\pi^2} F_{\mu\nu}\tilde{F}^{\mu\nu} \tag{18.69}$$

Where N and E are general coefficients that depend on the model. In our toy model $N = \frac{1}{2}(X_R - X_L)$ and $E = q^2(X_R - X_L)$. Next, we spontaneously break $U(1)_{PQ}$ by allowing the field ϕ to take a vacuum expectation value (v.e.v.)

$$\phi = \frac{1}{\sqrt{2}}(v + \rho(x)) \exp \frac{i\alpha(x)}{v} \tag{18.70}$$

where v is the v.e.v. , ρ is a heavy real scalar field and $\alpha(x)$ is the Goldstone boson, which is the axion. From a scheme like this one can derive all the couplings of the axion to the regular fields. We just want to show that the anomalous coupling exists with the gluons, which solves the θ vacuum problem, and with the photons. This latter coupling is important experimentally. We note that

$$j_\mu^{PQ} = -\frac{(v+\rho)^2}{v}\partial_\mu \alpha + \dots \tag{18.71}$$

the dots contain all the other fields, of course and there is no term containing either α, nor the gauge fields. The take away is that the derivative of the Goldstone mode can be written in terms of the current. In the Lagrangian, from the kinetic term we obtain

$$\frac{(v+\rho)^2}{2v^2}\partial_\mu \alpha \partial^\mu \alpha \tag{18.72}$$

among other terms. This is the only term in the Lagrangian where the Goldstone boson can appear. It appears through a derivative. In fact the $U(1)_{PQ}$ symmetry is now implemented in the field ϕ through a shift in α. Wwe integrate by parts and write $\partial_\mu \partial^\mu \alpha$ in terms of $\partial_\mu j_{PQ}^\mu$. Among other terms one obtains

$$\frac{\alpha}{v}\partial_\mu j_{PQ}^\mu \tag{18.73}$$

Using the anomaly equation (18.58), which is valid at every order in perturbation theory, we obtain couplings of the form (18.61), to the gluons and photons. Thus, as promised, the PQ symmetry solves the strong CP problem. No detail has been confirmed experimentally so far, and the detection of the axion remains a topic of active research. The most promising avenue for detection is through its coupling to the photon

$$\mathcal{L}_{a\gamma\gamma} = \frac{G}{4} a F_{\mu\nu}\tilde{F}^{\mu\nu} \tag{18.74}$$

where $G \sim v^{-1}$ is some inverse mass scale typical of the PQ symmetry breaking. We have rewritten the coupling in this form as it is the most commonly used when comparing experiments. We should point out that this coupling appears even if the charge of the fermion which carries PQ is zero. Indeed, we have failed to discuss how the axion mixes with the axial QCD current, which causes the pion and axion to mix.

Furthermore, the anomaly breaks the PQ symmetry explicitly, so that the axion cannot be a perfect Goldstone boson, but instead acquires a small mass[201]. Indeed, this small mass is another typical feature of the axion. It is a Pseudo Nambu-Goldstone boson(PNGB). Due to the symmetry, we may expect its mass to be very small $m_a \sim f_\pi m_\pi / f_a$.

Axions arise in a large variety of contexts. We generally refer to a light pseudo-scalar particle as an Axion Like Particle (ALP). Axions that solve the strong CP problem may arise from string theory. Indeed, many string theory models predict more than one ALP.

18.3 Zero order dynamics of a cosmic scalar field

We shall now study the evolution of a cosmic scalar field in a curved spacetime. The starting point is the action for the field[171]

$$S_\chi = \int d^4x\sqrt{-g}\left[-\frac{1}{2}g^{\mu\nu}\partial_\mu\chi\partial_\nu\chi - V(\chi)\right] \tag{18.75}$$

Where the $\sqrt{-g}$ makes the four integral invariant under coordinate changes. The sign of the kinetic term depends on the metric signature $(-,+,+,+)$ we are using, and it makes the time derivatives positive. The prescription to pass from a flat spacetime to a curved space time would be to take $\partial_\mu \to \nabla_\mu$, the covariant derivative, however on the scalar field $\nabla_\mu\chi = \partial_\mu\chi$. We use the greek letter χ for the scalar field, since we have used the usual ϕ with other meanings in this text.

We derive the equations of motion. The variation of the action is given by

$$\partial S = \int d^4x\left[-\sqrt{-g} - g^{\mu\nu}\partial_\mu\chi\partial_\nu\delta\chi - \sqrt{-g}V(\chi)\delta\chi\right] \tag{18.76}$$

We integrate by parts. The first term picks up the derivative $\partial_\nu(\sqrt{-g}g^{\mu\nu})$. Using the equalities $\partial_\mu g = g g^{\alpha\beta}\partial_\mu g_{\alpha\beta}$[20] and $g_\mu g^{\mu\nu} = -g^{\alpha\nu}g^{\beta\mu}\partial_\mu g_{\beta\alpha}$[21] and some indexology we get

$$\partial_\nu(\sqrt{-g}g^{\mu\nu}) = -\sqrt{-g}g^{\alpha\beta}\Gamma^\mu_{\alpha\beta} \tag{18.77}$$

With the fundamental theorem of variational calculus, ∂S is zero for any $\delta\chi$ when

$$g^{\mu\nu}\partial_\mu\partial_\nu\chi - g^{\mu\nu}\Gamma^\alpha_{\mu\nu}\partial_\alpha\chi - \frac{dV}{d\chi} = 0 \tag{18.78}$$

Using the definition of a covariant derivative on a one-form and on the scalar

$$g^{\mu\nu}\nabla_\mu\nabla_\nu\chi - \frac{dV}{d\chi} = 0 \tag{18.79}$$

[20] This is given by the Jacobi formula for the derivative of the determinant. $g^{\alpha\beta}$ is the matrix inverse of $g_{\mu\nu}$.
[21] $\partial_\mu(g^{\alpha\beta}g_{\beta\gamma}) = 0$ implies this.

This is obviously the generalization of a Klein-Gordon equation to curved space times.

We can derive the energy-momentum tensor through the definition (7.13). This is precisely the term that couples to gravity via the Einstein equations.

$$T_{\mu\nu} = \partial_\mu\chi\partial_\nu\chi - g_{\mu\nu}\left(\frac{1}{2}g^{\alpha\beta}\partial_\alpha\chi\partial_\beta\chi + V(\chi)\right) \tag{18.80}$$

These equations are exact. We will now look at the scalar field equation at order zero in a homogeneous flat FRW universe. We will work in conformal time (6.24). At the lowest order, we assume the field is uniform and depends only on conformal time.

$$\ddot{\chi} + 2\mathcal{H}\dot{\chi} + a^2\frac{dV}{d\chi} = 0 \tag{18.81}$$

We have encountered equations of this form before. At early times, when the Hubble factor is large, the field is *frozen* around its initial value due to Hubble friction. Eventually, as the universe expands and $\mathcal{H}$ decreases, the field *thaws* and begins to oscillate around the minimum of the potential.

The density $\rho = -T^0_0$ and pressure $P = \frac{1}{3}T^i_i$ are given by

$$\rho = \frac{\dot{\chi}^2}{2a^2} + \frac{|\vec{\nabla}\chi|^2}{2a^2} + V(\chi) \tag{18.82}$$

$$P = \frac{\dot{\chi}^2}{2a^2} + \frac{|\vec{\nabla}\chi|^2}{2a^2} - V(\chi) \tag{18.83}$$

$$w = \frac{P}{\rho} = \frac{\dot{\chi}^2 + |\vec{\nabla}\chi|^2 - 2a^2V(\chi)}{\dot{\chi}^2 + |\vec{\nabla}\chi|^2 + 2a^2V(\chi)} \tag{18.84}$$

where we kept the spatial derivative term for completeness, but it is zero in our approximation. Now we find something very curious. When χ is frozen, and $\dot{\chi} \simeq 0$, the dominating term is the potential. As long as we are not at the minimum then

$$\rho \simeq -P \tag{18.85}$$

This is precisely the equation of state we expect from dark energy, $w = 1$. This observation opens up the possibility that dark energy may be a cosmic scalar field frozen by Hubble friction. Unlike a vacuum energy, which is a hypothetical explanation for dark energy, this field is dynamical and perhaps there may be some experimental consequence of this dynamic in the evolution of the universe. In addition, the field may have interactions with standard model particles. We will come back to this.

When the Hubble friction becomes small, the field begins oscillating quite rapidly. If the period of oscillation is much shorter than $\mathcal{H}^{-1}$, it makes sense to average over the fast oscillations, since the peculiar oscillations cannot affect the evolution of the universe on longer time scales. To do this, it is more convenient to pass to coordinate time t. Recalling

that $dt = ad\tau$ the equations of motion can be written as

$$\chi^{''} + 3H\chi^{'} + \frac{dV}{d\chi} = 0 \tag{18.86}$$

Where the primes denote derivatives with respect to time.

The density and pressure are given by

$$\rho = \frac{\chi^{'2}}{2} + V(\chi) \tag{18.87}$$

$$P = \frac{\chi^{'2}}{2} - V(\chi) \tag{18.88}$$

In particular, the time derivative of the density is

$$\rho^{'} = -3H\chi^{'2} \tag{18.89}$$

We now assume the potential is of the form

$$V(\chi) = \Lambda^4(\frac{\chi}{\mu})^{2n} \tag{18.90}$$

where Λ and μ are some parameters with dimensions of energy. This potential is quite generic, so we expect to approximately describe a large class of models. When the field is oscillating it must be doing so around a minimum of the potential, which therefore must be close to this form. Over one period of oscillation, we may neglect the damping due to the Hubble term

$$\chi^{''} + 2n\frac{\Lambda^4}{\mu}(\frac{\chi}{\mu})^{2n-1} = 0 \tag{18.91}$$

for $n = 1$ this is a regular harmonic oscillator, and the period of oscillation is independent of the amplitude. For $n > 1$ it is an anaharmonic oscillator and the period depends on the amplitude of oscillation. We know that, at least during one period of oscillation, the density ρ is an integral of the motion, so

$$\frac{\chi^{'2}}{2} + V(\chi) = V(\chi_M) \tag{18.92}$$

where χ_M is the maximum amplitude over one period, $\chi_M^{'} = 0$. If we suppose the period of oscillation is much faster than a Hubble time H^{-1}, the problem can be separated into distinct parts. The fast oscillations neglect the expansion of the universe and are given by (18.91). On the other hand we may follow the decrease in density through (18.89) by taking the average over several periods of the oscillations

$$\frac{d}{dt} < \rho >= -3H < \chi^{'2} > \tag{18.93}$$

It is clear that the average density over one period of oscillation is given by the potential

calculated at the maximum amplitude

$$< \rho >= V(\chi_M) \tag{18.94}$$

To obtain the average of $< \chi^{'2} >$ we resort to the virial theorem. For a potential $V(\chi) \propto \chi^{2n}$, the classical virial theorem states

$$< \frac{\chi^{'2}}{2} >= n < V(\chi) > \tag{18.95}$$

Which implies that

$$< \frac{\chi^{'2}}{2} > (1 + \frac{1}{n}) = V(\chi_M) =< \rho > \tag{18.96}$$

Now the time averaged continuity equation (18.93) reads

$$\frac{d}{dt} < \rho >= -6H \frac{n}{n+1} < \rho > \tag{18.97}$$

Of course, for any continuity equation we expect the right hand side to be equal to $-3H(1+w)\rho$, having used the equation of state $P = w\rho$. By comparing we find an effective equation of state for the fast oscillating scalar field with[152, 176]

$$3(1+w) = 6\frac{n}{n+1} \tag{18.98}$$

equivalently

$$w = \frac{n-1}{n+1} \tag{18.99}$$

What we have found is that at late times, when the field is oscillating quickly, it dilutes as a perfect fluid with w. For $n = 1$

$$w(n=1) = 0 \tag{18.100}$$

which means the field dilutes as if it were pressureless matter. In this case, it is possible that a scalar field contributes at least in part to dark matter. For $n = 2$

$$w(n=2) = \frac{1}{3} \tag{18.101}$$

the scalar field acts as a radiation when oscillating. If the oscillation begins early enough, this can have a similar effect as that of a light relic in the early universe. For $n \geq 3$ the scalar field dilutes faster than any known standard model component. However, even if it is absent today, this does not means it was not present in the early universe. Even in this case, its cosmological effects must be studied.

To recap, we have found that a cosmological scalar field has a rich phenomenology. There is enough freedom in its parameters to describe either a dark energy or dark matter component. It may even transition from acting as a dark energy component in the early universe to a dark matter one at late times.

18.4 First order perturbations of the scalar field

As with the rest of the universe, we will split the scalar field χ into a time independent, homogeneous, term and a small perturbation[38, 181]

$$\chi(\tau, \vec{x}) = \bar{\chi}(\tau) + \delta\chi(\tau, \vec{x}) \tag{18.102}$$

As usual, it is more convenient to work with Fourier transform of $\delta\chi$. We will do so, and henceforth drop the tilde indicating the Fourier transform. So we will evaluate $\delta\chi(\vec{k}, \tau)$. Working out the equations of motion at first order is straightforward, if lengthy. It is sufficient to use the equations of motion (18.79)(or (18.78)) together with the Christoffel symbols at zero order (6.41) and at first order , in the conformal-Newtonian (13.43) or synchronous (13.49) gauges. The zero order equation (18.81) can be used to eliminate the second derivative of $\ddot{\bar{\chi}}$. With this, the equations of motion are

(Newtonian)

$$\delta\ddot{\chi} + 2\mathcal{H}\delta\dot{\chi} + k^2\delta\chi + a^2V^{''}(\bar{\chi})\delta\chi \quad = \quad \dot{\bar{\chi}}(\dot{\psi} + 3\dot{\phi}) - 2a^2\psi V^{'}(\bar{\chi}) \tag{18.103}$$

(Synchronous)

$$\delta\ddot{\chi} + 2\mathcal{H}\delta\dot{\chi} + k^2\delta\chi + a^2V^{''}(\bar{\chi})\delta\chi \quad = \quad -\frac{1}{2}\dot{h}\dot{\bar{\chi}}$$

where we will now use the prime to denote derivatives with respect to the scalar field, $V^{'}(\chi) = \frac{dV}{d\chi}$. The perturbed energy momentum tensor can be found using (18.80).

$$\text{(Newtonian)}\delta\rho = \frac{\delta\dot{\bar{\chi}}\dot{\chi}}{a^2} + V^{'}(\bar{\chi})\delta\chi - \frac{\psi\dot{\bar{\chi}}^2}{a^2} \tag{18.104}$$

$$\text{(Synchronous)}\delta\rho = \frac{\delta\dot{\bar{\chi}}\dot{\chi}}{a^2} + V^{'}(\bar{\chi})\delta\chi \tag{18.105}$$

$$T^0_i = -\frac{ik^i}{a^2}\dot{\bar{\chi}}\delta\chi \tag{18.106}$$

$$\theta = k^2\delta\chi \tag{18.107}$$

$$\delta P = \frac{\delta\dot{\chi}\dot{\bar{\chi}}}{a^2} - V^{'}(\bar{\chi})\delta\chi \tag{18.108}$$

The last equations are of the same form in both gauges. These quantities, of course, are a source to the metric and must be included in the Einstein equations (13.117)-(13.124). We did not impose any constraint on the perturbations and yet we have found no anisotropic stresses, $T^i_j = \delta P\delta^i_j$. This implies that vector and tensor modes of the perturbation equations don't affect, and aren't affected, by the presence of a scalar field at first order.

As expected, the field is only sourced by the gravitational field. In absence of the gravitational fields, the field acts somewhat as a damped harmonic oscillator with time dependent coefficients. For the homogeneous solutions the same arguments that we made for the zero order term holds. At early times the homogeneous solutions are frozen, and when

the Hubble factor decreases sufficiently they become oscillating or damped solutions. For larger values of k this period should begin earlier, while we expect that the longer wavelength modes (small k) remain frozen for a longer time. However, with respect to the the zero-order solution there is an important qualitative difference. The zero order equation can be assumed to have an initial value of the field set by primordial physics which will be in general different than zero. On the other hand, as we shall see shortly, adiabatic initial conditions for a scalar field imply the that the initial value of the perturbation $\delta\chi = \delta\dot{\chi} = 0$. Qualitatively, the freezing and thawing of the zero order mode does not happen (it could if the initial conditions where different). At early times, it is the in-homogeneous mode that dominates[38].

Let's take a mode well outside the horizon $k\tau \ll 1$.. At very early times, the zero order field $\bar{\chi}$ is frozen due to Hubble friction. More specifically we assume the initial condition is $\dot{\bar{\chi}} = 0$. Any other initial condition, with a time derivative, would be a decaying mode and wouldn't be relevant. With this assumption, the equation of state is given by (18.84), $w \simeq -1$. We had argued that adiabatic initial conditions implied, see (13.332),

$$\frac{\delta\rho}{\rho(1+w)} = \frac{3\delta_\gamma}{4} \tag{18.109}$$

Now we are working in the limit of $w \to -1$. Taking this limit, $\delta\rho$ must also be zero, in order for the fraction to give a constant.

$$\delta\rho_i = 0 \tag{18.110}$$

In the synchronous gauge. Therefore $\delta\chi = \delta\dot{\chi} = 0$ for adiabatic initial conditions. The same holds for the conformal-Newtonian gauge. Indeed by using (13.86)

$$\frac{\delta\rho_S}{\rho} = \frac{\delta\rho_N}{\rho} - \frac{\dot{\rho}}{\rho}\dot{S} \tag{18.111}$$

In the earliest phase $\dot{\rho}$ is constant, as can be verified using the continuity equation (7.21), or the explicit expression (18.89). Therefore the initial conditions for the field are zero in the conformal-Newtonian gauge as well. We note that, even if the field had different initial conditions, the Hubble friction would damp the homogeneous solution. To understand the evolution at earliest times, we must study the in-homogeneous solution.

For definiteness, we will now work in the conformal-Newtonian gauge. Since $k \ll \frac{1}{\tau}$ and $\mathcal{H} \propto \frac{1}{\tau}$ in the radiation and matter dominated eras we can neglect the term $k^2\delta\chi$ in (18.103). We may also neglect the term in the potential, as at initial times $\delta\chi$ is very small. On the right hand side, the derivatives of the metric terms are very close to zero as well, due to the initial conditions (13.325)-(13.326). Therefore we obtain in the radiation era

$$\delta\ddot{\chi} + \frac{2}{\tau}\delta\dot{\chi} \simeq -2C^2\tau^2\psi V^{'}(\bar{\chi}) \tag{18.112}$$

and in the matter dominated era

$$\delta\ddot{\chi} + \frac{4}{\tau}\delta\dot{\chi} \simeq -2D^2\tau^4 V^{'}(\bar{\chi}) \tag{18.113}$$

where we used (7.43) and (7.44) to make the τ dependence explicit. The constants C, D are defined by $a = C\tau$ (during radiation domination) and $a = D\tau^2$ (during matter domination). In both cases a solution can be found in the form

$$\delta\chi = A\tau^n \tag{18.114}$$

Plugging in and matching the exponents and the prefactors, the solutions are given by

$$\delta\chi \simeq -\frac{1}{15}a^2\tau^2\psi V^{'}(\bar{\chi}) \tag{18.115}$$

in the radiation dominated era, and

$$\delta\chi \simeq -\frac{1}{27}a^2\tau^2\psi V^{'}(\bar{\chi}) \tag{18.116}$$

in the matter dominated one. Eventually the modes enter the horizon and oscillation begins.

These solutions are very important in our understanding of the experimental signatures of the scalar field perturbations. They are sourced by the matter potential at very early times, and later remain interacting with them. This implies there will be correlations in the power spectra today with the metric potential and other quantities, including the temperature of the CMB.

We can put these solutions into a form which doesn't depend on $V^{'}$ explicitly by using the slow roll approximation. The slow roll approximation can be applied whenever the field $\bar{\chi}$ is frozen by Hubble friction and $\ddot{\bar{\chi}} \ll 2\mathcal{H}\dot{\bar{\chi}}$. In this approximation

$$a^2V^{'}(\bar{\chi}) \simeq -2\mathcal{H}\dot{\bar{\chi}} \tag{18.117}$$

Using $(\rho + P)a^2 = a^2\rho(1 + w) = \dot{\bar{\chi}}^2$ and $\Omega_\chi = \rho\frac{8\pi G a^2}{3\mathcal{H}^2}$

$$a^2V^{'}(\bar{\chi}) \simeq -2\mathcal{H}^2\sqrt{\frac{3}{8\pi G}\Omega_\chi(1 + w_\chi)} \tag{18.118}$$

where Ω_χ and $w \to w_\chi$ must be evaluated as functions of conformal time. For a scalar field the equation of state is not constant. Therefore, we obtain, for the growing mode during radiation domination

$$\delta\chi \simeq \frac{2}{15}\sqrt{\frac{3}{8\pi G}\Omega_\chi(1 + w_\chi)}\psi \tag{18.119}$$

and for the matter dominated era

$$\delta\chi \simeq \frac{8}{27}\sqrt{\frac{3}{8\pi G}\Omega_\chi(1 + w_\chi)}\psi \tag{18.120}$$

These equations show that, in practice, the perturbations are very small while the field is frozen. In fact $(1+w_x) \sim 10^{-10}$ in a typical model before the oscillations of the background field begin.

Let's look at later times, when the the Hubble friction term is smaller and we must consider all terms. Then, equation (18.103) is a damped and forced harmonic oscillator with

time dependent coefficients. All terms are relevant and a numerical technique is employed to get an exact solution. We want to gain a qualitative understanding of the form of the solutions, in particular, when they are oscillating and when they are damped. We introduce the damping ratio

$$\varsigma(\tau) = \frac{\mathcal{H}^2}{\sqrt{k^2 + a^2 V''(\bar{\chi})}} \tag{18.121}$$

In absence of time dependence this would simply indicate whether the oscillator is underdamped $\varsigma < 1$, and the solutions are oscillations with decreasing amplitudes, or overdamped $\varsigma > 1$, and the solutions are exponentially decreasing. This description is valid once the field begins to thaw and move towards the minimum of the potential. This happens when either $\mathcal{H}^2 \sim k^2$ or $\mathcal{H}^2 \sim a^2 V''$. Since $\mathcal{H}$ is a decreasing function, and $a^2 V''$ depends only on the background terms, for small values of k we expect all the modes to begin thawing at the same time. This time depends specifically on the form of the potential and the evolution of the background field. At the opposite limit, when $k^2 \gg a^2 V''$, we expect the particular form of the potential to play a very small role in the evolution. Any observable related to large values of k does not depend strongly on V.

As long as the quantities in (18.121) change slowly compared to the typical time of an oscillation $T \sim \frac{2\pi}{\sqrt{k^2 + a^2 V''(\bar{\chi})}}$ then the damped harmonic oscillator intuition is correct. For a potential of the form $V(\chi) \propto \chi^{2n}$ this limit is

$$(n-2)\frac{d \ln \bar{\chi}}{d\tau} \ll \frac{1}{\sqrt{k^2 + a^2 V''(\bar{\chi})}} \tag{18.122}$$

and we may presume the field decays exponentially if $\varsigma(\tau) > 1$ and oscillates if $\varsigma(\tau) < 1$.

One may believe that since the oscillation begins at $\mathcal{H} \sim k^2$ or $\mathcal{H}^2 \sim a^2 V''$ that $\varsigma(\tau) > 1$ always when oscillating. However, the oscillation begins at this stage, while both terms contribute in the damping ratio which could make the ratio less than one and allow oscillations at a later time.

This analysis assumes that the forcing term due to the gravitational potentials does not alter the situation. At least in a qualitative description of the phenomena, this turns out to be a good description. Indeed, the matter potentials are not expected to change over the time scale T, which arises only from the scalar field.

18.5 Solving the H_0 tension with early dark energy

The Hubble tension is the open problem of reconciling measurements of H_0[104, 160]. Local measurements from supernovas[180] disagree with the value inferred from the CMB[61]. In particular, it seems the local measurements give a larger value of H_0 than what is inferred from the CMB. If the discrepancy cannot be accounted for by systematics, some new physics could play a role in the discrepancy. Late-time physics, happening after decoupling, is constrained by several cosmological observables. Some new physics before decoupling may be able to account for the discrepancy and be consistent with a vast set of observables.

The H_0 tension may almost be solved by extra ultra-relativitistic degrees of freedom ΔN_{eff}[22]. As we pointed out in section 17.2, if we increase ΔN_{eff} while changing other cosmological

parameters to keep the equivalence scale factor a_{eq} and the sound horizon r_s fixed, we have a decrease in the power at large ℓ in the CMB power spectrum. Let's explore this fact a little more[177]. We will come to understand why an extra ultra-relativistic degree of freedom cannot solve the tension, but an early scalar field may.

The position of the first peak in the CMB is the angle subtended by the sound horizon at decoupling, as seen today. This must be the ratio between the sound horizon at decoupling (13.295) $r_s(\tau_{LSS})$ and the angular distance to the last scattering surface (7.119) $d_A(\tau_{LSS})$. If we increase ΔN_{eff}, we increase the Hubble factor in the early universe, and this causes a *decrease* of $r_s(a_{LSS})$, since $r_s(a_{LSS}) = \int c_s da/\mathcal{H}$. In order to keep the position of the first CMB peak fixed, we must decrease the angular distance to the last scattering surface. This can be accomplished by an *increase* in H_0. In a matter dominated universe, to simplify the discussion, the angular distance is

$$d_A(\tau_{LSS}) = \frac{(\tau_0 - \tau_{LSS})}{1 + z_{LSS}} = \frac{2}{H_0}\frac{1}{1 + z_{LSS}}(1 - \sqrt{a_{LSS}}) \tag{18.123}$$

z_{LSS}(equivalently a_{LSS}) is well fixed by theory and experiment, since $a_{LSS}T_{LSS} = a_0 T_{CMB}$ and T_{LSS} is well known from the theory of recombination. Therefore, we conclude that increasing N_{eff} causes the inferred value of H_0 from experiment to increase. There is a second, subdominant effect of N_{eff}: to decrease the damping scale (17.28), adding more power on small scales. However, when one changes H_0 (and Ω_{cdm} to keep a_{eq} consistent with matter power spectrum measurements), we obtain a net *decrease in power* on small scales. This is the peculiar effect of N_{eff} which cannot be mimicked by changing other parameters. All that remains is to check whether the increase in N_{eff} we need, in order to infer a higher value of H_0 from the data, does not cause too much damping of the small scales. In fact, an MCMC analysis shows that it does. We cannot increase sufficiently N_{eff} to solve the Hubble tension[22].

A scalar field could work[177], so long as it begins oscillating and losing energy density around the era of matter-radiation equality. If the scalar field would remain frozen, it would give some effect similar to that of the increased N_{eff}, including its degeneracies, as we just described. If it begins oscillating much sooner than matter radiation equality, it would dilute as a regular fluid also give similar effects, but of a different magnitude. It becomes interesting when the oscillation happens between matter-radiation equality and decoupling. The reason is that the damping scale depends on the Hubble factor until $\sim a_{eq}$, whereas the sound horizon on the value until decoupling. If the field decays after a_{eq} then H will be anomalously large before, but less so after. Including such a scalar field would cause both the sound horizon and the damping scale to decrease, while the ratio $\frac{r_s}{r_d}$ would decrease faster than it would with a simple change in N_{eff}. The scalar field affects the damping scale relatively more than r_s. When we remove the change in sound horizon by fixing the dark matter density and Hubble constant, the damping scale increases again. However, since the decrease was larger than it was with N_{eff}, it now becomes closer to its usual value. The predicted power spectrum at small scales is consistent with observations of the CMB, and we have raised the inferred H_0. All this qualitative discussion is confirmed by a detailed MCMC analysis[177].

There is some dependence on the potential and the simplest cases are not possible. If we

consider "axion-like" potentials

$$V(\chi) = \Lambda^4(1 - \cos\frac{\chi}{f})^n \tag{18.124}$$

with Λ and f being some parameters to be found by experiment, then $n = 1$ is not a viable solution to the problem. This can be understood by the fact that a scalar field with a potential $\sim \chi^{2n}$, as this one is around the minimum, dilutes as cold dark matter at late times, after it has begun oscillating. It would then contribute in an unacceptable way to the evolution of the universe at late times.

Potentials with $n \geq 2$ are viable candidates. Such potentials can be generated by multi-instantons and often arise in the context of string theory. A detailed analysis is performed in [177] using a MCMC as we described in section 16.5.

Following [176, 177], we define the critical scale factor a_c as the value of the scale factor when the homogeneous field is equal to $\frac{7}{8}\bar{\chi}_i$, where $\bar{\chi}_i$ is its initial value. And $f_{ede}(a_c) = \frac{\Omega_{EDE}}{\Omega_{Tot}}$ the fraction of scalar field, or early dark matter, at a_c. Then it is found that the Hubble tension is ameliorated for

$$(n = 2): \begin{cases} \log_{10}(a_c) = -4.136^{+0.56}_{-0.013} \\ f_{ede}(a_c) = 0.028^{+0.011}_{-0.016} \end{cases} \tag{18.125}$$

$$(n = 3): \begin{cases} \log_{10}(a_c) = -3.737^{+0.110}_{-0.094} \\ f_{ede}(a_c) = 0.050^{+0.024}_{-0.019} \end{cases} \tag{18.126}$$

where the errors represent a 1σ interval on the posterior distributions obtained on the values.

With this new model, the predicted power spectrum is well consistent with experiment. Future cosmic variance limited experiments may be able to detect signatures of the scalar field dynamics in the CMB T, E spectra.

19 CMB Rotation Spectra

19.1 Vacuum birefringence from a Pseudo Nambu-Goldstone boson

A scalar field which arises through the spontaneous breaking of a symmetry is a massless Goldstone boson[170]. If the broken symmetry is not exact, then the scalar field acquires a small mass. We had argued in section 18.2 that the axion is such a *Pseudo Nambu-Goldstone boson* (PNGB). Indeed, any scalar field with a small potential may be thought to arise from some broken high energy symmetry[137, 6]. We had argued that, if the current of this symmetry is anomalous, there is a generic coupling to the photon of the type (18.74)

$$\mathcal{L}_{\chi\gamma\gamma} = \frac{G_{\chi\gamma\gamma}}{4}\chi F_{\mu\nu}\tilde{F}^{\mu\nu} \tag{19.1}$$

where χ is the PNGB field, $F_{\mu\nu}$ the electromagnetic strength tensor and $\tilde{F}^{\mu\nu} = \frac{1}{2}\varepsilon^{\mu\nu\rho\sigma}F_{\rho\sigma}$ its dual. G is a coupling constant with dimensions of negative energies and, up to factors

of order unity, is the inverse of the energy scale of symmetry breaking. Most experimental searches focus on detecting this coupling[39]. Indeed, the properties of electromagnetism are very well tested and the form of the photon coupling is unique and does not appear otherwise. If there is a background uniform axion field, this coupling is an effective parity-violating operator for photons. This provides a unique phenomenon known as *vacuum birefringence[44, 41]* .

The insight into birefringence is due to the fact that the interaction Lagrangian (19.1) can be written as

$$\mathcal{L}_{\chi\gamma\gamma} = -G_{\chi\gamma\gamma}\chi\vec{E}\cdot\vec{H} \tag{19.2}$$

with $\vec{H}$ and $\vec{E}$ being the magnetic and electric vectors of the field, defined by $F^{i0} = E^i$ and $F^{ij} = -\varepsilon^{ijk}H^k$. This coupling can be studied by observing photons passing through either a strong external field χ or external magnetic field $\vec{H}$. Suppose there is a photon with wave vector $\vec{k}$ traveling through a strong magnetic field and we choose, for simplicity, $\vec{k} \perp \vec{H}$. The photon has two possible polarization states in the plane orthogonal to $\vec{k}$. The coupling preferentially selects the polarization which is parallel to the magnetic field. It is possible to show that the two orthogonal polarizations evolve with different refractive indexes, or that they acquire a relative phase. This causes optical birefringence in a vacuum, as the only "medium" present would be the magnetic field. The effect is that of producing a circular polarization starting from a linear one. This phenomena exhibits many interesting properties which are similar to those of neutrino oscillations[112], such as a possible electromagnetic MSW effect, and adiabatic evolution of the polarization eigenstates as the external field changes[41].

Although the case of an external magnetic field is interesting, we will focus on the case of an external scalar field χ. If such a field is present in cosmology, it may allow for vacuum birefringence in the CMB polarization. In this case, the effect of the coupling is to provide a preferential basis for the polarization of a photon in its transverse plane. In general, one is allowed to make any arbitrary choice of basis in the transverse plane. However, the form of the coupling is telling us that in a vertex $\gamma+\chi\to\gamma$, the polarization of the photon from the initial to final state flips from one linear polarization to the orthogonal direction. In order for the interaction to be non-zero, $\vec{E}_i \parallel \vec{H}_f$, which implies that $\vec{E}_i \perp \vec{E}_f$. The polarization is rotated by $\frac{\pi}{2}$ through a single interaction vertex. An eigenstate of the evolution through a scalar field must be such that the polarization does not change through an interaction. Clearly the only possibilities are circular polarizations. These eigenstates will evolve with different effective refractive indexes and, as we shall see, the net result is a rotation of the polarization vector along the world-line of the photon. We shall now calculate this effect.

To start, we write the action of all relevant electromagnetic terms

$$S = \int d^4x \left[-\frac{1}{4}F_{\mu\nu}F^{\mu\nu} + \frac{G_{\chi\gamma\gamma}}{4}\chi F_{\mu\nu}\tilde{F}^{\mu\nu}\right] \tag{19.3}$$

We neglect interactions with matter since we will be interested in the equation of motion through a vacuum. We also work in flat spacetime, as we are only interested in the polarization. On a curved space-time, polarization of a photon is simply parallel transported along its worldline. Replacing $\partial_\mu \to \nabla_\mu$ in the definition of $F_{\mu\nu}$, as is the prescription of working in a curved space time, has no effect. Passing to the covariant derivative means

adding Christoffel symbols which are symmetric in $\mu \leftrightarrow \nu$ and cancel out. This is true so long as spacetime is *torsionless*, which mathematically implies the Christoffel symbols are symmetric in the lower indexes $\Gamma^{\alpha}_{\mu\nu} = \Gamma^{\alpha}_{\nu\mu}$. Furthermore, the field χ appears without any derivative, so there is no change in the interaction term as well when considering curved spacetimes. The only addition we'd get for a curved spacetime would be the overall $\sqrt{-g}$ factor which would provide the relevant parallel transport, but not affect polarization otherwise.

Using the above action, we derive the equations of motion of the electromagnetic field in the usual way

$$\Box A^{\mu} - \partial^{\mu}\partial_{\nu}A^{\nu} = -G_{\chi\gamma\gamma}\varepsilon^{\mu\nu\rho\sigma}\partial_{\nu}\chi\partial_{\sigma}A_{\rho} \tag{19.4}$$

We will work in the Lorenz gauge $\partial_{\mu}A^{\mu} = 0$. The equation reduces to

$$\Box A^{\mu} = -G_{\chi\gamma\gamma}\varepsilon^{\mu\nu\rho\sigma}\partial_{\nu}\chi\partial_{\sigma}A_{\rho} \tag{19.5}$$

We want to solve these equations for the photon polarization. To be definite we will choose a wave-vector in the $\hat{z}$ direction $k^{\mu} = (\omega, 0, 0, \omega)$. The world-line of the photon has an everywhere tangent vector $(1, 0, 0, 1)$ and we denote λ the affine parameter of the worldline. With this in mind, we make the following plane wave ansatz

$$A^{\mu} = \xi^{\mu}(\lambda)e^{ik_{\mu}x^{\mu}} \tag{19.6}$$

allowing the amplitude to depend on the affine parameter only. Through λ, ξ^{μ} depends on z and t. Then the D'Alembert operator is

$$\Box = (\partial_t + \partial_z)(\partial_z - \partial_t) \tag{19.7}$$

We may note that $\partial_z + \partial_t = \frac{d}{d\lambda}$ is the total derivative along the worldline. The equation of motion is

$$\frac{d\xi^{\mu}}{d\lambda} = -\frac{G_{\chi\gamma\gamma}}{2\omega}\varepsilon^{\mu\nu\rho\sigma}\partial_{\nu}\chi k_{\sigma}A_{\rho} \tag{19.8}$$

By the Lorentz gauge condition $\xi^{\mu}k_{\mu} = 0$ and so there are only two components to consider $\mu = 1, 2$. ξ^{μ} is the polarization vector. This equation can be put in a form[22], typical of neutrino oscillations as well,

$$\frac{d}{d\lambda}\begin{pmatrix}\xi^1\\ \xi^2\end{pmatrix} = -\frac{G_{\chi\gamma\gamma}}{2}\frac{d\chi}{d\lambda}\begin{pmatrix}0 & 1\\ -1 & 0\end{pmatrix}\begin{pmatrix}\xi^1\\ \xi^2\end{pmatrix} \tag{19.9}$$

Clearly the two linear polarizations are coupled. This matrix may be diagonalized by

$$\begin{pmatrix}\xi^1\\ \xi^2\end{pmatrix} = R\begin{pmatrix}\xi^+\\ \xi^-\end{pmatrix} \tag{19.10}$$

with

$$R = \frac{1}{\sqrt{2}}\begin{pmatrix}1 & 1\\ i & -i\end{pmatrix} \tag{19.11}$$

[22]Note that $\varepsilon_{0123} = 1$, and so in Minkoski space $\varepsilon^{0123} = -1$.

The eigenvectors are circular polarizations, as we had anticipated. The above equation becomes

$$R^{-1}\frac{d}{d\lambda}(R\begin{pmatrix}\xi^+\\ \xi^-\end{pmatrix}) = -\frac{G_{\chi\gamma\gamma}}{2}\frac{d\chi}{d\lambda}R^{-1}\begin{pmatrix}0 & 1\\ -1 & 0\end{pmatrix}R\begin{pmatrix}\xi^+\\ \xi^-\end{pmatrix} \tag{19.12}$$

We chose to write this intermediate step explicitly so we may note that on the left hand side, the derivative may pick up a term $\frac{dR}{d\lambda}$, which could lead to mixing of modes *even in the diagonal basis*. This may be present in more general problems. In our simple case, R does not depend on λ and we simplify to

$$\frac{d}{d\lambda}\begin{pmatrix}\xi^+\\ \xi^-\end{pmatrix} = -\frac{G_{\chi\gamma\gamma}}{2}\frac{d\chi}{d\lambda}\begin{pmatrix}i & 0\\ 0 & -i\end{pmatrix}\begin{pmatrix}\xi^+\\ \xi^-\end{pmatrix} \tag{19.13}$$

The solutions are given by

$$\xi^{\pm}(\lambda) = \xi_i^{\pm}\exp\mp i\alpha(\lambda) \tag{19.14}$$

where

$$\alpha(\lambda) = \frac{G_{\chi\gamma\gamma}}{2}\int_0^{\lambda} d\lambda'\frac{d\chi}{d\lambda} = \frac{G_{\chi\gamma\gamma}}{2}\Delta\chi \tag{19.15}$$

with $\Delta\chi$ being the difference between the scalar field at the initial and final positions. $\xi_i^{\pm}$ are the initial conditions. To illustrate the phenomena, we choose an initial linear polarization. Without loss of generalization this may be $\xi_i^1 = \xi_0$, $\xi_i^2 = 0$ and therefore $\xi_i^{\pm} = \xi_0/\sqrt{2}$. Returning to the basis of linear polarizations the solutions are then given by

$$\xi^1(\lambda) = \xi_0\cos\alpha(\lambda) \tag{19.16}$$

$$\xi^2(\lambda) = \xi_0\sin\alpha(\lambda) \tag{19.17}$$

The net effect of the birefringence due to scalar field is to *rotate the polarization by an angle α which only depends on the initial and final values of the field*. The field may have any complicated space and time dependence.

19.2 CMB rotation spectra

Suppose a cosmic scalar field were present at around decoupling time and has a coupling to the photon as in (19.1). Then we can expect the polarization from the CMB to be rotated by an amount which depends on the difference between the field at the emission of the photons and today at the Earth, given by (19.15)[38, 175, 114, 52, 174]. This picture is complicated by the fact that the time of emission of photons of the CMB is not τ_{LSS} but rather a distribution $g(\tau)$ centered around τ_{LSS}. $g(\tau)$ is the visibility function, defined in (13.366)[40].

Let's work at zero-order first, and assume the scalar field is a uniform $\bar{\chi}$ only dependent on conformal time. This provides a uniform rotation angle across the sky given by

$$\alpha_0 = \frac{G_{\chi\gamma\gamma}}{2}\left(\bar{\chi}(\tau_0) - \int_0^{\tau_0} d\tau g(\tau)\bar{\chi}(\tau)\right) \tag{19.18}$$

In writing this, we have made an approximation. Not every photon of the CMB can be considered to be polarized. So we should consider a distribution $g(\tau)$ which describes the probability distribution of emission of *polarized* photons. This change is a smaller order effect and we neglect it. It is small not just because polarization itself is first order, but because $G_{\chi\gamma\gamma}$ is expected to be very small.

The zero order birefringence angle contains two terms. The first is proportional to the field today at Earth. If the field is oscillating with time, it may induce a time dependence in polarization measurements at Earth, irrespective of the original cosmological source. This time dependence has not been found, but future experiments may be sensitive to it. The second term is of cosmological origin

$$\bar{\alpha}(\tau) = -\frac{G_{\chi\gamma\gamma}}{2}\int_0^\tau d\tau^{'} g(\tau^{'})\bar{\chi}(\tau^{'}) \tag{19.19}$$

Once one has solved the zero order equations of the universe and recombination, to obtain the free electron density and thus $g(\tau)$, this integral can be calculated and the uniform rotation angle estimated. We did this with the Boltzmann code CLASS[26] as described in section 15.5. Of course the measurable angle today is $\bar{\alpha}(\tau_0)$.

We may also realize that the integral is of the line of sight form as, for example, (13.365). We can unpack it and write it as a Boltzmann like equation. Deriving both side with respect to conformal time and adding $g(\tau)$ explicitly

$$\dot{\bar{\alpha}} + an_e\sigma\bar{\alpha} = -\frac{G_{\chi\gamma\gamma}}{2}an_e\sigma_T\bar{\chi} \tag{19.20}$$

This may solved numerically as an alternative to integrating (19.19). The equation may be written in a covariant form. In the homogeneous frame, there is no velocity term for the electrons, and so the electron four-current is

$$j^\mu = \frac{1}{a}(n_e, 0, 0, 0) \tag{19.21}$$

We know that this is a proper four-vector which transforms in the usual way under coordinate transformations. Since $\bar{\alpha}$ is a scalar under coordinate transformations, we may replace the usual derivatives with covariant derivatives. In addition, the zero order $\bar{\alpha}$ is independent of position. Combining all this, the differential equation can be written as

$$\nabla_\mu\bar{\alpha} - \sigma_T j_\mu\bar{\alpha} = \frac{G_{\chi\gamma\gamma}}{2}\sigma_T j_\mu\bar{\chi} \tag{19.22}$$

The usefulness of this equation is that it is covariant. We may now take the first order perturbations in a gauge invariant manner. We assume the quantities are perturbed by small position dependent quantities $\bar{\chi} \to \bar{\chi}+\delta\chi$, $\bar{\alpha} \to \bar{\alpha}+\delta\alpha$, $n_e \to n_e+\delta n_e$. The perturbed current is now $j^\mu = \frac{1}{a}(n_e + \delta n_e, \vec{v_e})$ where $\vec{v}_e$ is the velocity field of the electrons. We consider the equation with $\mu = 0$. Care must be taken as the metric may enter into the lowered index current j_μ. The linearized equations are, in the synchronous (13.26) and

conformal-Newtonian (13.24) gauges,

$$\begin{aligned} &\text{(Synchronous)} \\ \delta\dot{\alpha} + an_e\sigma_T\delta\alpha + a\sigma_T\delta n_e\bar{\alpha} &= \\ &= -\frac{G_{\chi\gamma\gamma}}{2}a\sigma_T(n_e\delta\chi + \delta n_e\bar{\chi}) \end{aligned} \tag{19.23}$$

$$\begin{aligned} &\text{(Newtonian)} \\ \delta\dot{\alpha} + an_e\sigma_T\delta\alpha &= \\ &= -\frac{G_{\chi\gamma\gamma}}{2}a\sigma_T n_e\delta\chi - a\sigma_T(\bar{\alpha} + \frac{1}{2}G_{\chi\gamma\gamma}\bar{\chi})(2\psi n_e + \delta n_e) \end{aligned} \tag{19.24}$$

The perturbation δn_e is essentially a perturbation to the visibility function. This quantity does not usually appear in the first order equations since $g(\tau)$ is always multiplying other first order terms. In this situation, it is unavoidable that a perturbation to $g(\tau)$ contribute. In fact, a perturbation to the vacuum birefringence angle may arise because there is a fluctuation in the scalar field *or because there is a fluctuation in the position of emission*. This latter fact was neglected in previous literature where the approximation $g(\tau) \simeq \delta(\tau - \tau_{LSS})$ was often used.

The equations for the rotation angle do not depend on spatial derivatives, as such equations (19.23) and (19.24) are valid in Fourier space as well, where all the quantities that appear are the Fourier transforms. This equation can then be solved together with the rest of the Boltzmann equations to get $\delta\alpha(\vec{k}, \tau_0)$. A typical Boltzmann code, such as CLASS[26] or CAMB[151], does not track the δn_e perturbation, nor $\vec{v}_e$, since these never appear in the ordinary equations. Integration could proceed by including full Boltzmann equations for electrons. However we know that electrons and baryons(protons and Helium) are extremely tightly coupled, which is the reason why the electrons are not tracked on their own. Indeed, in the cosmology community by "baryons" one means precisely the coupled fluid. The tight coupling would also result in numerically difficult to solve equations, since the collision term would be very large. This issue is already present when dealing with the coupled baryons and photons equations, and is solved precisely by using a tight coupled limit. With these considerations, we make the approximation $\delta n_e \simeq \delta_b n_e$, which is valid at all the relevant times, when these particles are non-relativistic.

On the numerical side, we note that we do not need to know the full evolution of $\delta\alpha$ in general, since $\delta\alpha$ does not enter into any other Boltzmann equations. So we can formally integrate the differential equations to obtain an integral over a source term

$$\delta\alpha(\tau_0, \vec{k}) = \int_0^{\tau_0} S_\alpha(\tau, \vec{k})d\tau \tag{19.25}$$

In the conformal-Newtonian gauge the source term is given by

$$S_\alpha(\tau, \vec{k}) = -g(\tau)\left(\frac{1}{2}G_{\chi\gamma\gamma}\delta\chi + (\bar{\alpha} + \frac{1}{2}G_{\chi\gamma\gamma}\bar{\chi})(2\psi + \delta_b)\right) \tag{19.26}$$

whereas the same form is valid in the synchronous gauge by setting $\psi = 0$. With these last formulas, it is clear one can solve the usual Boltzmann equations with the addition of a scalar field, at zero and first order. Once all the equations have been solved, the integral of the source can be calculated to obtain the perturbation to the rotation angle. We describe how this is achieved in the CLASS code in section 15.6.

We now relate the calculated perturbations to the main observable, which is the CMB rotation power spectrum $C_\ell^{\alpha\alpha}$. The polarization on the sky today is rotated by an angle with respect to the case $G_{\chi\gamma\gamma} \to 0$ by

$$\delta\alpha(\hat{n}) = \sum_{\ell=1,M} a^{\alpha}_{\ell m} Y^m_\ell(\hat{n}) \tag{19.27}$$

Then

$$< a^{\alpha}_{\ell m} a^{\alpha *}_{\ell' m'} >= \delta_{\ell\ell'}\delta_{mm'} C^{\alpha\alpha}_\ell \tag{19.28}$$

Repeating the same calculations that give us the temperature power spectrum C_ℓ^{TT}, through the underlying perturbations, equation (14.17), we obtain

$$C^{\alpha\alpha}_\ell = \frac{2}{\pi}\int k^2 dk P_\psi(k) \left(\Delta_{\alpha,\ell}(k)\right)^2 \tag{19.29}$$

where the transfer function

$$\Delta_{\alpha,\ell}(k) = \int_0^{\tau_0} d\tau S(\vec{k},\tau) j_\ell(k(\tau_0 - \tau)) \tag{19.30}$$

Only scalar modes have been considered here, as there are no vector and tensor modes for the scalar field. P_ψ appears, as these are normalized to the Boltzmann equations solved with $\psi_i = 1$. Since the scalar field is sourced by the metric potential through (18.115)-(18.116) there must be cross-correlations with the temperature and E modes as well.

$$C^{\alpha X}_\ell = \frac{2}{\pi}\int k^2 dk P_\psi(k) \Delta_{\alpha,\ell}(k) \Delta^{(0)}_{X,\ell}(k) \tag{19.31}$$

where $X = T, E$ and $\Delta^{(0)}_{X,\ell}$ are given by eg. (14.14).

In addition to the rotation power spectrum, there is the uniform angle of rotation which is in principle observable[114, 52, 174]. The uniform angle is given by the sum of the zero order fields plus the monopole perturbation at earth, which is stochastic in nature. Since we can at most measure the single realization of the uniform angle, we couldn't say how much of it is given by the zero order term, and how much the stochastic perturbation has contributed. Therefore, it can serve to indicate that birefringence has taken place, but gaining more details from it would be challenging, save for possible order of magnitude estimates. This is why, although smaller, study of the rotation power spectra can give vastly more insight on the scalar field.

The source term (19.26) contains contribution both from the fluctuations of the scalar field *and* the matter and metric potential[40]. We expect to see the acoustic oscillations in the rotation spectra as well. This is actually a very precise prediction on the shape of the $C_\ell^{\alpha\alpha}$

at high ℓ. The normalization of the effect depends strongly on the value of $G_{\chi\gamma\gamma}\bar{\chi}$ at decoupling.

Of course, defining these quantities and being able to calculate them, does not mean we may measure them. It is a well posed question to understand how one can distinguish a change in polarization between last scattering and today, from an intrinsic polarization. After all, at any single point in the sky we only measure the Stoke parameters I, Q, U in some basis. Luckily, the correlations between angular modes on the sky enable us to separate the primordial Stoke parameters, from the change between decoupling and today[114, 134].

In the absence of any birefringence, the polarization today is described by the Stoke Parameters Q, U. The parametrization through the definite spin-s quantities (13.188) $(Q \pm iU)$ is most convenient. This also tells us that if the polarization in a direction $\hat{n}$ in the sky was rotated by $\alpha(\hat{n})$ counterclockwise around the direction of the incoming photon the spin-s quantities get transformed into

$$(\tilde{Q} \pm i\tilde{U})(\hat{n}) = e^{\mp 2i\alpha(\hat{n})}(Q \pm iU)(\hat{n}) \tag{19.32}$$

We wish to separate out the primordial contribution from the rotation due to birefringence. As usual, it is best to work with spherical harmonics. We define the observed angular coefficients through

$$(\tilde{Q} \pm i\tilde{U}) = \sum_{\ell m} (\tilde{a}^E_{\ell m} \pm i\tilde{a}^B_{\ell m})\, {}_{\pm 2}Y^m_\ell(\hat{n}) \tag{19.33}$$

analogously to (14.18). The primordial modes have the same expansion with the primordial $a^E_{\ell m}$ and $a^B_{\ell m}$. The angle is expanded as

$$\alpha(\hat{n}) = \sum_{LM} a^\alpha_{LM} Y^M_L(\hat{n}) \tag{19.34}$$

which includes the $L = 0$ term a^α_{00}, the uniform rotation across the sky.

Clearly, if we measure $(\tilde{Q} \pm i\tilde{U})$ and calculate the $\tilde{C}_\ell$ through $\tilde{a}^{E,B}_{\ell m}$ we will get a different value than what would be obtained in the absence of rotation. Our goal is now to relate the two quantities. For convenience we note

$$a^E_{\ell m} = \frac{1}{2}\int d\Omega({}_2Y^{m*}_\ell(\hat{n})(Q + iU) + {}_{-2}Y^{m*}_\ell(\hat{n})(Q - iU)) \tag{19.35}$$

$$a^B_{\ell m} = \frac{1}{2i}\int d\Omega({}_2Y^{m*}_\ell(\hat{n})(Q + iU) - {}_{-2}Y^{m*}_\ell(\hat{n})(Q - iU)) \tag{19.36}$$

with similar identities holding for the rotated quantities. Through these, we may calculate the rotated $\tilde{a}^B_{\ell m}$ in the small rotation angle approximation.

$$\tilde{a}^B_{\ell m} = \frac{1}{2i}\int d\Omega\,[{}_2Y^{m*}_\ell(Q + iU)(1 - 2i\alpha(\hat{n})) - {}_{-2}Y^{m*}_\ell(Q - iU)(1 + 2i\alpha(\hat{n}))] \tag{19.37}$$

Expanding everything into their own spherical coefficients

$$\begin{aligned}\tilde{a}^B_{\ell m} &= a^B_{\ell m} - \sum_{LM,\ell' m'} \int d\Omega a^\alpha_{LM} Y^M_L \times \\ &\left[(a^E_{\ell' m'} + i a^B_{\ell' m'})\, {}_2Y^{m'}_{\ell'}\, {}_2Y^{m*}_{\ell} + (a^E_{\ell' m'} - i a^B_{\ell' m'})\, {}_{-2}Y^{m'}_{\ell'}\, {}_{-2}Y^{m*}_{\ell}\right]\end{aligned}$$

The triple integral on spherical harmonics is expressed through Wigner-3j symbols (see appendix B) through

$$\begin{aligned}\int d\Omega\, {}_SY^{M*}_L(\hat{n})\, {}_{s1}Y^{m_1}_{\ell 1}(\hat{n})\, {}_{s_2}Y^{m_2}_{\ell_2}(\hat{n}) &= (-1)^{S+M}\sqrt{\frac{(2L+1)(2\ell_1+1)(2\ell_2+1)}{4\pi}} \times \\ &\begin{pmatrix} L & \ell_1 & \ell_2 \\ -M & m_1 & m_2 \end{pmatrix}\begin{pmatrix} L & \ell_1 & \ell_2 \\ S & -s_1 & -s_2 \end{pmatrix}\end{aligned}$$

Using this formula, and the fact that the Wigner-3j symbols possess the symmetry property

$$\begin{pmatrix} \ell_1 & \ell_2 & \ell_3 \\ m_1 & m_2 & m_3 \end{pmatrix} = (-1)^{\ell_1+\ell_2+\ell_3}\begin{pmatrix} \ell_1 & \ell_2 & \ell_3 \\ -m_1 & -m_2 & -m_3 \end{pmatrix} \tag{19.38}$$

We obtain

$$\begin{aligned}\tilde{a}^B_{\ell m} &= a^B_{\ell m} - 2(-1)^m \sum_{LM,\ell' m'} a^\alpha_{LM}\sqrt{\frac{(2\ell+1)(2\ell'+1)(2L+1)}{4\pi}}\begin{pmatrix} \ell & \ell' & L \\ -m & m' & M \end{pmatrix}\begin{pmatrix} \ell & \ell' & L \\ 2 & -2 & 0 \end{pmatrix} \times \\ &\times \left[a^E_{\ell' m'}\delta^{\text{even}}_{\ell+\ell'+L} + i a^B_{\ell' m'}\delta^{\text{odd}}_{\ell+\ell'+L}\right]\end{aligned}$$

A similar calculation gives

$$\begin{aligned}\tilde{a}^E_{\ell m} &= a^E_{\ell m} - 2i(-1)^m \sum_{LM,\ell' m'} a^\alpha_{LM}\sqrt{\frac{(2L+1)(2\ell+1)(2\ell'+1)}{4\pi}}\begin{pmatrix} \ell & \ell' & L \\ -m & m' & M \end{pmatrix}\begin{pmatrix} \ell & \ell' & L \\ 2 & -2 & 0 \end{pmatrix} \times \\ &\times \left[a^E_{\ell' m'}\delta^{\text{odd}}_{\ell+\ell'+L} + i a^B_{\ell' m'}\delta^{\text{even}}_{\ell+\ell'+L}\right]\end{aligned}$$

The rotation mixes the E and B modes, so we expect B modes even if primordial ones are not present. Let's calculate the $\tilde{C}^{EE}_L$. To do so we will assume a *fixed realization* of a^α_{LM}. It is not that a^α_{LM} is not stochastic, we are interested in separating out the C^{EE}_L from a^α_{LM}. Assuming there are no primordial B modes, $a^B_{\ell m} = 0$, we calculate $< \tilde{a}^B_{\ell m}\tilde{a}^{E*}_{\ell' m'} >= C^{EB}_{\ell m;\ell' m'}$. This quantity is zero in the Λ CDM model. It is so because of the conservation of parity. Noting that $< a^E_{\ell m}a^{E*}_{\ell' m'} >= C^{EE}_\ell \delta_{\ell\ell'}\delta_{mm'}$ and $< a^E_{\ell m} >= 0$, it is straightforward to obtain[134, 114]

$$\begin{aligned}< \tilde{a}^B_{\ell m}\tilde{a}^{E*}_{\ell' m'} > &= (-2)(-1)^m \sum_{LM}\sqrt{\frac{(2\ell+1)(2\ell'+1)(2L+1)}{4\pi}} \times \\ &\begin{pmatrix} \ell & \ell' & L \\ -m & m' & M \end{pmatrix}\begin{pmatrix} \ell & \ell' & L \\ 2 & -2 & 0 \end{pmatrix} C^{EE}_L \delta^{\text{even}}_{\ell+\ell'+L} a^\alpha_{LM}\end{aligned} \tag{19.39}$$

In particular, we can separate out the term at $L = 0$ ($M = 0$) using the identity

$$\begin{pmatrix} \ell & \ell^{'} & 0 \\ -m & m^{'} & 0 \end{pmatrix} = \frac{1}{\sqrt{2\ell+1}} \delta_{m,m'} \delta_{\ell,\ell'} (-1)^{\ell+\ell'} \tag{19.40}$$

The $C^{EB}_{\ell m;\ell' m'}$ due to only a uniform rotation is

$$-\frac{1}{\sqrt{\pi}} C^{EE}_{\ell} a^{\alpha}_{00} \delta_{\ell\ell'} \delta_{mm'} \tag{19.41}$$

Equation (19.39) is the main result. It shows that in the presence of vacuum birefringence a cross-correlation between E and B modes arises. A similar form is obtained for the TB cross-correlation, as well for changes in the TT and EE cross-correlations. At the lowest order in the rotation angle, there is no BB cross-correlation. This appears at second order in the rotation angle. In addition to these parity violating correlators, the covariance matrix is no longer diagonal in ℓ, m. These non-diagonal correlations are also useful information to fix a^{α}_{LM}. Assuming there is no other physics that gives rise to a BE and TE cross-correlation, using (19.39), and similar equations for C^{TT}_{ℓ}, C^{TE}_{ℓ}, C^{EE}_{ℓ} and C^{TB}_{ℓ} it is possible to invert and extract the a^{α}_{LM} from the data.

It is possible to find unbiased estimators for $C^{\alpha\alpha}_{\ell}$ in this manner[114] and a Fisher matrix analysis can be performed similarly to what we illustrated in section 16.3. The experimental noise can be found to be

$$C^{\alpha\alpha,\text{noise}}_{L} = \left[\sum_{\ell,\ell'} \frac{(2\ell+1)(\ell^{'}+1)(F^{L,BE}_{\ell\ell'})^2}{4\pi C^{BB,\text{map}}_{\ell} C^{EE,\text{map}}_{\ell}} \right]^{-1} \tag{19.42}$$

with

$$F^{L,BE}_{\ell\ell'} = 2C^{EE}_{\ell\ell'} \begin{pmatrix} \ell & L & \ell^{'} \\ 2 & 0 & -1 \end{pmatrix} W_{\ell} W_{\ell'} \tag{19.43}$$

and W_{ℓ} is the window function of the experiment (defined in (16.26) and (16.35) for a Gaussian beam). The $C^{XX,\text{map}}_{\ell}$ quantities are those given by (16.57)

$$C^{XY,\text{map}}_{\ell} = C^{XY}_{\ell} |W_{\ell}|^2 + C^{XY,\text{noise}}_{\ell} \tag{19.44}$$

The variance of the estimators $\hat{C}^{\alpha\alpha}_{L}$ are given by

$$(\Delta \hat{C}^{\alpha\alpha}_{L})^2 \simeq \frac{2}{f_{sky}(2L+1)} (C^{\alpha\alpha,\text{noise}}_{L})^2 \tag{19.45}$$

And a similar variance for the αT cross-correlation is approximately[38]

$$(\Delta \hat{C}^{\alpha T}_{L})^2 \simeq \frac{2}{f_{sky}(2L+1)} (C^{\alpha\alpha,\text{noise}}_{L})^2 \frac{C^{TT,\text{map}}_{L}}{|W_L|^2} \tag{19.46}$$

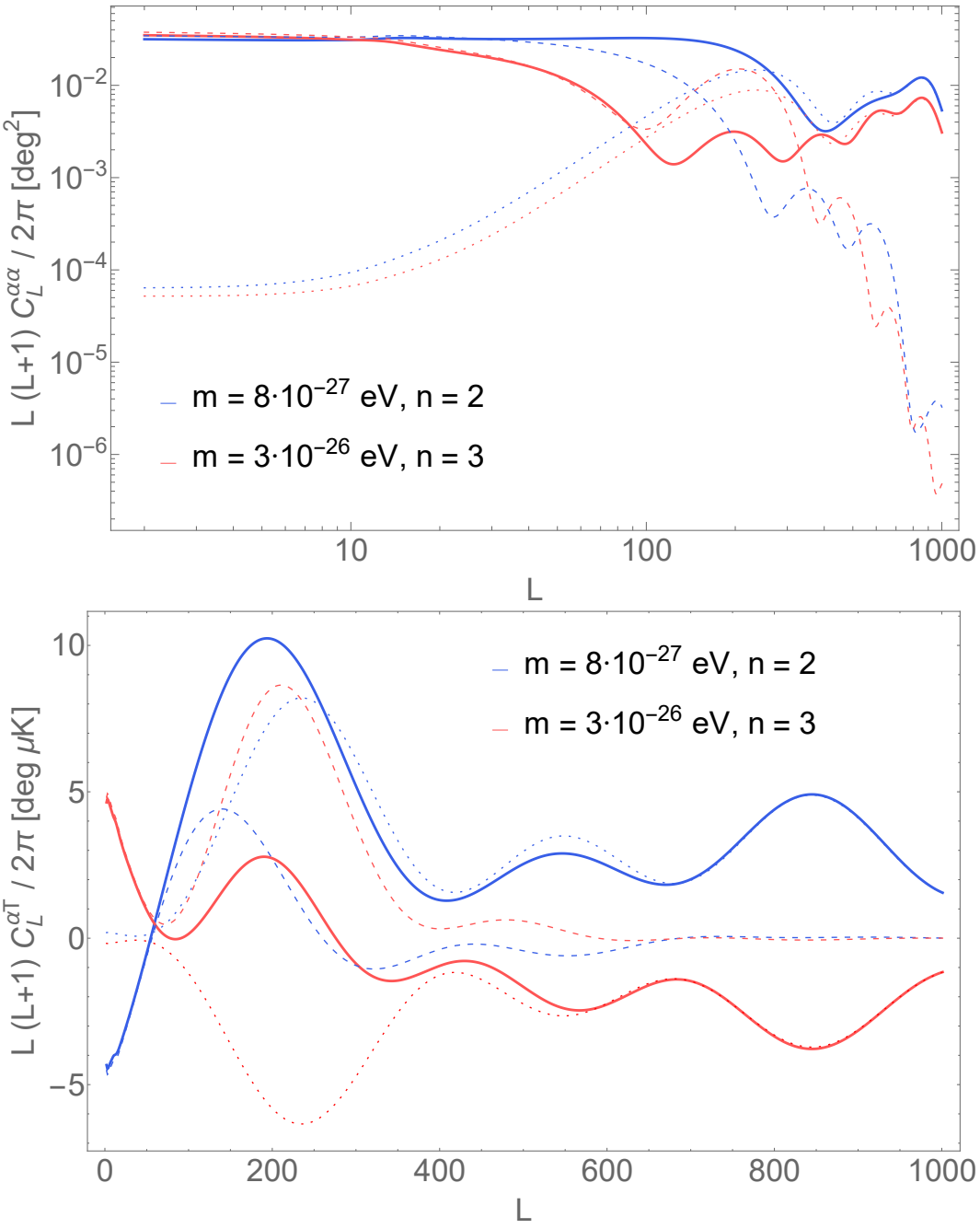

Figure 19.1: Rotation spectra $\frac{L(L+1)}{2\pi}C_L^{\alpha\alpha}$ and cross-correlation $\frac{L(L+1)}{2\pi}C_L^{\alpha T}$ for models indicated in table A.2, separating out the different contributions. The dashed lines indicate contributions of scalar field fluctuations only. The dotted lines are contributions from the density fluctuations. We used $G_{\chi\gamma\gamma} = 10^{-15} GeV^{-1}$ to normalize the spectra and parameters given by table 19.1.

n	$\chi_i[m_{Pl}]$	$m[eV]$	$\log_{10} a_c$	$f_{ede}(a_c)$	$H_0[km\ s^{-1}\ Mpc^{-1}]$
2	0.53	$8 \cdot 10^{-27}$	−3.7	0.04	70.0
3	0.58	$3 \cdot 10^{-26}$	−3.7	0.06	71.9

Table 19.1: Scalar field models under consideration. $m_{Pl} = \sqrt{\frac{1}{8\pi G}}$ is the reduced Planck mass.. These are the best fit model given in ref. [177].

19.3 Results for the rotation spectra and experimental forecasts

We compute the rotation power spectra and correlation with the temperature $C_L^{\alpha\alpha}$ and $C_L^{\alpha T}$. We consider two models with potentials given by

$$V(\chi) = \Lambda^4(1 - \cos\frac{\chi}{f})^n \tag{19.47}$$

where we fix $f = M_{Pl} = \sqrt{\frac{1}{G^2}} = 1.221 \cdot 10^{19} GeV$. We define the "mass" of the scalar field as

$$m = \frac{\Lambda^2}{f} \tag{19.48}$$

but take care to note that this is not a mass for $n \geq 2$. The expansion around the minimum of the potential is given by

$$V(\chi) \simeq \Lambda^4(\frac{\chi^2}{2f^2})^n = \frac{m^2}{2}\frac{\chi^{2n}}{f^{2n-2}} \tag{19.49}$$

The mass term is useful to compare across works in the literature. The two models under consideration are given in table 19.1. For each n, the values of the mass m and the initial value $\bar{\chi}_i$ of the background field are chosen so that the critical scale factor a_c and f_{ede} are those that best fit the cosmological data and reduce the Hubble tension[177]. They are compatible with (18.125)-(18.126) (which are confidence intervals). Adding early dark energy increases the value of H_0 inferred and this new value is used in the calculation.

The results for the total rotation spectra $C_L^{\alpha\alpha}$ and cross-correlation with temperature $C_L^{\alpha T}$ are shown in figure 19.1. In either case, the spectra shows the typical acoustic oscillations present in the CMB spectrum. The peaks and troughs being at similar positions as the temperature power spectrum C_L^{TT}. We have used $G_{\chi\gamma\gamma} = 10^{-15} GeV^{-1}$ as the coupling. The $C_L^{\alpha\alpha} \propto G_{\chi\gamma\gamma}^2$ while $C_L^{\alpha T} \propto G_{\chi\gamma\gamma}$.

As given by (19.26), the source of the spectra is given by the sum of two terms. The first term

$$S_\alpha^\chi(\tau, \vec{k}) = -\frac{1}{2}g(\tau)G_{\chi\gamma\gamma}\delta\chi \tag{19.50}$$

is due entirely to the fluctuations of the scalar field. Whereas the second term

$$S_\alpha^m(\tau, \vec{k}) = -g(\tau)(\bar{\alpha} + \frac{1}{2}G_{\chi\gamma\gamma}\bar{\chi})(2\psi + \delta_b) \tag{19.51}$$

is due to fluctuations of the matter content and the metric. We repeat the calculation with each term separately, plotted in figure 19.1.

Let's discuss the rotation spectra. It may seem that the overall sign, at small L, in the αT cross-correlation may be the defining difference between the two models. However this is misleading, as it depends on the sign of the initial value of the background field $\bar{\chi}_i$. If we choose the opposite initial sign, this spectra flips while all other observables, except the uniform rotation angle, remain unaltered. We have at the moment no way to choose this sign a-priori. The difference in the two models is actually encoded the change with L. In the $n = 2$ model the $C_L^{\alpha T}$ spectra changes sign from $L = 2$ to the first peak at $L \sim 200$. On

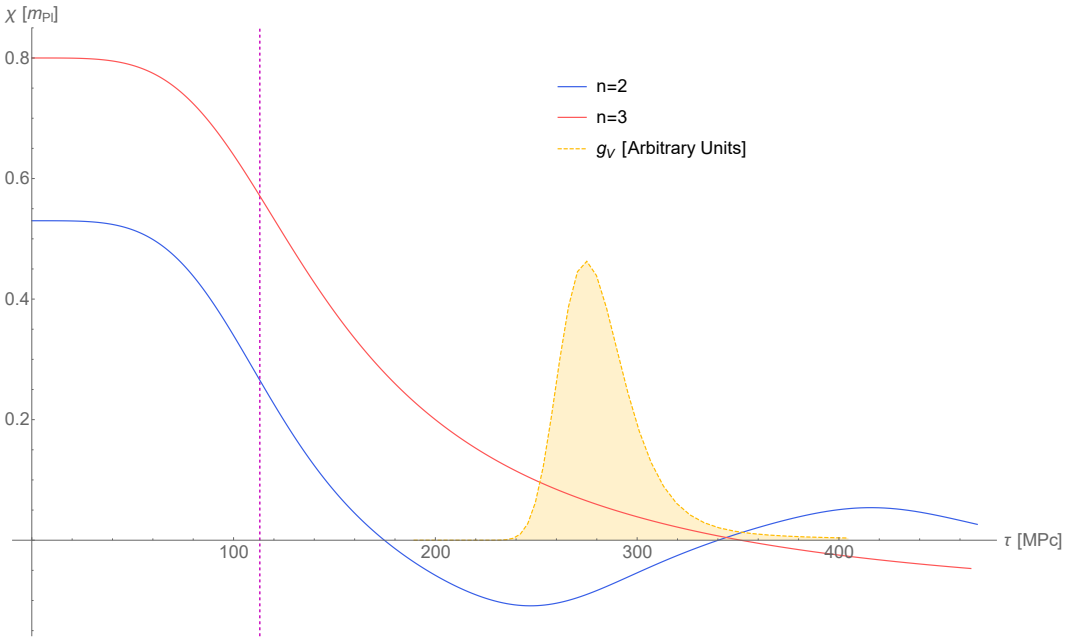

Figure 19.2: Evolution of the background field $\bar{\chi}$, in units of the reduced Planck mass m_{Pl}. The vertical dashed purple line is the moment of matter radiation equality. The fields begin oscillating at the same time, but their subsequent evolution depends on the form of the potential and so they contribute with opposite signs at decoupling. The model parameters are given in table 19.1. . Other used parameters are the best fit ΛCDM parameters[60].

the other hand, the $n = 3$ model does not change signs. Starting from $L \sim 20$ there is also a difference in magnitude between the two spectra. The reason can be gleaned from the $C_L^{\alpha T}$ spectra shown in figure 19.1. In the case of $n = 2$, the two sources give the same sign in the correlation spectra, and the amplitude is increased, whereas, in the $n = 3$ case, the two terms are out of phase and the amplitude is decreased.

To understand the origin of these differences we plot the evolution of the background scalar field $\bar{\chi}$ in both models, in figure 19.2, as well the evolution of the perturbation for several values of k, in figure 19.3. The scalar field does not affect the matter perturbations too much, as evidenced by the fact that the CMB power spectrum is nearly unaffected by the presence of an early scalar field. To understand the difference in rotation spectra for the two models, we analyze the field itself.

Low values of $k \sim 10^{-4} Mpc^{-1}$ contribute to low values of $L \sim k(\tau - \tau_{LSS})$. As can be seen in figure 19.3 (upper panel), for small times τ, there is a growing mode (18.116). While the background field is slowly rolling, the evolution is largely insensitive to the potential, and we expect similar contributions. As soon as the oscillations begin, the damping ratio (18.121) becomes relevant. It can be shown that in the $n = 2$ case $\varsigma < 1$ at the relevant times, while $\varsigma > 1$ for the $n = 3$, since when $\bar{\chi}$ moves toward the minimum of the potential, it is much flatter in this case, raising the damping ratio. When oscillations begin, the $n = 2$ perturbations begin to oscillate and become positive by recombination. On the other hand the $n = 3$ perturbations are damped and tend towards zero. In particular the sign does not change. This explains the difference in sign at low L in the $C_L^{\alpha T}$.

For larger values of $k \sim 10^{-2} Mpc^{-1}$ which contribute to $L \sim 10^2$, the value of k is becom-

Experiment	Beam θ	Power noise $T^{-1}w^{-\frac{1}{2}}[\mu K - \text{arcmin}]$	f_{sky}
LiteBIRD[98, 90]	$30'$	4.5	0.7
S3deep[2]	$1'$	4	0.06
S3wide[2]	$1.4'$	8	0.4
Simons Obs. SAT[94]	$17'$	2	0.1
Simons Obs. LAT[94]	$1.4'$	6	0.4
CMB-S4[53]	$3'$	1	0.4

Table 19.2: Specifications for different experimental configurations configured. For the polarization spectra the noise w^{-1} is multiplied by two. The beam is defined in (16.33), with θ being the Full Width at Half Maximum (16.36). The power noise is defined in (16.56). f_{sky} is the sky fraction observed.

ing larger the second derivative of the potential, which is very small in both cases, as $\bar{\chi}$ is moving towards the minimum at $\bar{\chi} = 0$ (figure 19.2). First of all, this implies that the differences between the model are less pronounced as the potential becomes a subdominant term to $k^2\delta\chi$. Furthermore, in this limit $\varsigma(\tau) < 1$ and we expect that both models oscillate. As can be seen in figure 19.3 (lower panel), this is the case. The spectra however is the sum of a term containing the perturbation, with one containing the background field. At recombination the background field has differing in signs in the two models, but the same sign in the perturbations at the same k. This implies that in one model, namely the $n = 2$ model , the two terms will sum constructively while in the other they sum destructively.

To conclude, we forecast the ability of future experiments to detect the birefringence signal using a Fisher matrix analysis. The relevant experimental parameters are in table 19.2.

Using the variance of the unbiased estimators for $C_L^{\alpha\alpha}$ (19.45) and $C_L^{\alpha T}$ (19.46) we calculate the minimum value of $G_{\chi\gamma\gamma}$ necessary in order for the signal-to-noise ration S/N to be larger than 3. This is given by[40]

$$(\frac{S}{N})_{\alpha\alpha} = \left(\sum_L (\frac{C_L^{\alpha\alpha,\text{fiducial}}}{\Delta\hat{C}_L^{\alpha\alpha}})^2\right)^{\frac{1}{2}} \tag{19.52}$$

$$(\frac{S}{N})_{\alpha T} = \left(\sum_L (\frac{C_L^{\alpha T,\text{fiducial}}}{\Delta\hat{C}_L^{\alpha T}})^2\right)^{\frac{1}{2}} \tag{19.53}$$

The minimum values that gives the $S/N > 3$ ratios are given in table 19.3. The αT cross-correlation can probe smaller values of the coupling. This is not surprising as the cross-correlation scales more slowly with $G_{\chi\gamma\gamma}$ then the self-correlation. We also note that the experimental reach is weakly dependent on the form of the potential.

We noted the spectra for the two potentials are quite different. Indeed, the rotation spectra may be a powerful tool to probe the fundamental parameters of the scalar field, should such a signal be discovered. For this reason, we also compute the minimum value of the coupling $G_{\chi\gamma\gamma}$ such that the different spectra may be discriminated. The results are tabulated in 19.4.

Should the Hubble tension persist in the data, the rotation spectra could be the crucial observable to determine whether an explanation through early dark energy is possible. If it is, it then becomes a remarkable tool to probe the physical parameter space of the cosmic scalar field χ.

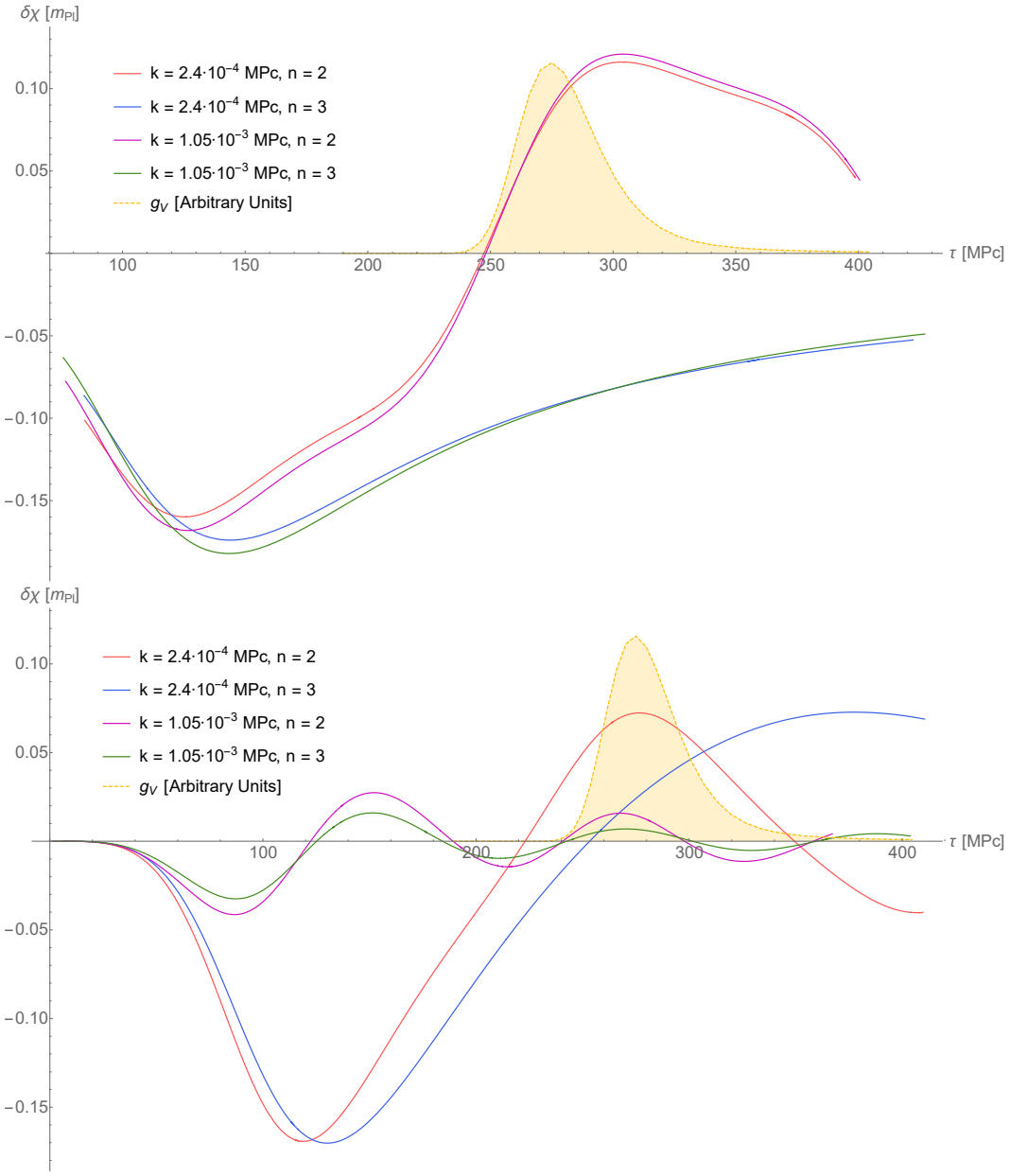

Figure 19.3: Evolution of longer wavelength modes (upper figure) and shorter ones (lower figure). At small values of k, for smaller n the solution presents oscillations while the $n = 3$ is a damped solution, causing different relative contributions at recombination. For shorter wavelengths the k term dominates over the potential term in (18.103) and so the evolution in the two models is similar. The calculation was done using a modifed version of the CLASS Boltzmann code, using best fit ΛCDM parameters and the addition of those given in table A.2. The conformal-Newtonian gauge was used.

Experiment	$G_{\chi\gamma\gamma}, C_L^{\alpha\alpha}, n=2$	$G_{\chi\gamma\gamma}, C_L^{\alpha T}, n=2$	$G_{\chi\gamma\gamma}, C_L^{\alpha\alpha}, n=3$	$G_{\chi\gamma\gamma}, C_L^{\alpha T}, n=3$
LiteBIRD[98, 90]	0.145	0.078	0.137	0.087
S3wide[2]	0.11	0.068	0.11	0.077
S3deep[2]	0.0345	0.034	0.033	0.038
Simons Obs. SAT[94]	0.031	0.027	0.03	0.029
Simons Obs. LAT[94]	0.085	0.052	0.08	0.058
CMB-S4[53]	0.014	0.0088	0.0135	0.0096

Table 19.3: Forecasts for the minimal value of the coupling $G_{\chi\gamma\gamma}$, in units of $10^{-15}GeV^{-1}$, such that future experiments could detect a cosmic birefringence signal due to early dark energy, with a signal-to-noise ratio of $(S/N) = 3$, using either the $C_L^{\alpha\alpha}$ or $C_L^{\alpha T}$ spectra.

Experiment	$G_{\chi\gamma\gamma}, C_L^{\alpha\alpha}$	$G_{\chi\gamma\gamma}, C_L^{\alpha T}$
LiteBIRD[98, 90]	0.44	0.042
S3wide[2]	0.34	0.038
S3deep[2]	0.205	0.019
Simons Obs. SAT[94]	0.095	0.015
Simons Obs. LAT[94]	0.25	0.029
CMB-S4[53]	0.043	0.005

Table 19.4: Forecasts for the minimal value of the coupling $G_{\chi\gamma\gamma}$, in units of $10^{-15}GeV^{-1}$, such that future experiments could distinguish between potentials with $n = 2$ and $n = 3$ with a signal-to-noise ratio of $(S/N) = 3$, using either the $C_L^{\alpha\alpha}$ or the $C_L^{\alpha T}$ spectra.

Part VI

Conclusions

It is a lengthy journey to make inferences from the CMB, but certainly a satisfying one. Describing the current state of CMB anisotropies requires putting together ideas from general relativity, quantum mechanics and statistical mechanics. It is certainly a culmination of the last 120 years of physics.

We began by considering the Einstein equations in a universe governed by the Copernican principle. Using the Friedmann equations and equilibrium thermodynamics, we are able to describe the evolution of the average, homogeneous universe, from the earliest times until today. This evolution is the adiabatic, or isentropic one. Through this, it was realized that the largest part of the universe is composed of two quantities which are not described in the standard model, dark energy and dark matter. This sets the stage for using cosmology to understand the nature of fundamental physics.

Doing so is easier said than done. If we wish to analyze cosmological observables, and the CMB is the prime example, we must understand the evolutions of the small perturbations. We have dedicated a large part of this work in deriving these equations *ab initio* in a consistent manner. We have done so in a pedagogical way, hoping to make the introduction to the subject easy for any student who is interested. In doing so, we have found that oscillations of the primordial plasma leave a very distinct imprint on the CMB anisotropies. It is these anisotropies that we have probed.

Unraveling them is also a big challenge. Many parameters affect the final result in not obvious ways. Numerical and statistical methods must be employed and care must be taken in not forgetting to include any physical effect. Indeed, this was one of the main results of the work: the practical impact of the theoretical assumptions one makes when analyzing the data (sec. 17.3). We found that neglecting the running of the spectral index can cause a bias in the analysis of the data, which would mimic the experimental signature of a light relic (sec. 17.3.1). Furthermore, unless the lifetime of the neutron is measured with more precision, the Big Bang Nucleosynthesis Helium fraction cannot be fixed and this introduces yet more error in the determination of the presence of a light relic (sec. 17.3.2).

We then turned to the physics of cosmic scalar fields, which may contribute to the dark energy phenomenon. We developed a new formalism for dealing with rotation of the polarization of the cosmic microwave background (sec. 19.2) and applied this to the study of an early dark energy which might solve the Hubble tension. We found how the rotation power spectrum is realistically measurable by next generation experiments and with non-trivial couplings to the photon. In addition, we showed how the rotation spectra can be used to gain a lot of insight on the the self-interactions of the early dark energy, through its potential. Finally, we demonstrated the possibility of a new phenomenon not previously noticed: the presence of acoustic oscillation signals in the rotation spectrum (sec. 19.3).

A Conventions and useful quantities

We use a metric signature

$$(-,+,+,+) \tag{A.1}$$

Corresponding to a flat Minkowski metric

$$\eta_{\mu\nu} = diag(-1,1,1,1) \tag{A.2}$$

Negative distances on a spacetime are *timelike.* Positive distances are *spacelike.* Greek letters (μ,ν,α,ρ,...) are used to indicate indexes which run from 0 to $d-1$, d being the dimension of the manifold. Latin letters (i,j,k,...) are used to indicate indexes which run on spacelike indexes.

The Levi-Civita connection[159], in a torsion free manifold[27], is implemented by the Christoffel symbols

$$\Gamma^{\rho}_{\mu\nu} = \frac{1}{2} g^{\rho\lambda}(g_{\mu\lambda,\nu} + g_{\nu\lambda,\mu} - g_{\mu\nu,\lambda}) \tag{A.3}$$

The Riemann tensor is defined as

$$R^{\rho}_{\sigma\mu\nu} = \Gamma^{\rho}_{\sigma\nu,\mu} - \Gamma^{\rho}_{\sigma\mu,\nu} + \Gamma^{\rho}_{\mu\lambda}\Gamma^{\lambda}_{\nu\sigma} - \Gamma^{\rho}_{\nu\lambda}\Gamma^{\lambda}_{\mu\sigma} \tag{A.4}$$

The Ricci tensor is

$$R_{\mu\nu} \equiv R^{\lambda}_{\mu\lambda\nu} \tag{A.5}$$

For the conformal-Newtonian and synchronous gauges we use the same definitions as given in [155]. For tensor modes we used the definitions given in [80].

The Stoke parameters on the celestial sphere are defined through a right handed basis around the incoming photon direction. Table A.1 indicates formulas and definitions in the text.

Natural units $c = \hbar = k_b = 1$ are used throughout, $1 = 198\text{MeV} \cdot \text{fm} = 198\text{eV} \cdot \text{nm}$. Main conversion are given in table A.2. We use eV and eV^{-1} as base units. In these units, specific values of importance to cosmology are

$$T_{CMB} = 2.7K = 2.328 \cdot 10^{-4} eV \tag{A.6}$$

$$H_0 = h \cdot 100 \frac{km}{s \cdot Mpc} = h \cdot 2.139 \cdot 10^{-33} eV \tag{A.7}$$

$$\rho_{\text{crit}} = h^2 \cdot 8.140 \cdot 10^{-11} eV^4 \tag{A.8}$$

$$G = \frac{1}{M_P^2} = 6.708 \cdot 10^{-39} GeV \tag{A.9}$$

where $M_P = 1.221 \cdot 10^{19} GeV$ is the Planck mass.

Symbol	Definition	Equation
a	Dimensionless scale factor	(6.22)
w	Equation of state parameter	(7.23)
f_i	Phase space density of species i	(8.1)
$\tilde{g}$	Relativistic d.o.f. of total density	(8.11)
s	Entropy density	(8.30)
g^*	Entropy d.o.f	(8.37)
$(\frac{\partial f}{\partial t})_C$	Collision term	(9.7)
X_A	Mass fraction of species A	(12.7)
σ_T	Thompson cross section	(A.10)
$\Phi_{H,A}$	Gauge invariant metric terms	(13.20), (13.21)
ψ, ϕ	Conformal-Newtonian metric	(13.24)
h, η	Synchronous gauge, Fourier space	(13.26), (13.27), (13.28), (13.35)
$h_{+,\times}$	Tensor metric	(13.62), (13.64)
R	Photon to Baryon ratio	(13.284)
P_ψ	Power spectrum definition	(13.334)
$\Theta^{(m)}_{\gamma,\ell}$	Photon temperature perturbations	(14.1)
$a^{E,B}_{\ell m}$	E and B modes on the sky	(14.18)
W_ℓ	Window function	(16.26)
k_D^{-1}	Damping scale	(17.28)
$G_{\chi\gamma\gamma}$	ALP-Photon coupling	(18.74)
δ, θ, σ	Matter perturbations	(13.76), (13.77), (13.78)
$F_{\nu,\gamma}$	Intensity perturbations	(13.155), (13.232)
$F^{(m)}_{\nu,\gamma,\ell}$	Angular expansions	(13.163), (13.239)
$E^{(m)}_\ell, B^{(m)}_\ell$	E and B photon perturbations	(13.240)
$\Pi^{(m)}$	Polarization source	(13.249)
τ_c	Optical depth	(13.280)

Table A.1: Definitions and formulas used throughout the text

$1nm = 5.05 \cdot 10^{-3} eV^{-1}$
$1Mpc = 1.5586 \cdot 10^{29} eV^{-1}$
$1K = 8.6207 \cdot 10^{-5} eV$
$1kg = 5.610 \cdot 10^{35} eV$
$1\frac{kg}{m^3} = 4.354 \cdot 10^{15} eV^4$
$1Gyear = 4.778 \cdot 10^{31} eV^{-1}$

Table A.2: Conversions between units

The Thompson cross section is given, in natural units, by

$$\sigma_T = \frac{8\pi}{3}\frac{\alpha^2}{m_e^2} = 6.652 \cdot 10^{-25} cm^2 \tag{A.10}$$

B Spin-Weighted Spherical Harmonics and Legendre Polynomials

On any given point on a sphere one can define three orthogonal vectors, two of which are chosen to lie in the tangent space of the sphere. The first is the outward pointing radial vector $\hat{n}$, the others are $\hat{e}_1$ and $\hat{e}_2$. The latter are defined up to a rotation around $\hat{n}$. A function ${}_sf(\theta, \phi)$ is said to be spin-s if under a right handed rotation of angle ψ around $\hat{n}$

it changes by ${}_sf(\theta,\phi)' = e^{-is\psi}\,{}_sf(\theta,\phi)$. An example of spin 1 function is, for some fixed vector $\vec{a}$:

$$\vec{a}\cdot(\hat{e}_1 + i\hat{e}_2) \tag{B.1}$$

In the sense that $\vec{a}\cdot(\hat{f}_1 + i\hat{f}_2) = e^{-i\psi}\vec{a}(\hat{e}_1 + i\hat{e}_2)$ where the basis $\{\hat{f}_i\}$ is the one obtained by a right handed rotation from the basis $\{\hat{e}_i\}$.

A spin-0 function is, for example,

$$\vec{a}\cdot\hat{n} \tag{B.2}$$

and a spin -1 function is

$$\vec{a}\cdot(\hat{e}_1 - i\hat{e}_2) \tag{B.3}$$

We now describe a procedure to raise or lower the spin. This method works on any smooth two-dimensional manifold, however we shall refer explicitly to the 2-sphere with the metric $ds^2 = d\theta^2 + \sin^2\theta d\phi^2$. On the tangent space at some point (θ,ϕ) it is natural to use the orthonormal basis $\{\hat{e}_\theta = (1,0), \hat{e}_\phi = (0, \frac{1}{\sin\theta})\}$. We define two complex vectors[128]

$$\vec{m} = \frac{1}{\sqrt{2}}(\hat{e}_\theta + i\hat{e}_\phi) \tag{B.4}$$

$$\vec{m}^* = \frac{1}{\sqrt{2}}(\hat{e}_\theta - i\hat{e}_\phi) \tag{B.5}$$

Under a right-handed rotation around the outgoing radial direction, $\vec{m}$ is spin-1 and $\vec{m}^*$ is spin-(-1). These vectors satisfy $\vec{m}\cdot\vec{m} = \vec{m}^*\cdot\vec{m}^* = 0$ and $\vec{m}\cdot\vec{m}^* = 1$. Given a spin-$s$ function ${}_sf$ we construct the spin-zero tensor

$$f_{i_1\ldots i_s} =_s f m^*_{i_1}\ldots m^*_{i_s} \tag{B.6}$$

if $s > 0$, or

$$f_{i_1\ldots i_{-s}} =_s f m_{i_1}\ldots m_{i_s} \tag{B.7}$$

if $s < 0$. Then a quantity with spin $s+1$ is

$${}_{(s+1)}f = f_{i_1\ldots i_s;j} m^{i_1 *}\ldots m^{i_s *} m^{j*} \tag{B.8}$$

where the ${}_{;}j$ indicates the covariant derivative on the sphere, which is expressed through the relevant Christoffel symbols. If $s < 0$ the terms $m^{i_1*}\ldots m^{i_s*}$ above are replaced with their complex conjugates. If we wish to lower the spin then we replace m^{j*} with m^j.

Up to an arbitrary constant, this procedure can be expressed through the raising and lowering operators for spin ð(eth) and $\bar{\text{ð}}$ (eth-bar)[115, 87]:

$$\text{ð}_s f(\theta,\phi) = -\sin^s\theta[\frac{\partial}{\partial\theta} + \frac{i}{\sin\theta}\frac{\partial}{\partial\phi}]\sin^{-s}\theta\,{}_sf(\theta,\phi) \tag{B.9}$$

$$\bar{\text{ð}}_s f(\theta,\phi) = -\sin^{-s}\theta[\frac{\partial}{\partial\theta} - \frac{i}{\sin\theta}\frac{\partial}{\partial\phi}]\sin^{s}\theta\,{}_sf(\theta,\phi) \tag{B.10}$$

where $(\text{ð}_s f)' = e^{-i(s+1)\psi}\,{}_sf$ and $(\bar{\text{ð}}_s f)' = e^{-i(s-1)\psi}\,{}_sf$.

Useful formulas when dealing with polarizations, which are spin 2 and -2 functions such that $\partial f/\partial\phi = imf$ and writing $\mu = \cos\theta$

$$\bar{\eth}^2\,{}_2f = (-\frac{\partial}{\partial\mu} + \frac{m}{1-\mu^2})^2((1-\mu^2)_2f) \tag{B.11}$$

$$\eth^2\,{}_{-2}f = (-\frac{\partial}{\partial\mu} - \frac{m}{1-\mu^2})^2((1-\mu^2)_{-2}f) \tag{B.12}$$

Where the repeated application of $\eth$ and $\bar{\eth}$ must keep in mind the changing value of the spin s on each application.

We can define the *spin-weighted spherical harmonics* ${}_sY_\ell^m$ in terms of the regular spherical harmonics by raising or lowering the spin[87, 128]. For $s > 0$

$${}_sY_\ell^m(\theta,\phi) \propto [\frac{(\ell-s)!}{(\ell+s)!}]^{\frac{1}{2}}\bar{\eth}^s Y_\ell^m \; 0 \le s \le \ell \tag{B.13}$$

While for $s < 0$

$${}_sY_\ell^m(\theta,\phi) \propto [\frac{(\ell+s)!}{(\ell-s)!}]^{\frac{1}{2}}\bar{\eth}^s Y_\ell^m \; -\ell \le s \le 0 \tag{B.14}$$

Note that for $\ell < |s|$ the ${}_sY_\ell^m = 0$ in each case. The overall factor makes the the spherical harmonics orthonormal. There exist various phase conventions in the literature, which is why we specified the proportionality. We shall use the convention as in [128]. Then the harmonics are explicitly given by

$$\begin{aligned} {}_sY_\ell^m(\theta,\phi) \;=\; & \left[\frac{2\ell+1}{4\pi}\frac{(\ell+m)!}{(\ell+s)!}\frac{(\ell-m)!}{(\ell-s)!}\right]^{\frac{1}{2}} \sin^{2\ell}\frac{\theta}{2}\sum_r \binom{\ell-s}{r}\binom{\ell+s}{r+s-m} \\ & (-1)^{\ell-r-s}e^{im\phi}\left(\cot\frac{\theta}{2}\right)^{2r+s-m} \end{aligned} \tag{B.15}$$

They can be related to the Wigner D-matrix[216], which links them to the rotation group

$$D^\ell_{-m,s}(\theta,\phi,\psi) = \sqrt{\frac{4\pi}{2\ell+1}}\,{}_sY_\ell^m(\theta,\phi)e^{-is\psi} \tag{B.16}$$

The Euler angles (θ,ϕ,ψ) are defined in the $z-y-z$ convention[116] often used in quantum mechanics[185].

The spin-weighted spherical harmonics are a complete and orthonormal set of spin-s functions

$$\int d\Omega\,{}_sY_\ell^{m*}\,{}_sY_{\ell'}^{m'} = \delta_{\ell,\ell'}\delta_{m,m'} \tag{B.17}$$

$$\sum_{\ell m}\,{}_sY_\ell^{m*}(\theta,\phi)\,{}_sY_\ell^m(\theta',\phi') = \delta(\cos\theta-\cos\theta')\delta(\phi-\phi') \tag{B.18}$$

Under a parity transformation $\theta \to \pi-\theta$, $\phi \to \phi+\pi$

$${}_sY_\ell^m \to (-1)^\ell\,{}_{-s}Y_\ell^m \tag{B.19}$$

Under complex conjugation

$$ {}_sY_\ell^{m*} = (-1)^{s+m}\ {}_{-s}Y_\ell^{-m} \tag{B.20} $$

Just like regular spherical harmonics, the outer product of two harmonics can be decomposed with the Clebsch-Gordan coefficients

$$ \begin{aligned} {}_{s1}Y_{\ell_1}^{m_1}\ {}_{s_2}Y_{\ell_2}^{m_2} &= \frac{\sqrt{(2\ell_1+1)(2\ell_2+1)}}{4\pi}\sum_{\ell,m,s} < \ell_1, m_1; \ell_2, m_2 | \ell_1, \ell_2, \ell, m > \\ &< \ell_1, -s_1; \ell_2, -s_2 | \ell_1, \ell_2, \ell, -s > \sqrt{\frac{4\pi}{2\ell+1}}\ {}_sY_\ell^m \end{aligned} $$

With the sum extending to $|\ell_1 - \ell_2| \leq \ell \leq |\ell_1 + \ell_2|$, $|m| \leq \ell$ and $|s| \leq \ell$. This relationship is easily proven by passing through the Wigner D-matrix. The Clebsch-Gordan coefficients are related to the Wigner-3j symbols. So alternatively to the above, we have the generalized Gaunt integral

$$ \begin{aligned} \int d\Omega\ {}_SY_L^{M*}(\hat{n})\,{}_{s1}Y_{\ell 1}^{m_1}(\hat{n})\,{}_{s_2}Y_{\ell_2}^{m_2}(\hat{n}) &= (-1)^{S+M}\sqrt{\frac{(2L+1)(2\ell_1+1)(2\ell_2+1)}{4\pi}} \times \\ & \begin{pmatrix} L & \ell_1 & \ell_2 \\ -M & m_1 & m_2 \end{pmatrix} \begin{pmatrix} L & \ell_1 & \ell_2 \\ S & -s_1 & -s_2 \end{pmatrix} \end{aligned} $$

It can be shown they satisfy a generalized addition relation[128]

$$ \sum_m\ {}_{s_1}Y_\ell^{m*}(\theta', \phi')\ {}_{s_2}Y_\ell^{m*}(\theta, \phi) = \sqrt{\frac{2\ell+1}{4\pi}}\ {}_{s_2}Y_\ell^{-s_1}(\beta, \alpha)e^{-is_2\gamma} \tag{B.21} $$

where α, β, γ are the Euler angles which rotate (θ', ϕ') into (θ, ϕ) using the $z-y-z$ convention. In the simple case of $s_1 = s_2 = 0$

$$ \sum_m Y_\ell^m(\hat{n})Y_\ell^{m*}(\hat{n}') = \frac{2\ell+1}{4\pi}P_\ell(\hat{n}\cdot\hat{n}') \tag{B.22} $$

Legendre Polynomials

Full discussion of Legendre polynomials can be found in ref. [85]. We discuss properties useful in the text.

The Legendre Polynomials are an infinite set of real polynomials $P_n(x)$ which on the closed interval $[-1, 1]$ satisfy the following orthogonality property:

$$ \int_{-1}^{1} P_n(x)P_m(x)\frac{dx}{2} = \frac{1}{2n+1}\delta_{nm} \tag{B.23} $$

They also are a complete set in the closed interval $[-1, 1]$ so that any function $f(x)$ defined in this interval can be expanded in this basis:

$$ f(x) = \sum_{\ell=0}^{\infty}(-i)^\ell(2\ell+1)f_\ell P_\ell(x) \tag{B.24} $$

where $(-i)^\ell$ is an arbitrary addition, which simplifies cosmological formula, and $(2\ell+1)$ is included to get a simple normalization. In this manner

$$f_\ell = \frac{1}{(-i)^\ell} \int_{-1}^{1} \frac{dx}{2} P_\ell(x) f(x) \tag{B.25}$$

For a plane wave $f(x) = e^{-ikx}$ the $f_\ell = j_\ell(k)$, the spherical Bessel functions. The first few Legendre polynomials are

$$\begin{aligned} P_0(x) &= 1 \\ P_1(x) &= x \\ P_2(x) &= \frac{1}{2}(3x^2 - 1) \\ P_3(x) &= \frac{1}{2}(5x^3 - 3x) \\ P_4(x) &= \frac{1}{8}(35x^4 - 30x^2 + 3) \end{aligned}$$

Conversely the first few powers of x can be expanded into Legendre Polynomials

$$\begin{aligned} x^2 &= \frac{1}{3}(P_0(x) + 2P_2(x)) \\ x^3 &= \frac{1}{5}(3P_1(x) + 2P_3(x)) \\ x^4 &= \frac{1}{35}(7P_0(x) + 20P_2(x) + 8P_4(x)) \end{aligned}$$

They satisfy a recurrence relation

$$(\ell+1)P_{\ell+1}(x) - (2\ell+1)xP_\ell(x) + \ell P_{\ell-1}(x) = 0 \tag{B.26}$$

And can be written in terms of sum of products of spherical harmonics

$$P_\ell(\hat{\alpha}\cdot\hat{\beta}) = \frac{4\pi}{2\ell+1} \sum_{m=-\ell}^{\ell} Y_{\ell m}(\hat{\alpha}) Y^*_{\ell m}(\hat{\beta}) \tag{B.27}$$

In the limit of large ℓ the Legendre polynomials tend to the Bessel function of order zero

$$\lim_{\ell\to+\infty} P_\ell(\cos\theta) = \sqrt{\frac{\theta}{\sin\theta}} J_0((\ell+\frac{1}{2})\theta) \tag{B.28}$$

References

[1] B. Allgood, R. A. Flores, J. R. Primack, A. V. Kravtsov, R. H. Wechsler, A. Faltenbacher, and J. S. Bullock. The shape of dark matter haloes: dependence on mass, redshift, radius and formation. *Monthly Notices of the Royal Astronomical Society*, 367(4):1781–1796, apr 2006.

[2] R. Allison, P. Caucal, E. Calabrese, J. Dunkley, and T. Louis. Towards a cosmological neutrino mass detection. *Physical Review D*, 92(12), dec 2015.

[3] Roger Apery. Irrationalite de $\zeta(2)$ et $\zeta(3)$. *Asterisque*, 61:11,13, 1979.

[4] T. M. Apostol. *NIST Handbook of Mathematical functions*. Cambridge University Press, 2010.

[5] Peter Arnold and Larry McLerran. Sphalerons, small fluctuations, and baryon-number violation in electroweak theory. *Physical Review D*, 36(2):581–595, jul 1987.

[6] Asimina Arvanitaki, Savas Dimopoulos, Sergei Dubovsky, Nemanja Kaloper, and John March-Russell. String axiverse. *Physical Review D*, 81(12), jun 2010.

[7] J. J. Aubert, U. Becker, P. J. Biggs, J. Burger, M. Chen, G. Everhart, P. Goldhagen, J. Leong, T. McCorriston, T. G. Rhoades, M. Rohde, Samuel C. C. Ting, Sau Lan Wu, and Y. Y. Lee. Experimental observation of a heavy ParticleJ. *Physical Review Letters*, 33(23):1404–1406, dec 1974.

[8] Benjamin Audren and Julien Lesgourgues. Non-linear matter power spectrum from time renormalisation group: efficient computation and comparison with one-loop. *Journal of Cosmology and Astroparticle Physics*, 2011(10):037–037, oct 2011.

[9] J. N. Bahcall, R. M. Soneira, and M. Schmidt. The galactic spheroid. *The Astrophysical Journal*, 265:730, feb 1983.

[10] Neta A. Bahcall and Xiaohui Fan. The most massive distant clusters: determining omega and sigma8. *The astrophysical journal*, 504(1), 1998.

[11] C. A. Baker, D. D. Doyle, P. Geltenbort, K. Green, M. G. D. van der Grinten, P. G. Harris, P. Iaydjiev, S. N. Ivanov, D. J. R. May, J. M. Pendlebury, J. D. Richardson, D. Shiers, and K. F. Smith. Improved experimental limit on the electric dipole moment of the neutron. *Physical Review Letters*, 97(13), sep 2006.

[12] Riccardo Barbieri, Paolo Creminelli, Alessandro Strumia, and Nikolaos Tetradis. Baryogenesis through leptogenesis. *Nuclear Physics B*, 575(1-2):61–77, may 2000.

[13] James M. Bardeen. Gauge-invariant cosmological perturbations. *Physical Review D*, 22(8):1882–1905, oct 1980.

[14] William A. Bardeen. Anomalous currents in gauge field theories. *Nuclear Physics B*, 75(2):246–258, jun 1974.

[15] N. Bartolo, S. Matarrese, and A. Riotto. Non-gaussianity and the cosmic microwave background anisotropies. *Advances in Astronomy*, 2010:1–68, 2010.

[16] Daniel Baumann. Tasi lectrues on inflation [arxiv:0907.5424 [hep-th]]. 2009.

[17] Daniel Baumann. Tasi lectures on primordial cosmology [arxiv:1807.03098 [hep-th]]. 2018.

[18] Daniel Baumann, Daniel Green, and Benjamin Wallisch. New target for cosmic axion searches. *Physical Review Letters*, 117(17), oct 2016.

[19] Daniel Baumann, Daniel Green, and Benjamin Wallisch. Searching for light relics with large-scale structure. *Journal of Cosmology and Astroparticle Physics*, 2018(08):029–029, aug 2018.

[20] K. G. Begeman, A. H. Broeils, and R. H. Sanders. Extended rotation curves of spiral galaxies: dark haloes and modified dynamics. *Monthly Notices of the Royal Astronomical Society*, 249(3):523–537, apr 1991.

[21] A.A. Belavin, A.M. Polyakov, A.S. Schwartz, and Yu.S. Tyupkin. Pseudoparticle solutions of the yang-mills equations. *Physics Letters B*, 59(1):85–87, oct 1975.

[22] José Luis Bernal, Licia Verde, and Adam G. Riess. The trouble with H_0. *Journal of Cosmology and Astroparticle Physics*, 2016(10):019–019, oct 2016.

[23] Gianfranco Bertone, editor. *Particle Dark Matter*. Cambridge University Press, 2009.

[24] Gianfranco Bertone, Dan Hooper, and Joseph Silk. Particle dark matter: evidence, candidates and constraints. *Physics Reports*, 405(5-6):279–390, jan 2005.

[25] James D. Bjorken and Sydney D. Drell. *Relativistic Quantum Fields*. McGraw-Hill College, 1965.

[26] Diego Blas, Julien Lesgourgues, and Thomas Tram. The cosmic linear anisotropy solving system (CLASS). part II: Approximation schemes. *Journal of Cosmology and Astroparticle Physics*, 2011(07):034–034, jul 2011.

[27] Matthias Blau. *Lecture notes on general relativity*.

[28] Ludwig Boltzmann. Further studies on the thermal equilibrium of gas molecules. In *History of Modern Physical Sciences*, pages 262–349. Imperial college press, jul 2003.

[29] R. Bowen, S. H. Hansen, A. Melchiorri, Joseph Silk, and R. Trotta. The impact of an extra background of relativistic particles on the cosmological parameters derived from the cosmic microwave background. *Monthly Notices of the Royal Astronomical Society*, 334(4):760–768, aug 2002.

[30] A. Boyarsky, M. Drewes, T. Lasserre, S. Mertens, and O. Ruchayskiy. Sterile neutrino dark matter. *Progress in Particle and Nuclear Physics*, 104:1–45, jan 2019.

[31] M. Breidenbach, J. I. Friedman, H. W. Kendall, E. D. Bloom, D. H. Coward, H. DeStaebler, J. Drees, L. W. Mo, and R. E. Taylor. Observed behavior of highly inelastic electron-proton scattering. *Physical Review Letters*, 23(16):935–939, oct 1969.

[32] U. G. Briel, J. P. Henry, and H. Boehringer. Observation of the coma cluster of galaxies with rosat during the all-sky survey. *Astronomy and Astrophysics*, 1992.

[33] W. Buchmuller, K. Hamaguchi, and M. Ratz. Gauge couplings at high temperature and the relic gravitino abundance. *Physics Letters B*, 574(3-4):156–161, nov 2003.

[34] Esra Bulbul, Maxim Markevitch, Adam Foster, Randall K. Smith, Michael Loewenstein, and Scott W. Randall. Detection of an unidentified emission line in the stacked x-ray spectrum of galaxy clusters. *The Astrophysical Journal*, 789(1):13, jun 2014.

[35] Scott Burles, Kenneth M. Nollett, and Michael S. Turner. Big bang nucleosynthesis predictions for precision cosmology. *The Astrophysical Journal*, 552(1):L1–L5, may 2001.

[36] Dario Buttazzo, Giuseppe Degrassi, Pier Paolo Giardino, Gian F. Giudice, Filippo Sala, Alberto Salvio, and Alessandro Strumia. Investigating the near-criticality of the higgs boson. *Journal of High Energy Physics*, 2013(12), dec 2013.

[37] Giovanni Cabass, Alessandro Melchiorri, and Enrico Pajer. μ distortions or running: A guaranteed discovery from CMB spectrometry. *Physical Review D*, 93(8), apr 2016.

[38] Robert R. Caldwell, Vera Gluscevic, and Marc Kamionkowski. Cross-correlation of cosmological birefringence with CMB temperature. *Physical Review D*, 84(4), aug 2011.

[39] L.M. Capparelli, G. Cavoto, J. Ferretti, F. Giazotto, A.D. Polosa, and P. Spagnolo. Axion-like particle searches with sub-THz photons. *Physics of the Dark Universe*, 12:37–44, jun 2016.

[40] Ludovico M. Capparelli, Robert R. Caldwell, and Alessandro Melchiorri. Cosmic birefringence test of the hubble tension. *Submitted to PRL*.

[41] Ludovico M. Capparelli, A. Damiano, L. Maiani, and A. D. Polosa. A note on polarized light from magnetars. *The European Physical Journal C*, 77(11), nov 2017.

[42] Ludovico M. Capparelli, Eleonora Di Valentino, Alessandro Melchiorri, and Jens Chluba. Impact of theoretical assumptions in the determination of the neutrino effective number from future CMB measurements. *Physical Review D*, 97(6), mar 2018.

[43] Bradley Carroll and Dale Ostlie. *An introduction to modern astrophysics*. Cambridge University Press, 2017.

[44] Sean M. Carroll, George B. Field, and Roman Jackiw. Limits on a lorentz- and parity-violating modification of electrodynamics. *Physical Review D*, 41(4):1231–1240, feb 1990.

[45] Renyue Cen, Patrick McDonald, Hy Trac, and Abraham Loeb. Probing the epoch of reionization with the Lyα forest at $z \sim 4 - 5$. *The Astrophysical Journal*, 706(1):L164–L167, nov 2009.

[46] Daniel J. H. Chung, Gary Shiu, and Mark Trodden. Running of the scalar spectral index from inflationary models. *Physical Review D*, 68(6), sep 2003.

[47] S. A. Colgate. Supernovae as a standard candle for cosmology. *The Astrophysical Journal*, 232:404, sep 1979.

[48] ACBAR Collaboration. High-resolution observations of the cosmic microwave background power spectrum with ACBAR. *The Astrophysical Journal*, 600(1):32–51, jan 2004.

[49] ADMX Collaboration. Search for invisible axion dark matter with the axion dark matter experiment. *Physical Review Letters*, 120(15), apr 2018.

[50] ALEPH Collaboration. A precise determination of the number of families with light neutrinos and of the Z boson partial widths. *Physics Letters B*, 235(3-4):399–411, feb 1990.

[51] ATLAS Collaboration. Observation of a new particle in the search for the Standard Model Higgs boson with the ATLAS detector at the LHC. *Physics Letters B*, 716(1):1–29, sep 2012.

[52] BICEP2 Collaboration. BICEP2 / keck array IX: New bounds on anisotropies of CMB polarization rotation and implications for axionlike particles and primordial magnetic fields. *Physical Review D*, 96(10), nov 2017.

[53] CMB-S4 Collaboration. *CMB-S4 Science Book, First Edition*. arXiv:1610.02743 [astroph.CO], 2016.

[54] CMS Collaboration. Observation of a new boson at a mass of 125 GeV with the CMS experiment at the LHC. *Physics Letters B*, 716(1):30–61, sep 2012.

[55] CMS Collaboration. Search for direct production of supersymmetric partners of the top quark in the all-jets final state in proton-proton collisions at sqrt(s) = 13 TeV. *Journal of High Energy Physics*, 2017(10), oct 2017.

[56] CMS Collaboration. Search for natural supersymmetry in events with top quark pairs and photons in pp collisions at $\sqrt{s} = 8$ TeV. *Journal of High Energy Physics*, 2018(3), mar 2018.

[57] PLANCK Collaboration. https://www.cosmos.esa.int/web/planck/picture-gallery.

[58] PLANCK Collaboration. Planck2013 results. I. overview of products and scientific results. *Astronomy & Astrophysics*, 571:A1, oct 2014.

[59] PLANCK Collaboration. Planck 2015 results. XIII. cosmological parameters. *Astronomy & Astrophysics*, 594:A13, sep 2016.

[60] PLANCK Collaboration. https://wiki.cosmos.esa.int/planck-legacy-archive/images/9/9c/baseline_params_table_2018_95pc.pdf. 2018.

[61] PLANCK Collaboration. Planck 2018 results. VI. cosmological parameters. *[arXiv:1807.06209]*, 2018.

[62] SDSS Collaboration. The fourth data release of the sloan digital sky survey. *The Astrophysical Journal Supplement Series*, 162(1):38–48, jan 2006.

[63] UA1 Collaboration. Experimental observation of isolated large transverse energy electrons with associated missing energy at sqrt(s)=540 GeV. *Physics Letters B*, 122(1):103–116, feb 1983.

[64] UA2 Collaboration. Observation of single isolated electrons of high transverse momentum in events with missing transverse energy at the CERN pp collider. *Physics Letters B*, 122(5-6):476–485, mar 1983.

[65] XENON Collaboration. First dark matter search results from the XENON1t experiment. *Physical Review Letters*, 119(18), oct 2017.

[66] C. B. Collins and S. W. Hawking. Why is the universe isotropic? *The Astrophysical Journal*, 180:317, mar 1973.

[67] R. Consiglio, P.F. de Salas, G. Mangano, G. Miele, S. Pastor, and O. Pisanti. PArthENoPE reloaded. *Computer Physics Communications*, 233:237–242, dec 2018.

[68] E. Corbelli and P. Salucci. The extended rotation curve and the dark matter halo of m33. *Monthly Notices of the Royal Astronomical Society*, 311(2):441–447, jan 2000.

[69] Laura Covi, Jihn E. Kim, and Leszek Roszkowski. Axinos as cold dark matter. *Physical Review Letters*, 82(21):4180–4183, may 1999.

[70] Laura Covi, David Lyth, Alessandro Melchiorri, and Carolina Odman. Running-mass inflation model and WMAP. *Physical Review D*, 70(12), dec 2004.

[71] Richard H. Cyburt, Brian D. Fields, Keith A. Olive, and Tsung-Han Yeh. Big bang nucleosynthesis: Present status. *Reviews of Modern Physics*, 88(1), feb 2016.

[72] Sacha Davidson, Enrico Nardi, and Yosef Nir. Leptogenesis. *Physics Reports*, 466(4-5):105–177, sep 2008.

[73] W. J. G. de Blok. The core-cusp problem. *Advances in Astronomy*, 2010:1–14, 2010.

[74] W. J. G. de Blok, Stacy S. McGaugh, Albert Bosma, and Vera C. Rubin. Mass density profiles of low surface brightness galaxies. *The Astrophysical Journal*, 552(1):L23–L26, may 2001.

[75] Frederik Denef, Arthur Hebecker, and Timm Wrase. de sitter swampland conjecture and the higgs potential. *Physical Review D*, 98(8), oct 2018.

[76] R. H. Dicke, P. J. E. Peebles, P. G. Roll, and D. T. Wilkinson. Cosmic black-body radiation. *The Astrophysical Journal*, 142:414, jul 1965.

[77] Michael Dine. TASI lectures on the strong CP problem. *arXiv:hep-ph/0011376*, 2000.

[78] Michael Dine, Willy Fischler, and Mark Srednicki. A simple solution to the strong CP problem with a harmless axion. *Physics Letters B*, 104(3):199–202, aug 1981.

[79] Michael Dine and Alexander Kusenko. Origin of the matter-antimatter asymmetry. *Reviews of Modern Physics*, 76(1):1–30, dec 2003.

[80] Scott Dodelson. *Modern Cosmology*. Academic Press, 2003.

[81] Scott Dodelson and Lawrence M. Widrow. Sterile neutrinos as dark matter. *Physical Review Letters*, 72(1):17–20, jan 1994.

[82] Michael Doran. CMBEASY: an object oriented code for the cosmic microwave background. *Journal of Cosmology and Astroparticle Physics*, 2005(10):011–011, oct 2005.

[83] H. Dreiner and G.G. Ross. Sphaleron erasure of primordial baryogenesis. *Nuclear Physics B*, 410(1):188–216, dec 1993.

[84] Leanne D Duffy and Karl van Bibber. Axions as dark matter particles. *New Journal of Physics*, 11(10):105008, oct 2009.

[85] T. M. Dunster. *NIST Handbook of Mathematical functions [https://dlmf.nist.gov/14]*. Cambridge University Press, 2010.

[86] Richard Easther and Hiranya V Peiris. Implications of a running spectral index for slow roll inflation. *Journal of Cosmology and Astroparticle Physics*, 2006(09):010–010, sep 2006.

[87] Michael Eastwood and Paul Tod. Edth-a differential operator on the sphere. *Mathematical Proceedings of the Cambridge Philosophical Society*, 92(2):317–330, sep 1982.

[88] John Ellis and Kazuki Sakurai. Search for sphalerons in proton-proton collisions. *Journal of High Energy Physics*, 2016(4):1–15, apr 2016.

[89] Richard I. Epstein, James M. Lattimer, and David N. Schramm. The origin of deuterium. In *The Big Bang and Other Explosions in Nuclear and Particle Astrophysics*, pages 87–91. World scientific, jun 1996.

[90] A. Suzuki et al. The LiteBIRD satellite mission: Sub-kelvin instrument. *Journal of Low Temperature Physics*, 193(5-6):1048–1056, may 2018.

[91] Adam G. Riess et al. Observational evidence from supernovae for an accelerating universe and a cosmological constant. *The Astronomical Journal*, 116(3):1009–1038, sep 1998.

[92] G. Hinshaw et al. Nine-year Wilkinson microwave anisotropy probe (WMAP) observations: cosmological parameter results. *The Astrophysical Journal Supplement Series*, 208(2):19, sep 2013.

[93] M. E. Abroe et al. Frequentist estimation of cosmological parameters from the MAXIMA-1 cosmic microwave background anisotropy data. *Monthly Notices of the Royal Astronomical Society*, 334(1):11–19, jul 2002.

[94] Peter Ade et al. The simons observatory: science goals and forecasts. *Journal of Cosmology and Astroparticle Physics*, 2019(02):056–056, feb 2019.

[95] R. W. Pattie et al. Measurement of the neutron lifetime using a magneto-gravitational trap and in situ detection. *Science*, 360(6389):627–632, may 2018.

[96] Saul Perlmutter et al. Measurements of ω and λ from 42 high-redshift supernovae. *The Astrophysical Journal*, 1999.

[97] T. J. Pearson et al. The anisotropy of the microwave background to ℓ = 3500: Mosaic observations with the cosmic background imager. *The Astrophysical Journal*, 591(2):556–574, jul 2003.

[98] T. Matsumura et al. Mission design of LiteBIRD. *Journal of Low Temperature Physics*, 176(5-6):733–740, jan 2014.

[99] M. Tanabashi et al. (Particle Data Group). The review of particle physics. *Physical Review D*, 98, 2018.

[100] James Evans. Anaximander. *Encyclopaedia Britannica*.

[101] Brian D. Fields. The primordial lithium problem. *Annual Review of Nuclear and Particle Science*, 61(1):47–68, nov 2011.

[102] D. J. Fixsen. The temperature of the cosmic microwave background. *The Astrophysical Journal*, 707(2):916–920, nov 2009.

[103] Grant R. Fowles. *Introduction to Modern Optics*. Dover Publications, 2009.

[104] Wendy L. Freedman. Cosmology at a crossroads. *Nature astronomy*, 2017.

[105] K. Freese. Review of observational evidence for dark matter in the universe and in upcoming searches for dark stars. *EAS Publications Series*, 36:113–126, 2009.

[106] A. Friedmann. Uber die moglichkeit einer welt mit konstanter negativer krummung des raumes. *Zeitschrift fur Physik*, 21(1):326–332, dec 1924.

[107] Paul A. Gagniuc. *Markov Chains*. John Wiley and Sons Ltd, 2017.

[108] G. Gamow. Quantum theory of the atomic nucleus. *Zeitschrift fur physik*, 1928.

[109] Andrew Gelman and Donald B. Rubin. Inference from iterative simulation using multiple sequences. *Statistical science*, 7(4):457–511, 1992.

[110] Howard Georgi and S. L. Glashow. Unity of all elementary-particle forces. *Physical Review Letters*, 32(8):438–441, feb 1974.

[111] G.F. Giudice and M. Shaposhnikov. Strong sphalerons and electroweak baryogenesis. *Physics Letters B*, 326(1-2):118–124, apr 1994.

[112] Carlo Giunti. *Fundamentals of Neutrino Physics and Astrophysics*. OUP Oxford, 2007.

[113] S. L. Glashow, J. Iliopoulos, and L. Maiani. Weak interactions with lepton-hadron symmetry. *Physical Review D*, 2(7):1285–1292, oct 1970.

[114] Vera Gluscevic, Marc Kamionkowski, and Asantha Cooray. Derotation of the cosmic microwave background polarization: Full-sky formalism. *Physical Review D*, 80(2), jul 2009.

[115] J. N. Goldberg, A. J. Macfarlane, E. T. Newman, F. Rohrlich, and E. C. G. Sudarshan. Spin-s spherical harmonics and ð. *Journal of Mathematical Physics*, 8(11):2155–2161, nov 1967.

[116] Herbert Goldstein. *Classical mechanics*. Addison-Wesley, 1980.

[117] Christophe Grojean, Géraldine Servant, and James D. Wells. First-order electroweak phase transition in the standard model with a low cutoff. *Physical Review D*, 71(3), feb 2005.

[118] David J. Gross and Frank Wilczek. Ultraviolet behavior of non-abelian gauge theories. *Physical Review Letters*, 30(26):1343–1346, jun 1973.

[119] James E. Gunn and Bruce A. Peterson. On the density of neutral hydrogen in intergalactic space. *The Astrophysical Journal*, 142:1633, nov 1965.

[120] Alan H. Guth. Inflationary universe: A possible solution to the horizon and flatness problems. *Physical Review D*, 23(2):347–356, jan 1981.

[121] W. K. Hastings. Monte carlo sampling methods using markov chains and their applications. *Biometrika*, 57(1):97–109, apr 1970.

[122] Peter W. Higgs. Broken symmetries and the masses of gauge bosons. *Physical Review Letters*, 13(16):508–509, oct 1964.

[123] Zhen Hou, Ryan Keisler, Lloyd Knox, Marius Millea, and Christian Reichardt. How massless neutrinos affect the cosmic microwave background damping tail. *Physical Review D*, 87(8), apr 2013.

[124] Cullan Howlett, Antony Lewis, Alex Hall, and Anthony Challinor. CMB power spectrum parameter degeneracies in the era of precision cosmology. *Journal of Cosmology and Astroparticle Physics*, 2012(04):027–027, apr 2012.

[125] Ester M. Hu and Lennox L. Cowie. High-redshift galaxy populations. *Nature*, 2006.

[126] Wayne Hu and Naoshi Sugiyama. Small-scale cosmological perturbations: An analytic approach. *The Astrophysical Journal*, 471(2):542–570, nov 1996.

[127] Wayne Hu, Naoshi Sugiyama, and Joseph Silk. The physics of microwave background anisotropies. *Nature*, 386(6620):37–43, mar 1997.

[128] Wayne Hu and Martin White. CMB anisotropies: Total angular momentum method. *Physical Review D*, 56(2):596–615, jul 1997.

[129] Michael J. Hudson. Cosmic flows: Toward an understanding of the large-scale structure in the universe. *Publications of the Astronomical Society of the Pacific*, 111(765):1469–1471, nov 1999.

[130] Lam Hui, Jeremiah P. Ostriker, Scott Tremaine, and Edward Witten. Ultralight scalars as cosmological dark matter. *Physical Review D*, 95(4), feb 2017.

[131] Dragan Huterer and Daniel L Shafer. Dark energy two decades after: observables, probes, consistency tests. *Reports on Progress in Physics*, 81(1):016901, dec 2017.

[132] Y. I. Izotov, T. X. Thuan, and N. G. Guseva. A new determination of the primordial he abundance using the he i λ10830 a emission line: cosmological implications. 445(1):778–793, sep 2014.

[133] Yuri I. Izotov, Trinh X. Thuan, and Grażyna Stasińska. The primordial abundance of4he: A self-consistent empirical analysis of systematic effects in a large sample of low-metallicity HiiRegions. *The Astrophysical Journal*, 662(1):15–38, jun 2007.

[134] Marc Kamionkowski. How to derotate the cosmic microwave background polarization. *Physical Review Letters*, 102(11), mar 2009.

[135] Marc Kamionkowski, Arthur Kosowsky, and Albert Stebbins. A probe of primordial gravity waves and vorticity. *Physical Review Letters*, 78(11):2058–2061, mar 1997.

[136] Marc Kamionkowski, Arthur Kosowsky, and Albert Stebbins. Statistics of cosmic microwave background polarization. *Physical Review D*, 55(12):7368–7388, jun 1997.

[137] Marc Kamionkowski, Josef Pradler, and Devin G. E. Walker. Dark energy from the string axiverse. *Physical Review Letters*, 113(25), dec 2014.

[138] Jihn E. Kim. Weak-interaction singlet and Strong CP Invariance. *Physical Review Letters*, 43(2):103–107, jul 1979.

[139] F. R. Klinkhamer and N. S. Manton. A saddle-point solution in the weinberg-salam theory. *Physical Review D*, 30(10):2212–2220, nov 1984.

[140] Lloyd Knox. Determination of inflationary observables by cosmic microwave background anisotropy experiments. *Physical Review D*, 52(8):4307–4318, oct 1995.

[141] Kazunori Kohri and Tomohiro Matsuda. Ambiguity in running spectral index with an extra light field during inflation. *Journal of Cosmology and Astroparticle Physics*, 2015(02):019–019, feb 2015.

[142] Edward W. Kolb and Michaesl S. Turner. The early universe.

[143] Edward W. Kolb and Michaesl S. Turner. *The Early Universe*. Frontiers in physics, 1990.

[144] Arthur Kosowsky. Cosmic microwave background polarization. *Annals of Physics*, 246(1):49–85, feb 1996.

[145] Holger G. Krapp. Polarization vision: How insects find their way by watching the sky. *Current Biology*, 17(14):R557–R560, jul 2007.

[146] Alexander Kusenko and Mikhail Shaposhnikov. Supersymmetric Q-balls as dark matter. *Physics Letters B*, 418(1-2):46–54, jan 1998.

[147] Julien Lesgourgues and Thomas Tram. The cosmic linear anisotropy solving system (CLASS) IV: efficient implementation of non-cold relics. *Journal of Cosmology and Astroparticle Physics*, 2011(09):032–032, sep 2011.

[148] Julien Lesgourgues and Thomas Tram. Fast and accurate CMB computations in non-flat FLRW universes. *Journal of Cosmology and Astroparticle Physics*, 2014(09):032–032, sep 2014.

[149] A Lewis and A. Challinor. Weak gravitational lensing of the CMB. *Physics Reports*, 429(1):1–65, jun 2006.

[150] Antony Lewis and Sarah Bridle. Cosmological parameters from CMB and other data: A monte carlo approach. *Physical Review D*, 66(10), nov 2002.

[151] Antony Lewis, Anthony Challinor, and Anthony Lasenby. Efficient computation of cosmic microwave background anisotropies in closed friedmann-robertson-walker models. *The Astrophysical Journal*, 538(2):473–476, aug 2000.

[152] Andrew R. Liddle and Robert J. Scherrer. Classification of scalar field potentials with cosmological scaling solutions. *Physical Review D*, 59(2), dec 1998.

[153] Andrei Linde. Inflationary cosmology after planck. In *Post-Planck Cosmology*, pages 230–316. Oxford University Press, feb 2015.

[154] S. Linden, J.-M. Virey, and A. Tilquin. Cosmological parameter extraction and biases from type ia supernova magnitude evolution. *Astronomy & Astrophysics*, 506(3):1095–1105, sep 2009.

[155] Chung-Pei Ma and Edmund Bertschinger. Cosmological perturbation theory in the synchronous and conformal newtonian gauges. *The Astrophysical Journal*, 455:7, dec 1995.

[156] David J. E. Marsh and Ana-Roxana Pop. Axion dark matter, solitons and the cusp–core problem. *Monthly Notices of the Royal Astronomical Society*, 451(3):2479–2492, jun 2015.

[157] N. Menci, A. Grazian, M. Castellano, and N. G. Sanchez. A stringent limit on the warm dark matter particle masses from the abundance of z = 6 galaxies in the Hubble frontier fields. *The Astrophysical Journal*, 825(1):L1, jun 2016.

[158] J. Miralda-Escude. The dark age of the universe. *Science*, 300(5627):1904–1909, jun 2003.

[159] C. W. Misner, K. S. Thorne, and J. A. Wheeler. *Gravitation*. W H Freeman & Co, 1973.

[160] Edvard Mörtsell and Suhail Dhawan. Does the hubble constant tension call for new physics? *Journal of Cosmology and Astroparticle Physics*, 2018(09):025–025, sep 2018.

[161] Keith A. Olive. Tasi lectures on dark matter. *[arXiv:astro-ph/0301505]*, 2003.

[162] Keith A. Olive and Evan D. Skillman. A realistic determination of the error on the primordial helium abundance: Steps toward nonparametric nebular helium abundances. *The Astrophysical Journal*, 617(1):29–49, dec 2004.

[163] Thanu Padmanbhan. *Structure formation in the universe*. Cambridge University Press, 1993.

[164] R. D. Peccei and Helen R. Quinn. Constraints imposed by CP conservation in the presence of pseudoparticles. *Physical Review D*, 16(6):1791–1797, sep 1977.

[165] R. D. Peccei and Helen R. Quinn. CP conservation in the presence of pseudoparticles. *Physical Review Letters*, 38(25):1440–1443, jun 1977.

[166] Roberto D. Peccei. The strong CP problem and axions. In *Lecture Notes in Physics*, pages 3–17. Springer Berlin Heidelberg, 2008.

[167] P. J. E. Peebles. Recombination of the primeval plasma. *The Astrophysical Journal*, 153:1, jul 1968.

[168] A. A. Penzias and R. W. Wilson. A measurement of excess antenna temperature at 4080 mc/s. *The Astrophysical Journal*, 142:419, jul 1965.

[169] Laurence Perotto, Julien Lesgourgues, Steen Hannestad, Huitzu Tu, and Yvonne Y Y Wong. Probing cosmological parameters with the CMB: forecasts from monte carlo simulations. *Journal of Cosmology and Astroparticle Physics*, 2006(10):013–013, oct 2006.

[170] Michael E. Peskin and Daniel V. Schroeder. *An Introduction to quantum field theory*. CRC Press, 2019.

[171] Patrick Peter and Jean-Philippe Uzan. *Primordial cosmology*. Oxford Graduate Texts, 2009.

[172] Max Pettini and David V. Bowen. A new measurement of the primordial abundance of deuterium: Toward convergence with the baryon density from the cosmic microwave background? *The Astrophysical Journal*, 560(1):41–48, oct 2001.

[173] Cyril Pitrou, Alain Coc, Jean-Philippe Uzan, and Elisabeth Vangioni. Precision big bang nucleosynthesis with improved helium-4 predictions. *Physics Reports*, 754:1–66, sep 2018.

[174] Levon Pogosian, Meir Shimon, Matthew Mewes, and Brian Keating. Future CMB constraints on cosmic birefringence and implications for fundamental physics. *Physical Review D*, 100(2), jul 2019.

[175] Maxim Pospelov, Adam Ritz, and Constantinos Skordis. Pseudoscalar perturbations and polarization of the cosmic microwave background. *Physical Review Letters*, 103(5), jul 2009.

[176] Vivian Poulin, Tristan L. Smith, Daniel Grin, Tanvi Karwal, and Marc Kamionkowski. Cosmological implications of ultralight axionlike fields. *Physical Review D*, 98(8), oct 2018.

[177] Vivian Poulin, Tristan L. Smith, Tanvi Karwal, and Marc Kamionkowski. Early dark energy can resolve the hubble tension. *Physical Review Letters*, 122(22), jun 2019.

[178] John P. Preskill. Cosmological production of superheavy magnetic monopoles. *Physical Review Letters*, 43(19):1365–1368, nov 1979.

[179] The MACHO Project. The MACHO project: Microlensing results from 5.7 years of large magellanic cloud observations. *The Astrophysical Journal*, 542(1):281–307, oct 2000.

[180] Adam G. Riess, Stefano Casertano, Wenlong Yuan, Lucas M. Macri, and Dan Scolnic. Large magellanic cloud cepheid standards provide a 1% foundation for the determination of the hubble constant and stronger evidence for physics beyond the lcdm. *The Astrophysical Journal*, 876(1), 2019.

[181] Antonio Riotto. Inflation and the theory of cosmological perturbations. *ICTP Lecture Notes Series 14 [https://arxiv.org/abs/hep-ph/0210162]*, (2003).

[182] Leslie J Rosenberg and Karl A. van Bibber. Searches for invisible axions. *Physics Reports*, 325(1):1–39, feb 2000.

[183] S.E. Rugh and H. Zinkernagel. The quantum vacuum and the cosmological constant problem. *Studies in History and Philosophy of Science Part B: Studies in History and Philosophy of Modern Physics*, 33(4):663–705, dec 2002.

[184] Bertrand Russell. *Why I am not a Christian*. Rationalist Press Association Limited, 1927.

[185] J. J. Sakurai and Jim Napolitano. *Modern quantum mechanics*. Addison-Wesley, 2010.

[186] L. Salvati, L. Pagano, R. Consiglio, and A. Melchiorri. Cosmological constraints on the neutron lifetime. *Journal of Cosmology and Astroparticle Physics*, 2016(03):055–055, mar 2016.

[187] Dimitar D. Sasselov. https://www.cfa.harvard.edu/ sasselov/rec/.

[188] T. Schafer and E. V. Shuryak. Instantons in QCD. *Reviews of Modern Physics*, 70(2):323–425, apr 1998.

[189] Matthew D. Schwartz. *Quantum field theory and the standard model*. Cambridge University Press, 2013.

[190] S. Seager, D. D. Sasselov, and D. Scott. A new calculation of the recombination epoch. *The Astrophysical Journal*, 523(1):L1–L5, sep 1999.

[191] Sara Seager, Dimitar D. Sasselov, and Douglas Scott. How exactly did the universe become neutral? *The Astrophysical Journal Supplement Series*, 128(2):407–430, jun 2000.

[192] Uros Seljak and Matias Zaldarriaga. A line-of-sight integration approach to cosmic microwave background anisotropies. *The Astrophysical Journal*, 469:437, oct 1996.

[193] A. P. Serebrov, V. E. Varlamov, A. G. Kharitonov, A. K. Fomin, Yu. N. Pokotilovski, P. Geltenbort, I. A. Krasnoschekova, M. S. Lasakov, R. R. Taldaev, A. V. Vassiljev, and O. M. Zherebtsov. Neutron lifetime measurements using gravitationally trapped ultracold neutrons. *Physical Review C*, 78(3), sep 2008.

[194] Ramamurti Shankar. *Principles of quantum mechanics*. Springer, 2013.

[195] Ilya L. Shapiro and Joan Solà. On the possible running of the cosmological "constant". *Physics Letters B*, 682(1):105–113, nov 2009.

[196] M.A. Shifman, A.I. Vainshtein, and V.I. Zakharov. Can confinement ensure natural CP invariance of strong interactions? *Nuclear Physics B*, 166(3):493–506, apr 1980.

[197] P. Sikivie. Axions, domain walls, and the early universe. *Physical Review Letters*, 48(17):1156–1159, apr 1982.

[198] Gabrielle Simard, Duncan Hanson, and Gil Holder. Prospects for delensing the cosmic microwave background for studying inflation. *The Astrophysical Journal*, 807(2):166, jul 2015.

[199] David Skinner. Mathematical methods lecture notes. *http://www.damtp.cam.ac.uk/user/dbs26/1BMethods/All.pdf*.

[200] Douglas Spolyar, Katherine Freese, and Paolo Gondolo. Dark matter and the first stars: A new phase of stellar evolution. *Physical Review Letters*, 100(5), feb 2008.

[201] Mark Srednicki. Axion couplings to matter. *Nuclear Physics B*, 260(3-4):689–700, oct 1985.

[202] A.A. Starobinsky. A new type of isotropic cosmological models without singularity. *Physics Letters B*, 91(1):99–102, mar 1980.

[203] Paul J. Steinhardt. A quintessential introduction to dark energy. *Philosophical transactions: mathematical, physical and engineering sciences*, 2003.

[204] Peter Svrcek and Edward Witten. Axions in string theory. *Journal of High Energy Physics*, 2006(06):051–051, jun 2006.

[205] G. 't Hooft. *Physical Review Letters*, 37(8), 1976.

[206] G. 't Hooft. *Physical Review D*, 1976.

[207] Ryuichi Takahashi, Masanori Sato, Takahiro Nishimichi, Atsushi Taruya, and Masamune Oguri. Revising the HALOFIT model for the nonlinear matter power spectrum. *The Astrophysical Journal*, 761(2):152, dec 2012.

[208] Thomas Tram and Julien Lesgourgues. Optimal polarisation equations in FLRW universes. *Journal of Cosmology and Astroparticle Physics*, 2013(10):002–002, oct 2013.

[209] J. Anthony Tyson, Greg P. Kochanski, and Ian P. DellAntonio. Detailed mass map of CL 0024+1654 from strong lensing. *The Astrophysical Journal*, 498(2):L107–L110, may 1998.

[210] Eleonora Di Valentino, Alessandro Melchiorri, and Olga Mena. Can interacting dark energy solve the H_0 tension? *Physical Review D*, 96(4), aug 2017.

[211] L. Verde. Statistical methods in cosmology. In *Lectures on Cosmology*, pages 147–177. Springer Berlin Heidelberg, 2010.

[212] Steven Weinberg. A model of leptons. *Physical Review Letters*, 19(21):1264–1266, nov 1967.

[213] Steven Weinberg. A new light boson? *Physical Review Letters*, 40(4):223–226, jan 1978.

[214] Steven Weinberg. *The quantum theory of fields, volume III: Supersymmetry*. Cambridge University Press, 2000.

[215] Fred E. Wietfeldt and Geoffrey L. Greene. Colloquium: The neutron lifetime. *Reviews of Modern Physics*, 83(4):1173–1192, nov 2011.

[216] Eugene P. Wigner and J. J. Griffin. *Group Theory and its Application to the Quantum Mechanics of Atomic Spectra*. Academic Press, 1959.

[217] F. Wilczek. Problem of strong P and T invariance in the presence of instantons. *Physical Review Letters*, 40(5):279–282, jan 1978.

[218] John H. Wise. Cosmic reionisation. *Contemporary Physics*, 60(2):145–163, apr 2019.

[219] Edward Witten. Cosmic separation of phases. *Physical Review D*, 30(2):272–285, jul 1984.

[220] A. T. Yue, M. S. Dewey, D. M. Gilliam, G. L. Greene, A. B. Laptev, J. S. Nico, W. M. Snow, and F. E. Wietfeldt. Improved determination of the neutron lifetime. *Physical Review Letters*, 111(22), nov 2013.

[221] Matias Zaldarriaga and Uroš Seljak. All-sky analysis of polarization in the microwave background. *Physical Review D*, 55(4):1830–1840, feb 1997.

[222] Matias Zaldarriaga and Uros Seljak. CMBFAST for spatially closed universes. *The Astrophysical Journal Supplement Series*, 129(2):431–434, aug 2000.

[223] Saleem Zaroubi. The epoch of reionization. In *The First Galaxies*, pages 45–101. Springer Berlin Heidelberg, sep 2012.

[224] F. Zwicky. On the masses of nebulae and of clusters of nebulae. *The Astrophysical Journal*, 86:217, oct 1937.