

Part-III Advanced Cosmology

Lent Term 2017

Lecture notes: Physics of the cosmic microwave background

Anthony Challinor

a.d.challinor@ast.cam.ac.uk

These notes accompany the section (around 12 lectures) of the Part-III course *Advanced Cosmology* that covers the physics of the cosmic microwave background (CMB). The CMB provides a powerful window into the early universe. It started life as high-energy photons in thermal equilibrium in the early universe. Once the temperature dropped to around 3000 K, electrons and protons recombined to form hydrogen atoms and the universe became essentially transparent to CMB photons. The CMB subsequently free-streamed to the present, preserving its blackbody (thermal) spectrum but with wavelengths stretched into the microwave range due to the 1000-fold expansion of the universe since recombination. Today, we observe the CMB as almost perfect blackbody radiation with a temperature $\bar{T}_{\text{CMB}} = 2.7255 \text{ K}$. Looking more closely, we see that the CMB temperature is not completely uniform on the sky. The largest departure from isotropy is a dipole, with amplitude around 3 mK. We interpret this as being due to our peculiar motion. Looking deeper still, we find small angular variations, or *anisotropies*, in the temperature with an r.m.s. of around $110 \mu\text{K}$. A map of the CMB temperature anisotropies based on measurements by the Planck satellite is shown in Fig. 1. As we shall learn, these temperature anisotropies are sourced mostly by density and velocity fluctuations in the pre-recombination plasma, and the associated gravitational potentials. The CMB provides the earliest picture we have of the primordial fluctuations.

The CMB temperature anisotropies and its polarization are a very powerful cosmological probe, and that they can tell us both about the composition of the universe (i.e., what the main constituents are) and the statistical properties of the primordial perturbations. The reason the CMB is such a powerful probe is that when most CMB photons decoupled from matter around recombination, at redshift $z \approx 1100$, the fluctuations were still accurately described by linear perturbation theory. The anisotropies are therefore (almost) a linearly-processed version of the primordial perturbations, where the linear processing depends on the acoustic physics of the pre-recombination plasma that you discussed in the *Cosmology* course.

Linearity makes it quite straightforward to calculate the statistics of the anisotropies to high accuracy, and we shall see how to do this in detail. To capture all of the effects demanded by the accuracy of current data, we must use a kinetic theory description of the CMB that allows us to seamlessly discuss the transition from tightly-coupled radiation, with fluid-like behaviour, in the pre-recombination era to the free-streaming

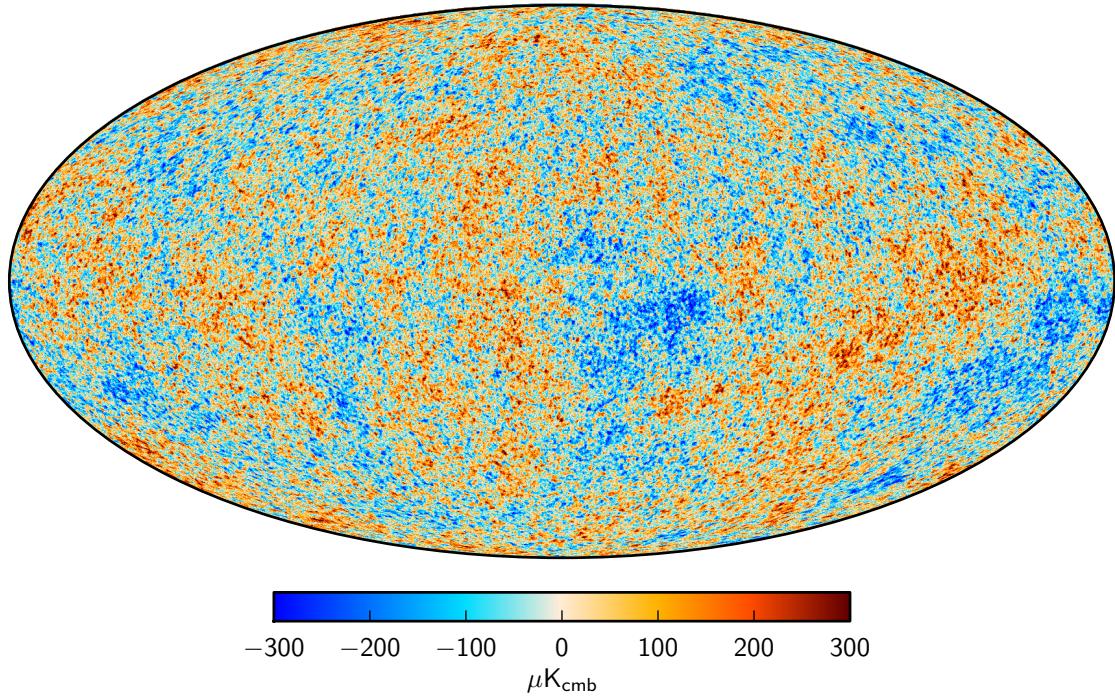


Figure 1: Temperature anisotropies of the CMB as measured by Planck.

radiation that we detect today. We shall therefore make extensive use of the ideas of relativistic kinetic theory, and these will be introduced as needed. A kinetic theory treatment is essential for an accurate calculation of the CMB anisotropies on small scales, where photon diffusion becomes significant.

The second main focus here is to develop further the idea that the CMB can constrain the physics of the early universe (e.g., inflation) that we believe generated the primordial perturbations. We shall see how gravitational waves produced during inflation affect the CMB, and introduce a new observable – the linear polarization of the CMB – which turns out to be a very promising route to detect such gravitational waves. At the end of the course, Paul Shellard will return to discuss a further key discriminant of inflationary models, the non-Gaussian statistics of the primordial perturbations. He will also discuss how the CMB and other cosmological probes can be used to constrain primordial non-Gaussianity. Paul Shellard’s lectures are not covered in these notes.

A roadmap for this part of the course, and these notes, is as follows:

- Statistics of random fields on the sphere – Sec. 1
- Relativistic kinetic theory and the Boltzmann equation – Sec. 2
- Temperature anisotropies from scalar perturbations – Sec. 3

- Scalar perturbations on small scales: diffusion damping and small-scale anisotropies – Sec. 4
- CMB anisotropies from gravitational waves – Sec. 5
- CMB polarization – Sec. 6

Recommended books

- *Modern Cosmology* by Dodelson (Academic Press). Treats many of the topics in this course in a pedagogical style.
- *The Cosmic Microwave Background* by Durrer (Cambridge University Press).
- *Angular Momentum* (third edition) by Brink & Satchler (Oxford Science Publications). Representations of the rotation group crop up frequently in CMB physics and this quantum mechanics text is a good introduction to the theory.
- Wayne Hu’s excellent website (<http://background.uchicago.edu/~whu/>). Not a book, but the CMB tutorials are well worth a visit.

1 Statistics of random fields on the sphere

Theory (e.g., quantum mechanics during inflation) only allows us to predict the statistical properties of cosmological fields (such as the matter overdensity $\delta\rho$). Here, we explore the basic statistical properties enforced on such fields by assuming the physics that generates the initial fluctuations, and subsequently processes them, respects the symmetries of the background cosmology, i.e., isotropy and homogeneity.

For 3D random fields, the situation is rather straightforward. Consider a random field $f(\mathbf{x})$ – i.e., at each point $f(\mathbf{x})$ is some random number – with zero mean, $\langle f(\mathbf{x}) \rangle = 0$. The probability of realising some field configuration is a *functional* $\text{Pr}[f(\mathbf{x})]$. *Correlators* of fields are expectation values of products of fields at different spatial points (and, generally, times). The correlators take a particularly simple form in Fourier space. You met the 2-point correlator, or power spectrum, in the Part-III *Cosmology* course¹:

$$\langle f(\mathbf{k}) f^*(\mathbf{k}') \rangle = \frac{2\pi^2}{k^3} \mathcal{P}_f(k) \delta^{(3)}(\mathbf{k} - \mathbf{k}') . \quad (1.2)$$

¹Our Fourier convention is that

$$f(\mathbf{k}) = \int \frac{d^3\mathbf{x}}{(2\pi)^{3/2}} f(\mathbf{x}) e^{-i\mathbf{k}\cdot\mathbf{x}} \quad \text{and} \quad f(\mathbf{x}) = \int \frac{d^3\mathbf{k}}{(2\pi)^{3/2}} f(\mathbf{k}) e^{i\mathbf{k}\cdot\mathbf{x}} . \quad (1.1)$$

Note that for real fields, $f(\mathbf{k}) = f^*(-\mathbf{k})$.

Here, $\mathcal{P}_f(k)$ is the *dimensionless* power spectrum. It depends only on $k = |\mathbf{k}|$ because of statistical isotropy. The Dirac delta function appears because of statistical homogeneity: the correlator has to be invariant under active translation of the field, which in Fourier space becomes a transformation $f(\mathbf{k}) \rightarrow f(\mathbf{k})e^{-i\mathbf{k}\cdot\mathbf{a}}$. Similarly, for the 3-point correlator, we have

$$\langle f(\mathbf{k}_1)f(\mathbf{k}_2)f(\mathbf{k}_3) \rangle \propto B_f(k_1, k_2, k_3)\delta^{(3)}(\mathbf{k}_1 + \mathbf{k}_2 + \mathbf{k}_3), \quad (1.3)$$

where $B_f(k_1, k_2, k_3)$ is the 3D *bispectrum* of f , and is a symmetric function of the three wavenumbers. Again, the Dirac delta function in Eq. (1.3) arises from statistical homogeneity. Why does the bispectrum depend only on the three wavenumbers? We start with nine degrees of freedom corresponding to the components of the three 3D wavevectors. However, three of these are eliminated by statistical homogeneity, and a further three by statistical isotropy. The 3-point function can therefore only depend on the three rotational-invariants associated with triangular configurations of the three wavevectors, which we can take to be the magnitudes of the wavevectors. The 3-point function (and all N -point functions for N odd) vanishes for zero-mean *Gaussian* random fields (see below for definition) and so provides a critical test of theories, such as simple inflation models, that predict very nearly Gaussian primordial perturbations.

Gaussian random fields: For a Gaussian (homogeneous and isotropic) random field, $\text{Pr}[f(\mathbf{x})]$ is a Gaussian functional of $f(\mathbf{x})$. If we think of discretising the field in N pixels, so it is represented by a N -dimensional vector $\mathbf{f} = [f(\mathbf{x}_1), f(\mathbf{x}_2), \dots, f(\mathbf{x}_N)]^T$, the probability density function for \mathbf{f} is a multi-variate Gaussian fully specified by the covariance matrix

$$\langle f_i f_j \rangle = C_{ij}, \quad (1.4)$$

where $f_i \equiv f(\mathbf{x}_i)$, so that

$$\text{Pr}(\mathbf{f}) = \frac{e^{-\mathbf{f}^T \mathbf{C}^{-1} \mathbf{f} / 2}}{\sqrt{\det(2\pi \mathbf{C})}}. \quad (1.5)$$

The covariance matrix elements C_{ij} depends only on $|\mathbf{x}_i - \mathbf{x}_j|$ by statistical homogeneity and isotropy. Since $f(\mathbf{k})$ is linear in $f(\mathbf{x})$, the probability distribution for $f(\mathbf{k})$ is also a multi-variate Gaussian. Since different Fourier modes are uncorrelated (see Eq. (1.2)), they are statistically *independent* for Gaussian fields.

As will be discussed at the end of this course, simple inflation models predict fluctuations that are very nearly Gaussian and this property is preserved by *linear* evolution. The cosmic microwave background probes fluctuations mostly in the linear regime and so the fluctuations look very Gaussian. Non-linear structure formation at late times destroys Gaussianity and gives rise to the filamentary cosmic web. Searching for primordial non-Gaussianity to probe departures from simple inflation is a very hot topic but no convincing evidence for primordial non-Gaussianity has yet been found.

In cosmology, what we observe is generally a projection of some 3D random field onto the celestial sphere. For example, as we shall learn the CMB anisotropies are mostly the projection of the photon density, bulk velocity and the gravitational potential over the surface of last-scattering. In the rest of this section, we shall develop some of the language used to describe random fields on the sphere.

We shall consider a generic zero-mean random field on the sphere, denoted $f(\hat{\mathbf{n}})$. Here, $\hat{\mathbf{n}}$ is a unit vector along the line-of-sight. We shall usually be interested in real fields f . Spherical harmonics form a basis for (square-integrable) functions on the sphere:

$$f(\hat{\mathbf{n}}) = \sum_{l=0}^{\infty} \sum_{m=-l}^l f_{lm} Y_{lm}(\hat{\mathbf{n}}). \quad (1.6)$$

The Y_{lm} are familiar from quantum mechanics as the position-space representation of the eigenstates of $\hat{L}^2 = -\nabla^2$ and $\hat{L}_z = -i\partial_\phi$ (setting $\hbar = 1$):

$$\begin{aligned} \nabla^2 Y_{lm} &= -l(l+1)Y_{lm} \\ \partial_\phi Y_{lm} &= imY_{lm}, \end{aligned} \quad (1.7)$$

with l an integer ≥ 0 and m an integer with $|m| \leq l$. The spherical harmonics are orthonormal over the sphere,

$$\int d\hat{\mathbf{n}} Y_{lm}(\hat{\mathbf{n}}) Y_{l'm'}^*(\hat{\mathbf{n}}) = \delta_{ll'} \delta_{mm'}, \quad (1.8)$$

so that the *spherical multipole coefficients* of $f(\hat{\mathbf{n}})$ are

$$f_{lm} = \int d\hat{\mathbf{n}} f(\hat{\mathbf{n}}) Y_{lm}^*(\hat{\mathbf{n}}). \quad (1.9)$$

There are various phase conventions for the Y_{lm} ; here we adopt $Y_{lm}^* = (-1)^m Y_{l-m}$ so that $f_{lm}^* = (-1)^m f_{l-m}$ for a real field.

The $m = 0$ spherical harmonics are axisymmetric about the z -axis. Their dependence on the polar angle θ is proportional to the Legendre polynomials, $P_l(\cos \theta)$, giving an angular wavelength of approximately $2\pi/l$.

1.1 Transformation under rotations

We have seen that statistical isotropy and homogeneity restrict the form of the correlations between random 3D fields. For fields over the sphere, the relevant symmetry is statistical isotropy. If we are to build statistically-isotropic correlators for the f_{lm} , we need to understand first how the f_{lm} transform under rotations.

A rotation in 3D is fully specified by three *Euler angles*, often denoted α , β and γ . The rotation $D(\alpha, \beta, \gamma)$ consists of the following sequence of rotations:

- first, rotate by γ about the z -axis;
- second, rotate by β about the original y -axis; and
- third, rotate by α about the original z -axis.

We see from Eq. (1.7) that the ϕ dependence of $Y_{lm}(\hat{\mathbf{n}})$ is $e^{im\phi}$, so under a rotation by γ about the z -axis,

$$\begin{aligned}\hat{D}(0, 0, \gamma)Y_{lm}](\theta, \phi) &= Y_{lm}(\theta, \phi - \gamma) \\ &= e^{-im\gamma}Y_{lm}(\theta, \phi) \\ &= e^{-i\gamma\hat{L}_z}Y_{lm}.\end{aligned}\tag{1.10}$$

It follows that \hat{L}_z is the *generator* of rotations about the z -axis and hence \hat{L}_i generates rotations about the i th axis. For a general rotation, we have

$$\hat{D}(\alpha, \beta, \gamma) = e^{-i\alpha\hat{L}_z}e^{-i\beta\hat{L}_y}e^{-i\gamma\hat{L}_z}.\tag{1.11}$$

Since \hat{L}^2 commutes with the \hat{L}_i , the rotated $\hat{D}Y_{lm}$ is still an eigenstate of \hat{L}^2 with eigenvalue $l(l+1)$ and so must be a linear combination of the $2l+1$ spherical harmonics at that l , i.e.,

$$\hat{D}Y_{lm} = \sum_{m'} D_{m'm}^l Y_{lm'}.\tag{1.12}$$

The $D_{m'm}^l$ are the *Wigner D matrices* and for a given l they form a $(2l+1)$ -dimensional representation of the rotation group.

We can extract the D matrices using orthonormality of the spherical harmonics:

$$\begin{aligned}D_{m'm}^l &= \int d\hat{\mathbf{n}} Y_{lm'}^* e^{-i\alpha\hat{L}_z} e^{-i\beta\hat{L}_y} e^{-i\gamma\hat{L}_z} Y_{lm} \\ &= \int d\hat{\mathbf{n}} \left(e^{i\alpha\hat{L}_z} Y_{lm'} \right)^* e^{-i\beta\hat{L}_y} e^{-i\gamma\hat{L}_z} Y_{lm} \\ &= e^{-im'\alpha} \underbrace{\int d\hat{\mathbf{n}} Y_{lm'}^* e^{-i\beta\hat{L}_y} Y_{lm}}_{d_{m'm}^l(\beta)} e^{-im\gamma},\end{aligned}\tag{1.13}$$

where we used the fact that the \hat{L}_i are Hermitian. The last line defines the *reduced d matrices*, $d_{m'm}^l$, which depend only on β . With our phase conventions, the $d_{m'm}^l$ are real. Explicit expressions can be found in, for example, Brink & Satchler.

The rotation operators are unitary,

$$\hat{D}^\dagger \hat{D} = \hat{D} \hat{D}^\dagger = 1,\tag{1.14}$$

and hence the $D_{m'm}^l$ are unitary matrices (at fixed l):

$$\begin{aligned}\hat{D}^\dagger \hat{D} = 1 \quad \Rightarrow \quad \delta_{mm'} &= \int d\hat{\mathbf{n}} (\hat{D}Y_{lm})^* \hat{D}Y_{lm'} \\ &= \sum_{nn'} \int d\hat{\mathbf{n}} D_{nm}^{l*} Y_{ln}^* D_{n'm'}^l Y_{ln'} \\ &= \sum_n D_{nm}^{l*} D_{nm'}^l.\end{aligned}\tag{1.15}$$

The inverse rotation has Euler angles $-\gamma$, $-\beta$ and $-\alpha$, so, by unitarity,

$$\hat{D}^{-1}(\alpha, \beta, \gamma) = \hat{D}(-\gamma, -\beta, -\alpha) = \hat{D}^\dagger(\alpha, \beta, \gamma).\tag{1.16}$$

The rotation matrices with one m equal to zero are related to the spherical harmonics. Taking $\hat{\mathbf{n}}$ to have polar coordinates (θ, ϕ) , we can write $\hat{\mathbf{n}} = \hat{D}(\phi, \theta, 0)\hat{\mathbf{z}}$ so that

$$\begin{aligned}Y_{lm}(\hat{\mathbf{n}}) &= Y_{lm}(\hat{D}(\phi, \theta, 0)\hat{\mathbf{z}}) \\ &= [\hat{D}^{-1}(\phi, \theta, 0)Y_{lm}](\hat{\mathbf{z}}) \\ &= D_{mm'}^{l*}(\phi, \theta, 0)Y_{lm'}(\hat{\mathbf{z}}) \\ &= \sqrt{\frac{2l+1}{4\pi}} D_{m0}^{l*}(\phi, \theta, 0),\end{aligned}\tag{1.17}$$

where we have used $Y_{lm}(\hat{\mathbf{z}}) = \sqrt{(2l+1)/4\pi} \delta_{m0}$. It follows that

$$D_{m0}^l(\phi, \theta, 0) = \sqrt{\frac{4\pi}{2l+1}} Y_{lm}^*(\hat{\mathbf{n}}).\tag{1.18}$$

Under an active rotation of the function $f(\hat{\mathbf{n}})$, the multipoles transform as

$$\begin{aligned}f_{lm} &\rightarrow \int d\hat{\mathbf{n}} Y_{lm}^* \sum_{l'm'} f_{l'm'} \hat{D}Y_{l'm'} \\ &= \int d\hat{\mathbf{n}} Y_{lm}^* \sum_{l'm'n'} f_{l'm'} D_{n'm'}^l Y_{l'n'} \\ &= \sum_{m'} D_{mm'}^l f_{lm'}.\end{aligned}\tag{1.19}$$

Note the order of indices on the $D_{mm'}^l$ here, compared to Eq. (1.12).

1.2 Two-point function

Consider the two-point function $\langle f_{lm} f_{l'm'}^* \rangle$. If this is to be invariant under all rotations, we must have

$$\langle f_{lm} f_{l'm'}^* \rangle = \sum_{nn'} D_{mn}^l D_{m'n'}^{l'*} \langle f_{ln} f_{l'n'}^* \rangle.\tag{1.20}$$

This is only the case if

$$\langle f_{lm} f_{l'm'}^* \rangle = C_l \delta_{ll'} \delta_{mm'}, \quad (1.21)$$

since then the right-hand side of Eq. (1.20) is

$$\sum_{nn'} D_{mn}^l D_{m'n'}^{l'*} C_l \delta_{ll'} \delta_{nn'} = C_l \delta_{ll'} \underbrace{\sum_n D_{mn}^l D_{m'n'}^{l'*}}_{\delta_{mm'}} \quad (1.22)$$

where we used Eq. (1.15). The real quantity C_l is the *angular power spectrum* of $f(\hat{\mathbf{n}})$. Zero-mean Gaussian random fields are fully determined by their two-point function (i.e., their covariance) and hence their angular power spectrum.

We can also consider the two-point function in real space:

$$\begin{aligned} \langle f(\hat{\mathbf{n}}) f(\hat{\mathbf{n}}') \rangle &= \sum_{lm} \sum_{l'm'} \underbrace{\langle f_{lm} f_{l'm'}^* \rangle}_{C_l \delta_{ll'} \delta_{mm'}} Y_{lm}(\hat{\mathbf{n}}) Y_{l'm'}^*(\hat{\mathbf{n}}') \\ &= \sum_l C_l \sum_m Y_{lm}(\hat{\mathbf{n}}) Y_{lm}^*(\hat{\mathbf{n}}') \\ &= \sum_l \frac{2l+1}{4\pi} C_l P_l(\hat{\mathbf{n}} \cdot \hat{\mathbf{n}}') \\ &= C(\theta), \end{aligned} \quad (1.23)$$

where $\hat{\mathbf{n}} \cdot \hat{\mathbf{n}}' = \cos \theta$ and we used the addition theorem for spherical harmonics²,

$$\sum_m Y_{lm}(\hat{\mathbf{n}}) Y_{lm}^*(\hat{\mathbf{n}}') = \frac{2l+1}{4\pi} P_l(\hat{\mathbf{n}} \cdot \hat{\mathbf{n}}'). \quad (1.26)$$

²To derive the addition theorem, note that the left-hand side of Eq. (1.26) is rotationally invariant. To see this, consider back-rotating $\hat{\mathbf{n}}$ to $\hat{D}^{-1}\hat{\mathbf{n}}$, and similarly for $\hat{\mathbf{n}}'$. Using $[\hat{D}f](\hat{\mathbf{n}}) = f(\hat{D}^{-1}\hat{\mathbf{n}})$ for any function $f(\hat{\mathbf{n}})$, we have

$$\begin{aligned} \sum_m Y_{lm}(\hat{D}^{-1}\hat{\mathbf{n}}) Y_{lm}^*(\hat{D}^{-1}\hat{\mathbf{n}}') &= \sum_{mn'n'} D_{nm}^l Y_{ln}(\hat{\mathbf{n}}) D_{n'm'}^{l*} Y_{ln'}^*(\hat{\mathbf{n}}') \\ &= \sum_n Y_{ln}(\hat{\mathbf{n}}) Y_{ln}^*(\hat{\mathbf{n}}'), \end{aligned} \quad (1.24)$$

where we used unitarity of the rotation matrices in the last equality. It follows that the left-hand side of Eq. (1.26) can only depend on the rotationally-invariant angle between $\hat{\mathbf{n}}$ and $\hat{\mathbf{n}}'$. Taking $\hat{\mathbf{n}}'$ along the z -axis, the left-hand side of Eq. (1.26) reduces to

$$\sqrt{\frac{2l+1}{4\pi}} Y_{l0}(\theta, \phi) = \frac{2l+1}{4\pi} P_l(\cos \theta), \quad (1.25)$$

where we have used the result that the $m = 0$ spherical harmonics are proportional to Legendre polynomials.

It follows that the two-point correlation function depends only on the angle between the two points, as required by statistical isotropy. Note that the variance of the field is

$$C(0) = \sum_l \frac{2l+1}{4\pi} C_l \approx \int d \ln l \frac{l(l+1)C_l}{2\pi}. \quad (1.27)$$

It is conventional to plot $l(l+1)C_l/(2\pi)$ which we see is (almost) the contribution to the variance per log range in l . Finally, we note that we can invert the correlation function to get the power spectrum by using orthogonality of the Legendre polynomials:

$$C_l = 2\pi \int_{-1}^1 d \cos \theta C(\theta) P_l(\cos \theta). \quad (1.28)$$

1.3 Projections of 3D random fields

The CMB decoupled from matter around recombination. Roughly, the temperature anisotropies that we observe along different directions are a projection of conditions at that time over the *last-scattering*, a sphere centred on our current location with comoving radius $\chi_* \approx 13800$ Mpc equal to the distance light can have travelled from last-scattering to the present.

Projecting a 3D statistically-isotropic and homogeneous random field gives a statistically-isotropic 2D field on the sphere. To see this explicitly, consider projecting a random field $F(\mathbf{x})$ over a sphere of radius r , centred on the origin, to form $f(\hat{\mathbf{n}}) = F(r\hat{\mathbf{n}})$. Expanding $F(\mathbf{x})$ in Fourier modes, we have

$$\begin{aligned} f(\hat{\mathbf{n}}) &= \int \frac{d^3\mathbf{k}}{(2\pi)^{3/2}} F(\mathbf{k}) e^{ikr\hat{\mathbf{k}} \cdot \hat{\mathbf{n}}} \\ &= 4\pi \sum_{lm} i^l \left(\int \frac{d^3\mathbf{k}}{(2\pi)^{3/2}} F(\mathbf{k}) j_l(kr) Y_{lm}^*(\hat{\mathbf{k}}) \right) Y_{lm}(\hat{\mathbf{n}}), \end{aligned} \quad (1.29)$$

where we have used the *Rayleigh plane-wave expansion*

$$\begin{aligned} e^{i\mathbf{k} \cdot \mathbf{x}} &= \sum_l i^l (2l+1) j_l(kr) P_l(\hat{\mathbf{k}} \cdot \hat{\mathbf{n}}) \\ &= 4\pi \sum_{lm} i^l j_l(kr) Y_{lm}^*(\hat{\mathbf{k}}) Y_{lm}(\hat{\mathbf{n}}). \end{aligned} \quad (1.30)$$

Here, $r = |\mathbf{x}|$ and $j_l(kr)$ are the spherical Bessel functions. Extracting the spherical multipoles of $f(\hat{\mathbf{n}})$ from Eq. (1.29), we have

$$f_{lm} = 4\pi i^l \int \frac{d^3\mathbf{k}}{(2\pi)^{3/2}} F(\mathbf{k}) j_l(kr) Y_{lm}^*(\hat{\mathbf{k}}). \quad (1.31)$$

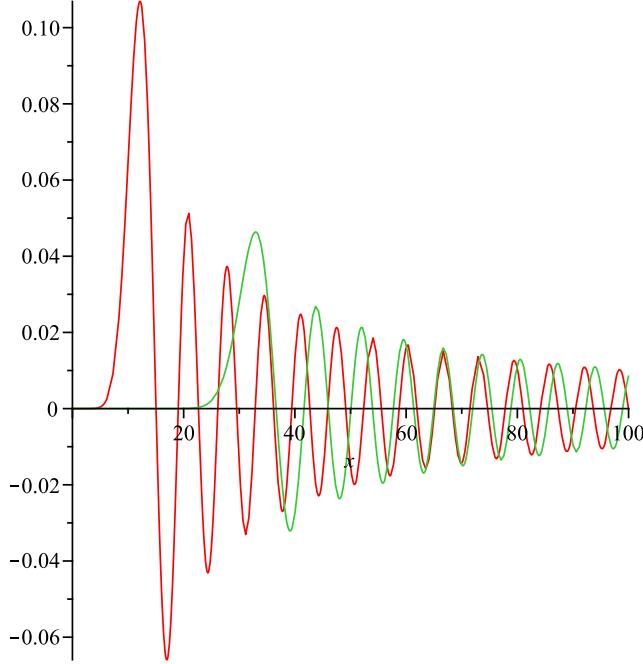


Figure 2: Spherical Bessel functions $j_l(x)$ for $l = 10$ (red) and $l = 30$ (green).

The two-point correlator of the f_{lm} is then

$$\begin{aligned}
 \langle f_{lm} f_{l'm'}^* \rangle &= (4\pi)^2 i^{l-l'} \int \frac{d^3\mathbf{k}}{(2\pi)^{3/2}} \frac{d^3\mathbf{k}'}{(2\pi)^{3/2}} \underbrace{\langle F(\mathbf{k}) F^*(\mathbf{k}') \rangle}_{\frac{2\pi^2}{k^3} \mathcal{P}_F(k) \delta^{(3)}(\mathbf{k}-\mathbf{k}')} j_l(kr) j_{l'}(k'r) Y_{lm}^*(\hat{\mathbf{k}}) Y_{l'm'}(\hat{\mathbf{k}}') \\
 &= 4\pi i^{l-l'} \left(\int d \ln k \mathcal{P}_F(k) j_l(kr) j_{l'}(kr) \right) \left(\int d \hat{\mathbf{k}} Y_{lm}^*(\hat{\mathbf{k}}) Y_{l'm'}(\hat{\mathbf{k}}) \right) \\
 &= 4\pi \delta_{ll'} \delta_{mm'} \int d \ln k \mathcal{P}_F(k) j_l^2(kr).
 \end{aligned} \tag{1.32}$$

It follows that $f(\hat{\mathbf{n}})$ is a statistically-isotropic field with angular power spectrum

$$C_l = 4\pi \int d \ln k \mathcal{P}_F(k) j_l^2(kr). \tag{1.33}$$

The spherical Bessel functions peak sharply at $kr = l$ for large l (see Fig. 2), and are $O(kr)^l$ for small arguments. This means that the observed multipoles l mainly probe spatial structure in the 3D field $F(\mathbf{x})$ with wavenumber $k \sim l/r$. The smallness of the spherical Bessel functions of high l for small arguments reflects the fact that no small-scale anisotropy is generated by projecting a plane wave over a sphere of radius much less than the wavelength. The oscillatory tails of the Bessel functions mean that

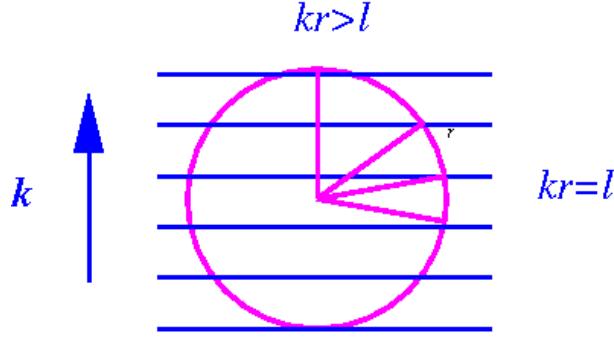


Figure 3: Projection of a plane-wave component of a scalar field $F(\mathbf{x})$ at distance r . Along lines of sight perpendicular to the wavevector, the Fourier component with comoving wavelength $2\pi/k$ projects to angular structure with angular wavelength $2\pi/(kr)$, which corresponds to a multipole $l = kr$. For lines of sight closer to the wavenumber, the angular wavelength of the projection is larger corresponding to smaller l . Conversely, along such lines of sight, angular structure at multipole l is sourced by shorter wavelength perturbations corresponding to the oscillatory tails of the spherical Bessel functions at large argument.

some power from a given k does also enter larger-scale anisotropies. Physically, this arises from Fourier modes that are not aligned with their wavevector perpendicular to the line of sight (see Fig. 3).

1.4 Three-point function

In non-Gaussian theories, we are led to consider functions on the sphere, $f(\hat{\mathbf{n}})$, that are non-Gaussian zero-mean random fields. The simplest statistic to characterise non-Gaussianity is the *3-point* function $\langle f_{l_1 m_1} f_{l_2 m_2} f_{l_3 m_3} \rangle$, since this necessarily vanishes for a zero-mean Gaussian field. What restrictions does statistical isotropy place on the form of the 3-point function? The answer, as we shall verify shortly, is that

$$\langle f_{l_1 m_1} f_{l_2 m_2} f_{l_3 m_3} \rangle = B_{l_1 l_2 l_3} \left(\begin{array}{ccc} l_1 & l_2 & l_3 \\ m_1 & m_2 & m_3 \end{array} \right), \quad (1.34)$$

where $B_{l_1 l_2 l_3}$ is the *angular bispectrum*, and the symbol in brackets is the Wigner $3j$ symbol, related to the vector-addition coefficients that are familiar to you from coupling of angular momenta in quantum mechanics (see below). We shall have much more to say about the bispectrum at the end of the course, when we discuss non-Gaussian perturbations from inflation and how to search for non-Gaussianity in the CMB. Higher-point functions can also be considered; for example the 4-point function is very important for discussing the effect of weak gravitational lensing on the CMB.

We now develop some of the formalism required to establish the rotational invariance of Eq. (1.34). Along the way we shall introduce a number of useful results for products of spherical harmonics that we shall use later in our discussion of CMB physics.

1.4.1 Coupling of two angular momenta (non-examinable)

The spherical harmonics are the position-space representation of the orbital angular momentum eigenstates $|lm\rangle$. For generality, in this section we consider general angular momentum states $|jm\rangle$ where j need not be an integer (but $2j+1$ is).

Consider a system with two parts (e.g., two particles) each with its own angular momentum. We can consider composite states that are tensor products of the angular momentum eigenstates for each part, which we write as

$$|j_1 j_2 m_1 m_2\rangle = |j_1 m_1\rangle \otimes |j_2 m_2\rangle. \quad (1.35)$$

These are eigenstates of \hat{J}_1^2 , $\hat{J}_{1,z}$, \hat{J}_2^2 and $\hat{J}_{2,z}$. They are also eigenstates of the z -component of the total angular momentum, $\mathbf{J} = \mathbf{J}_1 + \mathbf{J}_2$, with eigenvalue $m_1 + m_2$ (taking $\hbar = 1$). However, they are not generally eigenstates of \hat{J}^2 . It is often convenient to work with linear combinations of the $(2j_1+1)(2j_2+1)$ states that are eigenstates of \hat{J}^2 and \hat{J}_z (and \hat{J}_1^2 and \hat{J}_2^2). We write such states as $|j_1 j_2 JM\rangle$, so that

$$\begin{aligned} \hat{J}^2 |j_1 j_2 JM\rangle &= J(J+1) |j_1 j_2 JM\rangle \\ \hat{J}_z |j_1 j_2 JM\rangle &= M |j_1 j_2 JM\rangle \\ \hat{J}_1^2 |j_1 j_2 JM\rangle &= j_1(j_1+1) |j_1 j_2 JM\rangle \\ \hat{J}_2^2 |j_1 j_2 JM\rangle &= j_2(j_2+1) |j_1 j_2 JM\rangle. \end{aligned} \quad (1.36)$$

The allowed values of J range from $|j_1 - j_2|$ to $j_1 + j_2$ in integer steps. In this way, we account for all $(2j_1+1)(2j_2+1)$ of the original states.

We can express the $|j_1 j_2 JM\rangle$ as linear combinations of the $|j_1 j_2 m_1 m_2\rangle$ using

$$|j_1 j_2 JM\rangle = \underbrace{\sum_{m_1 m_2} |j_1 j_2 m_1 m_2\rangle \langle j_1 j_2 m_1 m_2|}_{\text{Resolution of identity}} |j_1 j_2 JM\rangle. \quad (1.37)$$

The *vector-addition* (or Clebsch–Gordan) coefficient $\langle j_1 j_2 m_1 m_2 | j_1 j_2 JM \rangle$ is real so that

$$\langle j_1 j_2 m_1 m_2 | j_1 j_2 JM \rangle = \langle j_1 j_2 JM | j_1 j_2 m_1 m_2 \rangle. \quad (1.38)$$

It vanishes for $m_1 + m_2 \neq M$. The inverse relation is

$$\begin{aligned} |j_1 j_2 m_1 m_2\rangle &= \sum_{JM} |j_1 j_2 JM\rangle \langle j_1 j_2 JM | j_1 j_2 m_1 m_2 \rangle \\ &= \sum_{JM} \langle j_1 j_2 m_1 m_2 | j_1 j_2 JM \rangle |j_1 j_2 JM\rangle. \end{aligned} \quad (1.39)$$

The sum over M here is redundant since the vector-addition coefficient forces $M = m_1 + m_2$.

Consider rotating the composite system. The composite states $|j_1 j_2 m_1 m_2\rangle$ transform as

$$\hat{D}|j_1 j_2 m_1 m_2\rangle = \sum_{n_1 n_2} D_{n_1 m_1}^{l_1} D_{n_2 m_2}^{l_2} |j_1 j_2 n_1 n_2\rangle, \quad (1.40)$$

i.e., they transform under the $(2j_1 + 1)(2j_2 + 1)$ product representation of the rotation group, denoted $D_{j_1} \otimes D_{j_2}$. However, the composite rotation operator can also be written as³

$$\hat{D}(\alpha, \beta, \gamma) = e^{-i\alpha \hat{J}_z} e^{-i\beta \hat{J}_y} e^{-i\gamma \hat{J}_z}, \quad (1.41)$$

so that the $2J + 1$ states $|j_1 j_2 J M\rangle$ with given J transform irreducibly as

$$\hat{D}|j_1 j_2 J M\rangle = \sum_N D_{NM}^J |j_1 j_2 J N\rangle. \quad (1.42)$$

In this way, the product representation has been written in terms of its irreducible components, which we can write as

$$D_{j_1} \otimes D_{j_2} = \sum_{J=|j_1-j_2|}^{j_1+j_2} D_J, \quad (1.43)$$

and the states $|j_1 j_2 J M\rangle$ for each J span the invariant subspaces of the reduced representation.

The rotation matrix elements for the $|j_1 j_2 J M\rangle$ states can be related to those for the $|j_1 m_1\rangle$ and $|j_2 m_2\rangle$ states using Eq. (1.37); for a given rotation

$$\begin{aligned} D_{MN}^J &= \langle j_1 j_2 J M | \hat{D} | j_1 j_2 J N \rangle \\ &= \sum_{m_1 m_2 n_1 n_2} \langle j_1 j_2 J M | j_1 j_2 m_1 m_2 \rangle \underbrace{\langle j_1 j_2 m_1 m_2 | \hat{D} | j_1 j_2 n_1 n_2 \rangle}_{D_{m_1 n_1}^{j_1} D_{m_2 n_2}^{j_2}} \langle j_1 j_2 n_1 n_2 | j_1 j_2 J N \rangle \\ &= \sum_{m_1 m_2 n_1 n_2} \langle j_1 j_2 J M | j_1 j_2 m_1 m_2 \rangle D_{m_1 n_1}^{j_1} D_{m_2 n_2}^{j_2} \langle j_1 j_2 n_1 n_2 | j_1 j_2 J N \rangle. \end{aligned} \quad (1.44)$$

We also have

$$\begin{aligned} D_{m_1 n_1}^{j_1} D_{m_2 n_2}^{j_2} &= \langle j_1 j_2 m_1 m_2 | \hat{D} | j_1 j_2 n_1 n_2 \rangle \\ &= \sum_{J M J' N} \langle j_1 j_2 m_1 m_2 | j_1 j_2 J M \rangle \underbrace{\langle j_1 j_2 J M | \hat{D} | j_1 j_2 J' N \rangle}_{D_{M N}^J \delta_{J J'}} \langle j_1 j_2 J' N | j_1 j_2 n_1 n_2 \rangle \\ &= \sum_{J M N} \langle j_1 j_2 m_1 m_2 | j_1 j_2 J M \rangle D_{M N}^J \langle j_1 j_2 J N | j_1 j_2 n_1 n_2 \rangle, \end{aligned} \quad (1.45)$$

³Recall that $\hat{\mathbf{J}}_1$ and $\hat{\mathbf{J}}_2$ commute.

where the sum over M and N is redundant.

It is often convenient to work with the Wigner $3j$ symbol rather than the vector-addition coefficients. These are related by⁴

$$\sqrt{2j_3 + 1} \begin{pmatrix} j_1 & j_2 & j_3 \\ m_1 & m_2 & m_3 \end{pmatrix} = (-1)^{j_1 - j_2 - m_3} \langle j_1 j_2 m_1 m_2 | j_1 j_2 j_3 - m_3 \rangle. \quad (1.46)$$

The $3j$ symbol vanishes for $m_1 + m_2 + m_3 \neq 0$, and has the same triangle restrictions on the j_i as the vector-addition coefficient. It is rather more symmetrical than the vector-addition coefficients; in particular it is invariant under cyclic permutation of columns and is multiplied by $(-1)^{j_1 + j_2 + j_3}$ under non-cyclic ones:

$$\begin{pmatrix} j_1 & j_2 & j_3 \\ m_1 & m_2 & m_3 \end{pmatrix} = \begin{pmatrix} j_3 & j_1 & j_2 \\ m_3 & m_1 & m_2 \end{pmatrix} = (-1)^{j_1 + j_2 + j_3} \begin{pmatrix} j_1 & j_3 & j_2 \\ m_1 & m_3 & m_2 \end{pmatrix} \quad \text{etc.} \quad (1.47)$$

Moreover, under $m_i \rightarrow -m_i$,

$$\begin{pmatrix} j_1 & j_2 & j_3 \\ -m_1 & -m_2 & -m_3 \end{pmatrix} = (-1)^{j_1 + j_2 + j_3} \begin{pmatrix} j_1 & j_2 & j_3 \\ m_1 & m_2 & m_3 \end{pmatrix}. \quad (1.48)$$

In terms of the $3j$ symbol, Eq. (1.45) becomes

$$D_{m_1 n_1}^{j_1} D_{m_2 n_2}^{j_2} = \sum_{JMN} (2J+1) \begin{pmatrix} j_1 & j_2 & J \\ m_1 & m_2 & M \end{pmatrix} \begin{pmatrix} j_1 & j_2 & J \\ n_1 & n_2 & N \end{pmatrix} D_{MN}^{J*}. \quad (1.49)$$

If we make use of

$$D_{m0}^l(\phi, \theta, 0) = \sqrt{\frac{4\pi}{2l+1}} Y_{lm}^*(\hat{\mathbf{n}}), \quad (1.50)$$

we can use Eq. (1.49) to reduce a product of spherical harmonics to a sum over spherical harmonics:

$$\begin{aligned} Y_{l_1 m_1}(\hat{\mathbf{n}}) Y_{l_2 m_2}(\hat{\mathbf{n}}) &= \sum_{LM} \sqrt{\frac{(2l_1+1)(2l_2+1)(2L+1)}{4\pi}} \begin{pmatrix} l_1 & l_2 & L \\ m_1 & m_2 & M \end{pmatrix} \\ &\quad \times \begin{pmatrix} l_1 & l_2 & L \\ 0 & 0 & 0 \end{pmatrix} Y_{LM}^*(\hat{\mathbf{n}}), \end{aligned} \quad (1.51)$$

a result that we shall make repeated use of later. For now, we note that Eq. (1.51) implies that

$$\int d\hat{\mathbf{n}} Y_{l_1 m_1} Y_{l_2 m_2} Y_{l_3 m_3} = \sqrt{\frac{(2l_1+1)(2l_2+1)(2l_3+1)}{4\pi}} \begin{pmatrix} l_1 & l_2 & l_3 \\ m_1 & m_2 & m_3 \end{pmatrix} \begin{pmatrix} l_1 & l_2 & l_3 \\ 0 & 0 & 0 \end{pmatrix}. \quad (1.52)$$

⁴Note that $j_1 - j_2 - m_3$ is always an integer, even when dealing with half-integer angular momenta, i.e., spin. The same is true for $j_1 + j_2 + j_3$.

This useful result for the integral of three spherical harmonics is known as the *Gaunt integral*. It appears naturally when computing the angular bispectrum from the projection on the sphere of a 3D non-Gaussian field (see Question 1 on Examples II).

The Gaunt integral is rotationally invariant, i.e., if we actively rotate the three $Y_{l_i m_i}$ and then perform the integral we get the same result. This follows since $[\hat{D} Y_{lm}](\hat{\mathbf{n}}) = Y_{lm}(\hat{D}^{-1}\hat{\mathbf{n}})$, and the back-rotation of the argument does not alter the integral over all angles (the Jacobian relating $d\hat{\mathbf{n}}$ and $d\hat{D}^{-1}\hat{\mathbf{n}}$ is unity). Invariance can also be seen easily from the important result,

$$\sum_{n_1 n_2 n_3} D_{n_1 m_1}^{j_1} D_{n_2 m_2}^{j_2} D_{n_3 m_3}^{j_3} \begin{pmatrix} j_1 & j_2 & j_3 \\ n_1 & n_2 & n_3 \end{pmatrix} = \begin{pmatrix} j_1 & j_2 & j_3 \\ m_1 & m_2 & m_3 \end{pmatrix}, \quad (1.53)$$

which follows from unitarity of \hat{D} :

$$\begin{aligned} \langle j_1 j_2 m_1 m_2 | j_1 j_2 JM \rangle &= \langle j_1 j_2 m_1 m_2 | \hat{D}^\dagger \hat{D} | j_1 j_2 JM \rangle \\ &= \sum_{n_1 n_2 N} D_{n_1 m_1}^{j_1 *} D_{n_2 m_2}^{j_2 *} \langle j_1 j_2 n_1 n_2 | j_1 j_2 JN \rangle D_{NM}^J. \end{aligned} \quad (1.54)$$

Equation (1.53) tells us how to construct rotationally-invariant combinations of three spherical multipoles (indeed, the $3j$ symbol was first introduced by Wigner to describe the coupling of three angular momenta to get zero resultant):

$$\begin{aligned} \sum_{m_i} f_{l_1 m_1} f_{l_2 m_2} f_{l_3 m_3} \begin{pmatrix} l_1 & l_2 & l_3 \\ m_1 & m_2 & m_3 \end{pmatrix} &\mapsto \sum_{m_i} (\hat{D}f)_{l_1 m_1} (\hat{D}f)_{l_2 m_2} (\hat{D}f)_{l_3 m_3} \begin{pmatrix} l_1 & l_2 & l_3 \\ m_1 & m_2 & m_3 \end{pmatrix} \\ &= \sum_{m_i n_i} D_{m_1 n_1}^{l_1} D_{m_2 n_2}^{l_2} D_{m_3 n_3}^{l_3} f_{l_1 n_1} f_{l_2 n_2} f_{l_3 n_3} \\ &\quad \times \begin{pmatrix} l_1 & l_2 & l_3 \\ m_1 & m_2 & m_3 \end{pmatrix} \\ &= \sum_{n_i} f_{l_1 n_1} f_{l_2 n_2} f_{l_3 n_3} \begin{pmatrix} l_1 & l_2 & l_3 \\ n_1 & n_2 & n_3 \end{pmatrix}. \end{aligned} \quad (1.55)$$

Equation (1.53) also establishes the rotational invariance of the right-hand side of the three-point function in Eq. (1.34).

1.5 Cosmic variance

Theory does not predict the exact form that cosmological fields should take. Instead, a given model makes definite predictions for the probability distribution from which the one realisation of the fields that we can observe is drawn. Any estimate that we make of properties of the underlying probability distribution, such as its power spectrum,

will inevitably have some random error called *cosmic variance* due to us attempting to estimate ensemble-averaged quantities from a single realisation. Nearly all inferences that we make in observational cosmology have a contribution to their error budget from cosmic variance. This may not currently be the dominant error if the measurements are noisy, but the cosmic-variance limit is an aspiration for all cosmological observations.

As a concrete example, consider estimating the angular power spectrum of a zero-mean random field $f(\hat{\mathbf{n}})$. Suppose that we have full-sky, noise-free observations of $f(\hat{\mathbf{n}})$. We can extract the spherical multipoles f_{lm} and use these to form the following estimator for C_l :

$$\hat{C}_l = \frac{1}{2l+1} \sum_m f_{lm} f_{lm}^*. \quad (1.56)$$

This is unbiased, i.e., $\langle \hat{C}_l \rangle = C_l$, and the sum over m ensures that we use all the available information in the $2l+1$ independent degrees of freedom in the multipoles with given l . However, the estimator has a non-zero variance. For a Gaussian random field, we can calculate this variance using Wick's theorem (see below) as follows:

$$\begin{aligned} \text{var}(\hat{C}_l) &= \langle \hat{C}_l \hat{C}_l \rangle - \langle \hat{C}_l \rangle^2 \\ &= \frac{1}{(2l+1)^2} \sum_{mm'} \langle f_{lm} f_{lm}^* f_{lm'} f_{lm'}^* \rangle - C_l^2 \\ &= \frac{1}{(2l+1)^2} \sum_{mm'} (\langle f_{lm} f_{lm'} \rangle \langle f_{lm}^* f_{lm'}^* \rangle + \langle f_{lm} f_{lm'}^* \rangle \langle f_{lm}^* f_{lm'} \rangle) \\ &= \frac{2}{2l+1} C_l^2. \end{aligned} \quad (1.57)$$

This irreducible cosmic variance arises from having only $2l+1$ independent modes at each multipole l with which to estimate the variance of their distribution.

Current measurements of the angular power spectrum of the CMB temperature anisotropies from Planck are shown in Fig. 4. The error bars on the data points include both cosmic variance and the effect of noise in the measurement, with the former dominating up to multipoles $l \sim 2000$.

*Wick's theorem.*⁵ For a zero-mean Gaussian random field, the N -point functions for even N can be reduced to sums of products of two-point functions, where the sum is over possible pairs. For example,

$$\langle f_i f_j f_k f_l \rangle = \langle f_i f_j \rangle \langle f_k f_l \rangle + \langle f_i f_k \rangle \langle f_j f_l \rangle + \langle f_i f_l \rangle \langle f_j f_k \rangle. \quad (1.58)$$

The N -point functions for odd N vanish.

⁵The theorem is due to Isserlis, but most cosmologists refer to it as Wick's theorem after the analogous result in quantum field theory (which is due to Wick).

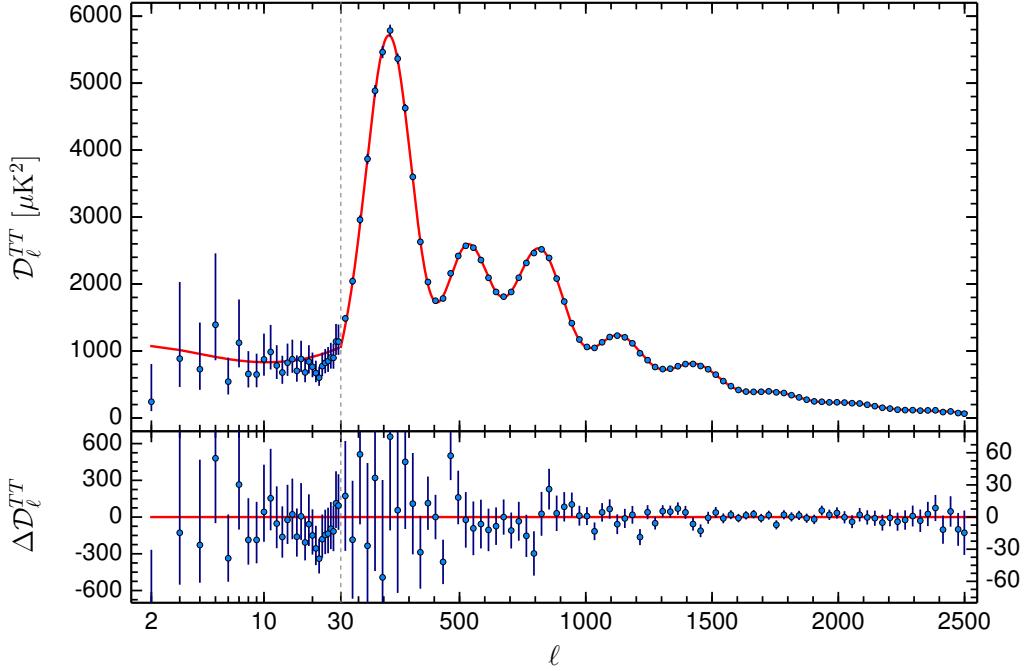


Figure 4: Measurements of the angular power spectrum of the CMB temperature anisotropies from Planck, where $\mathcal{D}_l \equiv l(l+1)C_l/(2\pi)$. The red line is a six-parameter Λ CDM model that is fit to the data points; the differences between the data and the fit are shown in the lower panel. Note the remarkable agreement of the measurements and the theory.

We can prove Wick's theorem easily by considering the *characteristic function* of the multivariate Gaussian probability density function. This is simply the Fourier transform, and for an N -dimensional discrete distribution is given by

$$\begin{aligned} \langle e^{i\mathbf{J}\cdot\mathbf{f}} \rangle &= \frac{1}{\sqrt{\det(2\pi\mathbf{C})}} \int d^N\mathbf{f} e^{i\mathbf{J}\cdot\mathbf{f}} e^{-\mathbf{f}^T\mathbf{C}^{-1}\mathbf{f}/2} \\ &= e^{-\mathbf{J}^T\mathbf{C}\mathbf{J}/2}. \end{aligned} \quad (1.59)$$

The moments of the distribution can be easily obtained from derivatives of the characteristic function:

$$\langle f_{i_1} f_{i_2} \dots f_{i_n} \rangle = (-i)^n \left. \frac{\partial}{\partial J_{i_1}} \frac{\partial}{\partial J_{i_2}} \dots \frac{\partial}{\partial J_{i_n}} \langle e^{i\mathbf{J}\cdot\mathbf{f}} \rangle \right|_{\mathbf{J}=0}. \quad (1.60)$$

Evaluating this for the characteristic function of Gaussian distribution with $n = 2N$, we have

$$\langle f_{i_1} f_{i_2} \dots f_{i_{2N}} \rangle = \frac{1}{2^N N!} \underbrace{(C_{i_1 i_2} C_{i_3 i_4} \dots C_{i_{2N-1} i_{2N}} + \text{perms.})}_{(2N)! \text{ terms}}. \quad (1.61)$$

The total number of ways of choosing N pairs from $2N$ indices is $(2N)!/(2^N N!)$, hence the

right-hand side of Eq. (1.61) is equivalent to a sum of products of 2-point functions over all pairs.

The same results holds for Gaussian *fields*; the proof is similar but we must work with a characteristic *functional* and its functional derivatives.

2 Relativistic kinetic theory and the Boltzmann equation

Kinetic theory provides a seamless description of the CMB from the tight-coupled era, where collisions suppress anisotropy in the CMB, through to the free-streaming era, where collisions are negligible. In this section, we develop relativistic kinetic theory and derive the Boltzmann equation for the CMB to first order in perturbations about a spatially-flat Friedmann–Robertson–Walker universe.

Our conventions for the perturbed spacetime metric will mostly follow those in the *Cosmology course*:

$$ds^2 = a^2(\eta) \{ -(1 + 2A)d\eta^2 + 2B_i dx^i d\eta + [(1 + 2C)\delta_{ij} + 2E_{ij}] dx^i dx^j \} , \quad (2.1)$$

where E_{ij} is trace-free. Note that we use η for conformal time rather than τ (since the latter is commonly used for both the optical depth and proper time). The scale factor is $a(\eta)$. We raise and lower Latin indices on spatial vectors (like B_i) and tensors (like E_{ij}) with δ_{ij} , for example, $B^i \equiv \delta^{ij} B_j$.

2.1 Distribution function

The CMB photons can be described by a *one-particle distribution function* $f(x^\mu, p^\mu)$, which is a function of the spacetime position x^μ and 4-momentum p^μ of a photon. We shall usually parameterise the 4-momentum in terms of components relative to an *orthonormal tetrad* $(E_0)^\mu$ and $(E_i)^\mu$ ($i = 1, 2, 3$), where

$$g_{\mu\nu}(E_0)^\mu(E_0)^\nu = -1, \quad g_{\mu\nu}(E_0)^\mu(E_i)^\nu = 0, \quad g_{\mu\nu}(E_i)^\mu(E_j)^\nu = \delta_{ij} . \quad (2.2)$$

Note that the lower index inside the bracket is a label for the particular tetrad vector. We use the following explicit form for the tetrad vectors, correct to first order in the metric perturbations:

$$(E_0)^\mu = a^{-1}(1 - A)\delta_0^\mu \quad \text{and} \quad (E_i)^\mu = a^{-1} [B_i\delta_0^\mu + (1 - C)\delta_i^\mu - E_i^j\delta_j^\mu] . \quad (2.3)$$

Note that in the absence of perturbations, these are simply along the coordinate directions.

Exercise: verify that the tetrad given in Eq. (2.3) satisfies the conditions in Eq. (2.2).

We write

$$p^\mu = E(E_0)^\mu + p^i(E_i)^\mu, \quad (2.4)$$

where E is the energy of the photon as measured by an observer whose 4-velocity is $(E_0)^\mu$ and p^i are the tetrad components of the 3-momentum⁶. We shall generally write the spatial tetrad components in 3-vector form, e.g., \mathbf{p} for the 3-momentum. Since the 4-momentum of a massless particle is null, we have $E = |\mathbf{p}|$.

The one-particle distribution function is defined such that the number of photons in a proper 3-volume element $d^3\mathbf{x}$ and with 3-momentum in a proper element $d^3\mathbf{p}$ is $f d^3\mathbf{x} d^3\mathbf{p}$. It can therefore be thought of as a density of photons in *phase space*. The key properties of the distribution function are as follows:

- f is Lorentz invariant; and
- f is conserved along the photon path in phase space in the absence of scattering (Liouville's theorem).

2.1.1 Lorentz invariance of the distribution function

We can establish Lorentz invariance of f most easily by considering the case of massive particles, with rest mass m , and then taking the limit as $m \rightarrow 0$. As seen by some observer, we isolate those particles that at some instant occupy a proper 3-volume $d^3\mathbf{x}$ and have 3-momentum in the range $d^3\mathbf{p}$ centred around \mathbf{p} . Some other observer will see these same particles as occupying a different volume, $d^3\mathbf{x}'$, and having a different spread of 3-momenta, $d^3\mathbf{p}'$. What we shall now show is that the *product* $d^3\mathbf{x} d^3\mathbf{p}$ is the same for both observers. This will establish invariance of the phase-space volume and hence, since all observers agree about the number of particles, invariance of the phase-space density.

We start with the momentum element. In the Part-III course *Quantum Field Theory*, you will have seen that $d^3\mathbf{p}/E(\mathbf{p})$ is a Lorentz-invariant momentum element. This

⁶We use a hat on indices to distinguish the tetrad components of tensors from their components in the coordinate basis.

follows most neatly by noting that the covariant integral over 4-momenta restricted to be on-shell (i.e., satisfying $p^\mu p_\mu = -m^2$) and in the forward light cone ($p^0 > 0$) is

$$\delta(p^\mu p_\mu + m^2)\Theta(p^0)\sqrt{-g}d^4p = \delta(E^2 - \mathbf{p}^2 - m^2)\Theta(E)dEd^3\mathbf{p}, \quad (2.5)$$

where $\Theta(x)$ is the Heaviside function that is unity for $x \geq 0$ and zero otherwise, and $-g$ is (minus) the determinant of the metric. On the right-hand side of Eq. (2.5) we have expressed the 4-momentum in terms of its tetrad components, so that $-g \rightarrow 1$. If we think of the δ -function on the right as a function of E , we have

$$\delta(E^2 - \mathbf{p}^2 - m^2)\Theta(E) = \frac{\delta(E - \sqrt{\mathbf{p}^2 + m^2})}{2E(\mathbf{p})}, \quad (2.6)$$

where the Heaviside function selects only the root of $E^2 - \mathbf{p}^2 - m^2 = 0$ with positive E . Finally, integrating over dE , we have

$$\delta(p^\mu p_\mu + m^2)\Theta(p^0)\sqrt{-g}d^4p = \frac{d^3\mathbf{p}}{2E(\mathbf{p})}. \quad (2.7)$$

Since the left-hand side is Lorentz-invariant, the same is true of the right-hand side.

For the transformation of the volume element, we consider those same particles selected above from the perspective of their rest frame. If the 3-volume element in the rest frame is V_0 , in some other frames S and S' in which the particles have Lorentz factors γ and γ' , the volumes in these frames are (by length contraction)

$$d^3\mathbf{x} = V_0/\gamma \quad \text{and} \quad d^3\mathbf{x}' = V_0/\gamma'. \quad (2.8)$$

It follows that $\gamma d^3\mathbf{x} = \gamma' d^3\mathbf{x}'$ and, since $E = \gamma m$ and $E' = \gamma' m$, we see that $E(\mathbf{p})d^3\mathbf{x}$ is Lorentz invariant. Finally, multiplying with the invariant $d^3\mathbf{p}/E(\mathbf{p})$, we see that the phase-space element $d^3\mathbf{p}d^3\mathbf{x}$ is Lorentz invariant.

2.1.2 Liouville's theorem

To establish that the distribution function is conserved along the photon path in phase space if there is no scattering, we shall show that the Lorentz-invariant phase-space element is conserved for geodesic (i.e., free) motion. To do so, it is convenient to work temporarily with a $3 + 1$ splitting, so that at any time t (defined by some foliation of spacetime) the photon is characterised by its spatial coordinates in some chart, $x^i(t)$, and the three spatial covariant components of the 4-momentum, $p_i(t)$. As usual, the 0th component of the 4-momentum is determined by the on-shell condition,

$$g^{00}(p_0)^2 + 2g^{0i}p_0p_i + g^{ij}p_ip_j = 0. \quad (2.9)$$

The reason for using p_i is that these components are conjugate to x^i , with the dynamics determined by Hamilton's equations

$$\frac{dx^i}{dt} = \frac{\partial H}{\partial p_i} \quad \text{and} \quad \frac{dp_i}{dt} = -\frac{\partial H}{\partial x^i}, \quad (2.10)$$

where the Hamiltonian should be thought of as a function of t , x^i and p_i . As we show below, a suitable Hamiltonian is

$$H(x^i, p_i, t) = -p_0 = \frac{g^{0i} p_i}{g^{00}} + \left[\left(\frac{g^{0i} p_i}{g^{00}} \right)^2 - \frac{g^{ij} p_i p_j}{g^{00}} \right]^{1/2}, \quad (2.11)$$

where the second equality follows from Eq. (2.9). Under a *time-independent* reparameterisation of the spatial coordinates, $x^i \rightarrow \tilde{x}^i$, the conjugate momentum transforms to

$$\tilde{p}_i = \frac{\partial x^j}{\partial \tilde{x}^i} p_j, \quad (2.12)$$

and the Hamiltonian is invariant.

In the $3+1$ splitting, the phase space variables are x^i and p_i . The (Lorentz-invariant) proper phase space volume can be expressed as $d^3x^i d^3p_i$. By construction, this is invariant under time-independent reparameterisations $x^i \rightarrow \tilde{x}^i$. If we consider an observer with 4-velocity along the normal to the constant- t surface through some event, their instantaneous rest space is tangent to the surface. Since the spatial metric g_{ij} determines the intrinsic geometry of the constant- t surfaces, the proper spatial volume element for the observer is

$$\text{Proper volume} = \sqrt{\det(g_{ij})} d^3x^i. \quad (2.13)$$

Similarly, the proper 3-momentum element for the observer is

$$\text{Proper momentum volume} = \frac{1}{\sqrt{\det(g_{ij})}} d^3p_i. \quad (2.14)$$

The product of these is simply $d^3x^i d^3p_i$. Conservation of the proper phase space volume then follows directly from this being a Hamiltonian system⁷.

⁷Consider a time increment dt in which a separation (dx^i, dp_i) in phase space maps to a new separation. For example, in a two-dimensional phase space (for simplicity), to first order in dt

$$\begin{pmatrix} dx \\ dp \end{pmatrix} \rightarrow \begin{pmatrix} dx \\ dp \end{pmatrix} + dt \begin{pmatrix} \frac{\partial^2 H}{\partial p \partial x} & \frac{\partial^2 H}{\partial p^2} \\ -\frac{\partial^2 H}{\partial x^2} & -\frac{\partial^2 H}{\partial x \partial p} \end{pmatrix} \begin{pmatrix} dx \\ dp \end{pmatrix}.$$

This is a linear transformation that is a volume preserving since the trace of the part proportional to dt vanishes. The generalisation of this argument to six dimensions is straightforward.

Hamiltonian for a free, point particle. We now show that Hamilton's equations of motion, with $H = -p_0$, give the correct geodesic equation of motion for freely propagating photons. In terms of an affine parameter λ , with $p^\mu = dx^\mu/d\lambda$, the geodesic equation is

$$dp^\mu/d\lambda = -\Gamma_{\nu\rho}^\mu p^\nu p^\rho. \quad (2.15)$$

Inserting the explicit expression for the Christoffel symbols, and lowering the index gives the alternative form

$$\frac{dp_\mu}{d\lambda} = \frac{1}{2} \frac{\partial g_{\nu\rho}}{\partial x^\mu} p^\nu p^\rho. \quad (2.16)$$

We want to show that this, plus $p^\mu = dx^\mu/d\lambda$, is equivalent to Hamilton's equations (2.10).

We start with $dx^i/dt = \partial H/\partial p_i$, with $H = -p_0$. Recall that we have to think of p_0 here as being a function of x^i , p_i and t . If we differentiate the on-shell condition, Eq. (2.9), with respect to p_i (at fixed x^i and t), we find

$$\begin{aligned} & \underbrace{(g^{00}p_0 + g^{0j}p_j)}_{p^0} \frac{\partial p_0}{\partial p_i} + \underbrace{(g^{0i}p_0 + g^{ij}p_j)}_{p^i} = 0 \\ & \Rightarrow \frac{\partial p_0}{\partial p_i} = -\frac{p^i}{p^0}. \end{aligned} \quad (2.17)$$

It follows that

$$\frac{dx^i}{dt} = \frac{p^i}{p^0}. \quad (2.18)$$

These are the spatial components of $p^\mu = dx^\mu/d\lambda$ after dividing through by $p^0 = dt/d\lambda$.

We now consider $dp_i/dt = -\partial H/\partial x^i$. Differentiating Eq. (2.9), with respect to x^i (at fixed p_j and t), we find

$$2\underbrace{(g^{00}p_0 + g^{0j}p_j)}_{p^0} \frac{\partial p_0}{\partial x^i} + \left[\frac{\partial g^{00}}{\partial x^i}(p_0)^2 + 2\frac{\partial g^{0j}}{\partial x^i}p_0p_j + \frac{\partial g^{jk}}{\partial x^i}p_jp_k \right] = 0. \quad (2.19)$$

We can simplify this by recalling that the derivative of the inverse metric is

$$\frac{\partial g^{\mu\nu}}{\partial x^i} = -g^{\mu\rho}\frac{\partial g_{\rho\tau}}{\partial x^i}g^{\tau\nu}, \quad (2.20)$$

so that

$$\begin{aligned} \frac{\partial g^{00}}{\partial x^i}(p_0)^2 + 2\frac{\partial g^{0j}}{\partial x^i}p_0p_j + \frac{\partial g^{jk}}{\partial x^i}p_jp_k &= -\frac{\partial g_{\rho\tau}}{\partial x^i} \left[g^{\rho 0}g^{\tau 0}(p_0)^2 + 2g^{\rho 0}g^{\tau j}p_0p_j + g^{\rho j}g^{\tau k}p_jp_k \right] \\ &= -\frac{\partial g_{\rho\tau}}{\partial x^i} (g^{\rho 0}p_0 + g^{\rho j}p_j) (g^{\tau 0}p_0 + g^{\tau k}p_k) \\ &= -\frac{\partial g_{\rho\tau}}{\partial x^i} p^\rho p^\tau, \end{aligned} \quad (2.21)$$

which, finally, gives

$$\frac{dp_i}{dt} = \frac{1}{2p^0} \frac{\partial g_{\rho\tau}}{\partial x^i} p^\rho p^\tau. \quad (2.22)$$

These are just the spatial components of the geodesic equation (2.16), after dividing through by $p^0 = dt/d\lambda$. (The time component of the geodesic equation is implied by the on-shell condition (2.9) and the equation for the spatial components; it does not appear directly in the Hamiltonian formulation.)

2.2 CMB dipole

As we noted earlier, the largest anisotropy observed in the CMB is a dipole with amplitude of around 3 mK. This dipole is due to our peculiar motion, something that we can understand using the Lorentz invariance of the photon distribution function.

Suppose that the CMB is isotropic in some reference frame, and has a blackbody spectrum with temperature \bar{T}_{CMB} as observed in that frame. Consider us moving relative to the CMB with 3-velocity \mathbf{v} . Then a photon that we detect at energy E moving along direction \mathbf{e} has energy E_{CMB} in the CMB rest-frame given by the usual Doppler shift

$$E_{\text{CMB}} = E\gamma(1 + \mathbf{v} \cdot \mathbf{e}), \quad (2.23)$$

where $\gamma \equiv (1 - \mathbf{v}^2)^{-1/2}$. Since the radiation is isotropic and blackbody in its rest frame, the distribution function is

$$f(p^\mu) \propto \frac{1}{e^{E_{\text{CMB}}/\bar{T}_{\text{CMB}}} - 1}. \quad (2.24)$$

If we express this in terms of energy and direction in our reference frame, we find

$$f(p^\mu) \propto \frac{1}{e^{E\gamma(1+\mathbf{v}\cdot\mathbf{e})/\bar{T}_{\text{CMB}}} - 1}. \quad (2.25)$$

This still looks like a blackbody along any direction but the observed temperature varies over the sky as

$$T(\mathbf{e}) = \frac{\bar{T}_{\text{CMB}}}{\gamma(1 + \mathbf{v} \cdot \mathbf{e})} \approx \bar{T}_{\text{CMB}}(1 - \mathbf{v} \cdot \mathbf{e}) \quad (2.26)$$

for $|\mathbf{v}| \ll 1$. This is a dipole anisotropy. The measured dipole implies the solar system barycentre has speed $3.68 \times 10^5 \text{ m s}^{-1}$ relative to the CMB. It is clear from Eq. (2.26) that relative motion also produces quadrupole anisotropies and alters the monopole temperature but only at $O(|\mathbf{v}|^2)$.

After subtracting this *kinematic dipole*, we are left with the true cosmological anisotropies that are the focus of the rest of this course.

2.3 Boltzmann equation

The Boltzmann equation describes the evolution of the distribution function in phase space. As we have seen, if there is no scattering, the distribution function is conserved along the photon path but, more generally, scattering terms also appear in the Boltzmann equation. In this section, we develop the Boltzmann equation for the CMB (Thomson) scattering off free electrons.

It is convenient to write the energy E and 3-momentum components (on the orthonormal tetrad) of a photon as

$$E = \epsilon/a, \quad p^i = (\epsilon/a)e^i. \quad (2.27)$$

We call ϵ the comoving energy and it is constant along the photon path *in the absence of perturbations* (recall that the energy E of a photon measured by a comoving observer in an FRW universe satisfies $E \propto 1/a$). It follows that $d\ln\epsilon/d\eta$ is first order in perturbations. We shall work out the explicit expression for $d\ln\epsilon/d\eta$ to first order later. The e^i are the tetrad components of the *direction of propagation*. We shall often write e^i in 3-vector form as \mathbf{e} , where \mathbf{e} is a unit 3-vector (i.e., $\delta_{ij}e^i e^j = 1$). Due to isotropy, the direction \mathbf{e} is constant in an FRW universe and so $d\mathbf{e}/d\eta$ is also first order in metric perturbations.

In terms of ϵ and \mathbf{e} , the coordinate components of the 4-momentum of the photon follow from the expressions for the tetrad in Eq. (2.3):

$$p^\mu = \epsilon a^{-2} [1 - A + e^i B_i, (1 - C)e^i - e^j E_j^i]. \quad (2.28)$$

Since $p^\mu = dx^\mu/d\lambda$, dividing the i th component by the 0th component, we have

$$dx^i/d\eta = (1 + A - C - e^j B_j)e^i - e^j E_j^i. \quad (2.29)$$

At zero order in perturbations, we obtain the simple result $dx^i/d\eta = e^i$.

We can now write the distribution function as a function of conformal time η , spatial position x^i (which we shall generally write as \mathbf{x}), comoving energy ϵ and direction \mathbf{e} . Since the distribution function is conserved in the absence of scattering, we have

$$\frac{df}{d\eta} = \frac{\partial f}{\partial \eta} + \frac{\partial f}{\partial \mathbf{e}} \cdot \frac{d\mathbf{e}}{d\eta} + \frac{\partial f}{\partial \mathbf{x}} \cdot \frac{\partial \mathbf{x}}{\partial \eta} + \frac{\partial f}{\partial \ln \epsilon} \frac{d \ln \epsilon}{d\eta} = \left. \frac{df}{d\eta} \right|_{\text{scatt.}}, \quad (2.30)$$

where the right-hand side describes the effects of scattering. The left-hand side is the total derivative along the photon path in phase space, and would vanish by Liouville's theorem if there were no scattering. We prefer to work with $d\ln\epsilon/d\eta$ rather than $d\epsilon/d\eta$ since gravitational redshifting is achromatic, which makes $d\ln\epsilon/d\eta$ independent of ϵ . We consider the last three terms on the left of Eq. (2.30) separately.

- $(\partial f / \partial \mathbf{e}) \cdot (d\mathbf{e} / d\eta)$: this term is second order in perturbations and so can be dropped in a linear calculation. This is because any dependence of f on \mathbf{e} is first order by isotropy of the background FRW model, and $d\mathbf{e} / d\eta$ is also first order. The term describes *gravitational lensing* of the CMB, and is actually important at the few per cent level for the small-scale CMB power spectrum. Exploiting CMB lensing is at the frontier of observational CMB research but the theory is beyond the scope of this course.
- $(\partial f / \partial \mathbf{x}) \cdot (d\mathbf{x} / d\eta)$: we can use the zero-order result $d\mathbf{x} / d\eta = \mathbf{e}$ in this term, since any inhomogeneity in the distribution function is already first order by homogeneity of the background FRW model.
- $(\partial f / \partial \ln \epsilon)(d \ln \epsilon / d\eta)$: we only need the distribution function at zero order in this term since $d \ln \epsilon / d\eta$ is first order.

Writing the zero-order distribution function as $\bar{f}(\eta, \epsilon)$, we have

$$\frac{\partial f}{\partial \eta} + \mathbf{e} \cdot \nabla f + \frac{\partial \bar{f}}{\partial \ln \epsilon} \frac{d \ln \epsilon}{d\eta} = \left. \frac{df}{d\eta} \right|_{\text{scatt.}}, \quad (2.31)$$

correct to linear order. As we shall show shortly, the scattering term on the right is first order, so that $\partial \bar{f} / \partial \eta = 0$. In other words, *the zero-order distribution function depends only on the comoving energy*. Since the CMB was in thermal equilibrium at early times, \bar{f} should be very close to a Planck (blackbody) distribution with a temperature $\bar{T}_{\text{CMB}}(\eta)$ that varies as $1/a$:

$$\begin{aligned} \bar{f}(\epsilon) &\propto \left[\exp\left(\frac{E}{k_B \bar{T}_{\text{CMB}}(\eta)}\right) - 1 \right]^{-1} \\ &\propto \left[\exp\left(\frac{\epsilon}{a k_B \bar{T}_{\text{CMB}}(\eta)}\right) - 1 \right]^{-1}. \end{aligned} \quad (2.32)$$

Indeed, the (angle-averaged) energy spectrum of the CMB has been measured to high precision by the FIRAS instrument that flew on the COBE satellite. The fit to a blackbody is exquisite, with a current temperature $\bar{T}_{\text{CMB}} = 2.7255$ K.

The third term on the left of Eq. (2.31) describes the effect of gravitational redshifting. Since this is achromatic, the fluctuations in the distribution function inherit the frequency dependence $\partial \bar{f} / \partial \ln \epsilon$ (we shall see that the same frequency dependence is implied by the scattering term). We can therefore write

$$f(\eta, \mathbf{x}, \epsilon, \mathbf{e}) = \bar{f}(\epsilon) \left[1 - \Theta(\eta, \mathbf{x}, \mathbf{e}) \frac{d \ln \bar{f}}{d \ln \epsilon} \right], \quad (2.33)$$

where $\Theta(\eta, \mathbf{x}, \mathbf{e})$ is the *fractional temperature fluctuation*, which is independent of ϵ . To see that this is the temperature fluctuation, consider a perturbed distribution function

that is Planckian but with a direction- and position-dependent fractional temperature fluctuation Θ ; then

$$\begin{aligned} f(\eta, \mathbf{x}, \epsilon, \mathbf{e}) &\propto \left[\exp\left(\frac{\epsilon}{ak_B\bar{T}_{\text{CMB}}(\eta)(1+\Theta)}\right) - 1 \right]^{-1} \\ &\propto \left[\exp\left(\frac{\epsilon(1-\Theta)}{ak_B\bar{T}_{\text{CMB}}(\eta)}\right) - 1 \right]^{-1} \\ &= \bar{f}(\epsilon) - \Theta\epsilon\frac{d\bar{f}}{d\epsilon}, \end{aligned} \quad (2.34)$$

to first order in Θ . The quadrupole and higher moments of the gauge-invariant temperature fluctuation are gauge-invariant.

We can now express the Boltzmann equation in terms of the temperature fluctuation as follows:

$$-\frac{d\bar{f}(\epsilon)}{d\ln\epsilon} \left(\frac{\partial\Theta}{\partial\eta} + \mathbf{e}\cdot\nabla\Theta - \frac{d\ln\epsilon}{d\eta} \right) = \frac{df}{d\eta} \Big|_{\text{scatt.}}. \quad (2.35)$$

2.3.1 Collision term for Thomson scattering of unpolarized radiation

We now consider the scattering term in the Boltzmann equation. For the generation of *anisotropies*, we really only care about the details of scattering around recombination where the universe transitioned from being tightly coupled to transparent. Scattering at much earlier times is essential to maintain thermal equilibrium, but the details are irrelevant for anisotropy generation. The only effect of these very efficient scattering processes on the CMB anisotropies is to force the distribution function to be isotropic in the rest frame of the scattering medium, thus providing the initial condition for the growth of anisotropies through recombination.

The dominant scattering mechanism that affects the anisotropies of the CMB is therefore Thomson scattering off free electrons. This is the non-relativistic limit of Compton scattering. The Thomson limit is appropriate around recombination since the typical photon energy is around 1 eV, much less than the rest-mass energy (around 0.5 MeV) of an electron. We can also ignore the non-zero temperature T_e of the electrons, since this only produces corrections of order $k_B T_e / (m_e c^2) \sim 10^{-5}$ to the scattering term. We shall make one further approximation here, which is to ignore polarization corrections. We shall fix this omission later in Sec. 6.

In the rest-frame of an electron, Thomson scattering does not alter the photon energy. For a single electron at rest, the rate at which it scatters photons per solid angle, per unit incident photon flux is called the *differential cross-section*. For Thomson scattering of unpolarized radiation, averaged over the outgoing polarization, the differential cross-

section is

$$\frac{d\sigma}{d\Omega} = \frac{3\sigma_T}{16\pi}(1 + \cos^2\theta), \quad (2.36)$$

where θ is the scattering angle, i.e., the angle between the incident photon direction \mathbf{e}_{in} and the scattered direction \mathbf{e} . The Thomson cross-section,

$$\sigma_T = \frac{8\pi}{3} \left(\frac{e_c^2}{4\pi\epsilon_0 m_e c^2} \right)^2 = 6.65 \times 10^{-29} \text{ m}^2, \quad (2.37)$$

where e_c is the electron charge, is the integral of $d\sigma/d\Omega$ over solid angle. Note that the Thomson cross-section scales as the inverse of the mass of the charged scatterer and so the effects of free protons can be ignored relative to the electrons.

If the electrons were at rest, so they have no bulk velocity relative to our chosen reference frame, the scattering rate *with respect to proper time τ for an observer at rest in our reference frame* from electrons with proper number density n_e is

$$\frac{df(\epsilon, \mathbf{e})}{d\tau} \Big|_{\text{scatt.}} = n_e \int d\mathbf{e}_{\text{in}} \frac{d\sigma}{d\Omega} [f(\epsilon, \mathbf{e}_{\text{in}}) - f(\epsilon, \mathbf{e})]. \quad (2.38)$$

Here, the first term in the integrand on the left describes in-scattering of photons with energy ϵ and direction \mathbf{e}_{in} *into* the direction \mathbf{e} , while the second term describes out-scattering of photons with direction \mathbf{e} (to direction \mathbf{e}_{in}). Clearly, for isotropic radiation, the scattering term vanishes for electrons at rest and so the scattering term is necessarily first order in perturbations, as advertised earlier. In the out-scattering term, we can take the distribution function out of the integral to obtain

$$\frac{df(\epsilon, \mathbf{e})}{d\tau} \Big|_{\text{scatt.}} = -n_e \sigma_T f(\epsilon, \mathbf{e}) + \frac{3n_e \sigma_T}{16\pi} \int d\mathbf{e}_{\text{in}} f(\epsilon, \mathbf{e}_{\text{in}}) [1 + (\mathbf{e} \cdot \mathbf{e}_{\text{in}})^2]. \quad (2.39)$$

We can obtain the scattering rate allowing for a bulk velocity of the electrons by Lorentz transforming Eq. (2.39); the details are as follows.

Accounting for the electron bulk velocity: In the rest frame of the electrons, the scattering rate (in terms of proper time $\tilde{\tau}$ in that frame) is given by Eq. (2.39). Denoting quantities in the electron rest-frame with a tilde, we have

$$\frac{d\tilde{f}(\tilde{\epsilon}, \tilde{\mathbf{e}})}{d\tilde{\tau}} \Big|_{\text{scatt.}} = -n_e \sigma_T \tilde{f}(\tilde{\epsilon}, \tilde{\mathbf{e}}) + \frac{3n_e \sigma_T}{16\pi} \int d\tilde{\mathbf{e}}_{\text{in}} \tilde{f}(\tilde{\epsilon}, \tilde{\mathbf{e}}_{\text{in}}) [1 + (\tilde{\mathbf{e}}_{\text{in}} \cdot \tilde{\mathbf{e}})^2]. \quad (2.40)$$

The photon directions in the rest-frame are expressed in terms of components on the Lorentz-boosted version of the tetrad in Eq. (2.3). In the original frame, the baryons and electrons (which are tightly coupled by Coulomb scattering) have 3-velocity with tetrad components

v_b^i that we shall write in 3-vector form as \mathbf{v}_b . To zero-order in perturbations, the distribution function is isotropic in the rest frame and the out-scattering rate (the first term on the right of Eq. 2.40) balances the in-scattering rate (the second term on the right). The scattering rate is therefore first-order in perturbations and we only need the electron density n_e at zero-order. Moreover, we can neglect the difference in proper time between the rest frame and our chosen reference frame in transforming the scattering rate, since this first-order difference multiplies the first-order scattering rate. With these observations, and using the Lorentz invariance of the distribution function, we can write the scattering rate with respect to conformal time in our original reference frame as

$$\frac{df(\epsilon, \mathbf{e})}{d\eta} \Big|_{\text{scatt.}} = -a\bar{n}_e\sigma_T f(\epsilon, \mathbf{e}) + \frac{3a\bar{n}_e\sigma_T}{16\pi} \int d\tilde{\mathbf{e}}_{\text{in}} \tilde{f}(\tilde{\epsilon}, \tilde{\mathbf{e}}_{\text{in}}) [1 + (\tilde{\mathbf{e}}_{\text{in}} \cdot \tilde{\mathbf{e}})^2] . \quad (2.41)$$

The in-scattering process on the right of Eq. (2.41) involves the following direction and energy changes in the two frames:

$$\begin{aligned} \text{Original:} & \quad (\epsilon_{\text{in}}, \mathbf{e}_{\text{in}}) \rightarrow (\epsilon, \mathbf{e}) \\ \text{Rest frame:} & \quad (\tilde{\epsilon}_{\text{in}}, \tilde{\mathbf{e}}_{\text{in}}) \rightarrow (\tilde{\epsilon}, \tilde{\mathbf{e}}) \quad \text{with } \tilde{\epsilon}_{\text{in}} = \tilde{\epsilon} . \end{aligned} \quad (2.42)$$

Note that the photon energy does change in the original frame (i.e., that in which the electron is moving). Applying standard Lorentz transformations gives

$$\tilde{\epsilon} = \gamma\epsilon(1 - \mathbf{e} \cdot \mathbf{v}_b), \quad (2.43)$$

and, using the inverse transformation,

$$\begin{aligned} \epsilon_{\text{in}} &= \gamma\tilde{\epsilon}_{\text{in}}(1 + \tilde{\mathbf{e}}_{\text{in}} \cdot \mathbf{v}_b) \\ &= \gamma^2\epsilon(1 + \tilde{\mathbf{e}}_{\text{in}} \cdot \mathbf{v}_b)(1 - \mathbf{e} \cdot \mathbf{v}_b) \\ &\approx \epsilon(1 + \tilde{\mathbf{e}}_{\text{in}} \cdot \mathbf{v}_b - \mathbf{e} \cdot \mathbf{v}_b), \end{aligned} \quad (2.44)$$

where γ is the Lorentz factor between the two frames. We now use the Lorentz invariance of f to write (to first order)

$$\begin{aligned} \tilde{f}(\tilde{\epsilon}, \tilde{\mathbf{e}}_{\text{in}}) &= f(\epsilon_{\text{in}}, \mathbf{e}_{\text{in}}) \\ &= f[\epsilon(1 + \tilde{\mathbf{e}}_{\text{in}} \cdot \mathbf{v}_b - \mathbf{e} \cdot \mathbf{v}_b), \mathbf{e}_{\text{in}}] \\ &= \bar{f}[\epsilon(1 + \tilde{\mathbf{e}}_{\text{in}} \cdot \mathbf{v}_b - \mathbf{e} \cdot \mathbf{v}_b)] - \frac{d\bar{f}}{d\ln\epsilon}\Theta(\mathbf{e}_{\text{in}}) \\ &= \bar{f}(\epsilon) + \frac{d\bar{f}}{d\ln\epsilon}(\tilde{\mathbf{e}}_{\text{in}} - \mathbf{e}) \cdot \mathbf{v}_b - \frac{d\bar{f}}{d\ln\epsilon}\Theta(\mathbf{e}_{\text{in}}) . \end{aligned} \quad (2.45)$$

Substituting into the in-scattering part of Eq. (2.41), the $\bar{f}(\epsilon)$ term integrates to give $a\bar{n}_e\sigma_T\bar{f}(\epsilon)$. The third term on the right of Eq. (2.45) is first-order and so we can replace $d\tilde{\mathbf{e}}_{\text{in}}$ with $d\mathbf{e}_{\text{in}}$ and $\tilde{\mathbf{e}}_{\text{in}} \cdot \tilde{\mathbf{e}}$ with $\mathbf{e}_{\text{in}} \cdot \mathbf{e}$ in the angular integral. Finally, for the remaining term in Eq. (2.45) the $\tilde{\mathbf{e}}_{\text{in}} \cdot \mathbf{v}_b$ part integrates to zero by (parity) symmetry while the $-\mathbf{e} \cdot \mathbf{v}_b$ part is independent of $\tilde{\mathbf{e}}_{\text{in}}$ and can be taken outside the angular integral. Putting these pieces together, we find

$$\frac{df(\epsilon, \mathbf{e})}{d\eta} \Big|_{\text{scatt.}} = \frac{d\bar{f}}{d\ln\epsilon} \left(a\bar{n}_e\sigma_T\Theta(\mathbf{e}) - \frac{3a\bar{n}_e\sigma_T}{16\pi} \int d\hat{\mathbf{m}} \Theta(\hat{\mathbf{m}}) [1 + (\mathbf{e} \cdot \hat{\mathbf{m}})^2] - a\bar{n}_e\sigma_T\mathbf{e} \cdot \mathbf{v}_b \right) \quad (2.46)$$

where $\hat{\mathbf{m}}$ is a unit vector. This is our required expression for the scattering rate in the presence of non-zero electron bulk velocity. Note that in-scattering into the direction of the electron velocity *increases* the temperature anisotropy along that direction.

As promised, the scattering term in Eq. (2.46) has the same energy dependence, $d\bar{f}/d\ln \epsilon$, as the gravitational-redshifting term. This allows us to cancel the energy-dependent term on the left of the Boltzmann equation (2.35) to obtain our final form of the Boltzmann equation for the temperature fluctuation:

$$\frac{\partial \Theta}{\partial \eta} + \mathbf{e} \cdot \nabla \Theta - \frac{d \ln \epsilon}{d \eta} = -a\bar{n}_e \sigma_T \Theta + \frac{3a\bar{n}_e \sigma_T}{16\pi} \int d\hat{\mathbf{m}} \Theta(\hat{\mathbf{m}}) [1 + (\mathbf{e} \cdot \hat{\mathbf{m}})^2] + a\bar{n}_e \sigma_T \mathbf{e} \cdot \mathbf{v}_b. \quad (2.47)$$

2.3.2 Line-of-sight solution

The Boltzmann equation for Θ is an integro-differential equation in seven dimensions. Despite this, the equation can be solved very efficiently by expanding in appropriate basis functions, as we shall study later in this course. Here, we simply note that the equation can be solved formally as an integral along the unperturbed line of sight $\mathbf{x}(\eta) = \mathbf{x}_0 - (\eta_0 - \eta)\mathbf{e}$ (choosing the observation point to be \mathbf{x}_0 and at time η_0). To see this, note that the first two terms on the left of Eq. (2.47) are the convective derivative of Θ along $\mathbf{x}(\eta)$. Introducing the zero-order optical depth τ (see below), the Boltzmann equation has an integrating factor $e^{-\tau}$, and we can write

$$\frac{d}{d\eta} (e^{-\tau} \Theta) = S, \quad (2.48)$$

where the source term at events along the line of sight is

$$S = e^{-\tau} \frac{d \ln \epsilon}{d \eta} + \frac{3}{16\pi} g \int d\hat{\mathbf{m}} \Theta(\hat{\mathbf{m}}) [1 + (\mathbf{e} \cdot \hat{\mathbf{m}})^2] + g \mathbf{e} \cdot \mathbf{v}_b, \quad (2.49)$$

and is a linear functional of $\Theta(\mathbf{e})$ at that event. Here, we have introduced the visibility function (see below) $g \equiv -\dot{\tau}e^{-\tau}$.

Optical depth and the visibility function. The (zero-order) *comoving mean free path* for a photon, i.e., the comoving distance it travels on average between scatterings, is $l_p = 1/(a\bar{n}_e \sigma_T)$. The mean free path rises rapidly through recombination as the free electrons recombine with protons to form hydrogen atoms. A related concept is the *optical depth* along the line-of-sight between conformal times η and η_0 , defined by

$$\tau(\eta) \equiv \int_{\eta}^{\eta_0} d\eta a\bar{n}_e \sigma_T. \quad (2.50)$$

This is an integral of the inverse of l_p with respect to conformal time, so that

$$d\tau/d\eta = -a\bar{n}_e\sigma_T. \quad (2.51)$$

Note that the optical depth is dimensionless. The interpretation of τ is that $e^{-\tau}$ is the probability of no scattering as a photon traverses the interval (η, η_0) . To see this, consider a subset of this interval of duration $\Delta\eta$, which is much shorter than the (comoving) Hubble time there. The mean free path is then almost constant over $\Delta\eta$. Let us now break $\Delta\eta$ up into N intervals, with N large enough such that there is either zero or one scattering per interval. The probability of one scattering is $\Delta\eta/(Nl_p)$ and so, in N trials, the probability of no scattering is

$$\text{Pr}(\text{no scattering}) = \left(1 - \frac{\Delta\eta}{Nl_p}\right)^N. \quad (2.52)$$

In the limit as $N \rightarrow \infty$, the probability of no scattering over $\Delta\eta$ tends to

$$\text{Pr}(\text{no scattering}) = e^{-\Delta\eta/l_p}. \quad (2.53)$$

If we now consider a number of such intervals, with $\sum \Delta\eta = \eta_0 - \eta$, we have

$$\text{Pr}(\text{no scattering}) = e^{-\sum(\Delta\eta/l_p)} \rightarrow e^{-\tau}. \quad (2.54)$$

In a universe without late-time reionization (i.e., one in which hydrogen and helium atoms were not reionized by ultraviolet radiation from the first stars and quasars), moving backwards in time, the optical depth is very small until recombination at which point it rises sharply. There is thus a very low probability of scattering after recombination to the present, but scattering is certain from a few mean free paths before recombination. However, we now know from quasar absorption spectra, and from the large-angle polarization of the CMB (see Sec. 6), that the intergalactic medium was actually reionized around $z = 10$, with an optical depth through reionization of $\tau_{\text{reion}} \approx 0.1$. This means that around 90% of CMB photons have not scattered since recombination.

The *visibility function* $g(\eta)$ is the probability density that a photon last-scattered at time η . Since the probability of scattering in a conformal time interval $d\eta$ at η is $d\eta/l_p(\eta)$, the visibility function is

$$g(\eta) = a\bar{n}_e\sigma_T e^{-\tau} = -\dot{\tau}e^{-\tau}, \quad (2.55)$$

where overdots denote differentiation with respect to η . The integral of the visibility function is simply

$$\int_\eta^{\eta_0} d\eta' g(\eta') = 1 - e^{-\tau(\eta)}. \quad (2.56)$$

Integrating over all time by setting the lower limit $\eta = 0$, we see that the visibility function is essentially normalised to unity, as expected from its definition. The visibility function (and ionization fraction) is plotted in Fig. 5. The visibility is quite sharply peaked around the time of recombination, with a width of few $\times 10$ Mpc. (There is an additional contribution at lower redshift from reionization that is not shown in the figure.) The peak of the visibility function can be used to define the time of *last scattering* η_* . This is determined by the cosmological parameters such as the baryon and cold-dark matter (CDM) density, since these determine

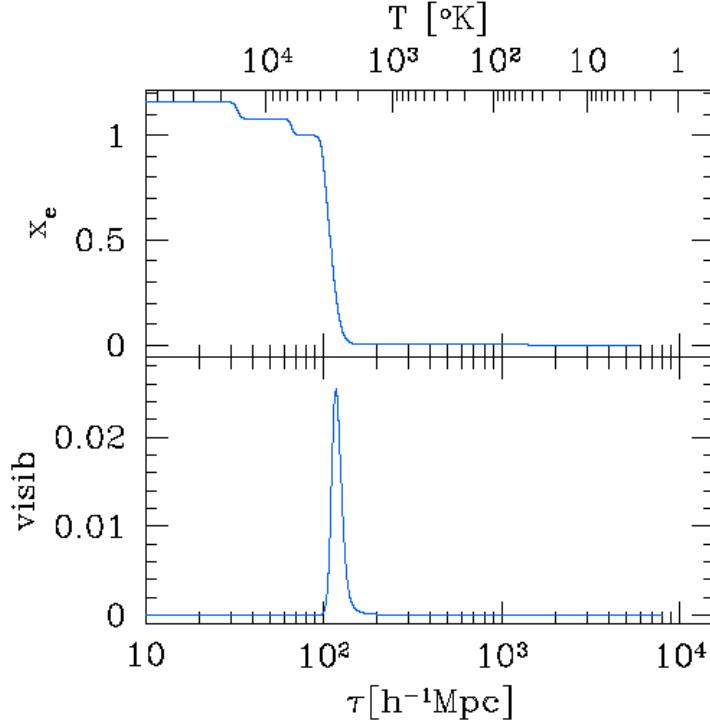


Figure 5: Ionization fraction x_e (top) and visibility function (bottom) as a function of conformal time (denoted by τ in this figure!). The ionization fraction is the ratio of unbound electrons to the total number of protons (excluding those in helium nuclei). The first two steps in x_e correspond to the two recombinations of helium, with $x_e = 1$ temporarily after helium recombination but before hydrogen recombination.

the ionization history. These parameters are determined accurately with recent observations. For example, the recent 2013 *Planck* results⁸ determine $\eta_* = 144.6 \text{ Mpc}$ and $z_* = 1090.1$.

We can integrate Eq. (2.48) along the background line-of-sight to obtain

$$\Theta(\eta_0, \mathbf{x}_0, \mathbf{e}) = \int_0^{\eta_0} d\eta' S(\eta', \mathbf{x}_0 - (\eta_0 - \eta')\mathbf{e}, \mathbf{e}), \quad (2.57)$$

where we have used $\tau(\eta_0) = 0$ and that $\tau(0)$ is essentially infinite. Terms in the source function that are multiplied by the visibility function are localised near last scattering. The other term, $e^{-\tau} d\ln \epsilon/d\eta$, contributes from last-scattering onwards.

⁸In the *Planck* papers, last scattering is identified not as the peak of the visibility function, but the time at which the optical depth to the present is unity ignoring the contribution from reionization.

2.4 Stress–energy tensor

The CMB couples gravitationally to other species, via its stress–energy tensor $T^{\mu\nu}$. This is given in terms of the distribution function by

$$T^{\mu\nu} = \int \frac{d^3\mathbf{p}}{E(\mathbf{p})} f p^\mu p^\nu. \quad (2.58)$$

Recalling that $d^3\mathbf{p}/E(\mathbf{p})$ is Lorentz-invariant, we see that the right-hand side is tensor valued. To see where Eq. (2.58) comes from, we note the physical interpretation of the tetrad components of the stress–energy tensor:

$T^{0\hat{0}} = \bar{\rho}(\eta) + \delta\rho = \bar{\rho}(\eta)(1 + \delta)$	energy density
$T^{0\hat{i}} = q^{\hat{i}} = [\bar{\rho}(\eta) + \bar{P}(\eta)]v^{\hat{i}}$	momentum density
$T^{\hat{i}\hat{j}} = [\bar{P}(\eta) + \delta P]\delta^{\hat{i}\hat{j}} - \Pi^{\hat{i}\hat{j}}$	flux of i th component of 3-momentum along j th direction ,

(2.59)

where we have written the energy density and pressure in terms of their background values (denoted by an overbar) and perturbations δ (fractional) and δP . We have also introduced the momentum density $q^{\hat{i}}$ and associated bulk velocity $v^{\hat{i}}$, and the symmetric trace-free anisotropic stress $\Pi^{\hat{i}\hat{j}}$. If we consider the 00 tetrad component of Eq. (2.58), we have

$$T^{0\hat{0}} = \int d^3\mathbf{p} f E, \quad (2.60)$$

which is clearly the energy density of the radiation given that $f d^3\mathbf{p}$ is the (proper) number density of photons. Similarly, for the $0i$ tetrad components,

$$T^{0\hat{i}} = \int d^3\mathbf{p} f p^{\hat{i}}, \quad (2.61)$$

which is indeed the momentum density. Finally, for the ij tetrad components, we have

$$T^{\hat{i}\hat{j}} = \int \frac{d^3\mathbf{p}}{E} f p^{\hat{i}} p^{\hat{j}}. \quad (2.62)$$

Since, even for a massive particle, $p^{\hat{i}}/E$ is the 3-velocity of the particle, the right-hand side of Eq. (2.62) gives the flux of 3-momentum, as required.

We can express δ , δP , $v^{\hat{i}}$ and $\Pi^{\hat{i}\hat{j}}$ in terms of the temperature perturbation $\Theta(\eta, \mathbf{x}, \mathbf{e})$ using Eq. (2.34). For example, for the 00 component, we have

$$\begin{aligned} \bar{\rho}(1 + \delta) &= \frac{1}{a^4} \int d\mathbf{e} \int d\epsilon f \epsilon^3 \\ &= \frac{4\pi}{a^4} \int d\epsilon \bar{f} \epsilon^3 - \frac{1}{a^4} \left(\int d\epsilon \frac{d\bar{f}}{d\epsilon} \epsilon^4 \right) \left(\int d\mathbf{e} \Theta \right) \\ &= \frac{4\pi}{a^4} \int d\epsilon \bar{f} \epsilon^3 \left(1 + 4 \int \frac{d\mathbf{e}}{4\pi} \Theta \right), \end{aligned} \quad (2.63)$$

where we have integrated by parts to obtain the final equality. It follows that

$$\bar{\rho} = \frac{4\pi}{a^4} \int d\epsilon \bar{f} \epsilon^3 \quad \text{and} \quad \delta = 4 \int \frac{d\mathbf{e}}{4\pi} \Theta . \quad (2.64)$$

Note that $\bar{\rho} \propto a^{-4}$, as expected for radiation. We see that the *monopole* moment of Θ determines the fractional energy density. We leave the other components as an exercise.

Exercise: show that $\bar{P} = \bar{\rho}/3$ and that

$$\begin{aligned} \delta P &= \bar{P}\delta \\ \mathbf{v} &= 3 \int \frac{d\mathbf{e}}{4\pi} \Theta \mathbf{e} \\ \Pi^{ij} &= -4\bar{\rho} \int \frac{d\mathbf{e}}{4\pi} \Theta e^{\langle i} e^{\rangle j}, \end{aligned} \quad (2.65)$$

where angle brackets denotes the symmetric trace-free part so that $e^{\langle i} e^{\rangle j} = e^i e^j - \delta^{ij}/3$. You will need to use the result that

$$\int d\mathbf{e} e^{\langle i} e^{\rangle j} = 0. \quad (2.66)$$

We see that the bulk velocity is determined by the *dipole* moment of Θ and the anisotropic stress by the *quadrupole* moment.

For freely-propagating radiation the stress-energy tensor is conserved, $\nabla_\mu T^{\mu\nu} = 0$, giving the relativistic continuity and Euler equations. More generally, the collisional Boltzmann equation can be used to derive the additions to the continuity and Euler equations due to energy and momentum exchange during scattering. We shall do this in Sec. 4 for scalar perturbations.

3 Temperature anisotropies from scalar perturbations

We now specialise to *scalar perturbations*. Recall from the *Cosmology* course that for scalar perturbations all perturbed 3-tensors can be written in terms of spatial derivatives of first-order scalar fields. We can use the gauge freedom to eliminate the metric perturbations B_i and E_{ij} ; this defines the *conformal Newtonian gauge*. We denote the metric perturbations A and $-C$ in the Newtonian gauge as ψ and ϕ , so that the perturbed metric is

$$ds^2 = a^2(\eta) [-(1+2\psi)d\eta^2 + (1-2\phi)\delta_{ij}dx^i dx^j]. \quad (3.1)$$

In this gauge, the spatial sections have an isotropic intrinsic geometry (since the induced metric is proportional to δ_{ij}), and the worldlines of constant x^i are orthogonal to the surfaces of constant time. The orthonormal tetrad vectors of Eq. (2.3) point along the coordinate basis vectors with

$$(E_0)^\mu = a^{-1}(1-\psi)\delta_0^\mu \quad \text{and} \quad (E_i)^\mu = a^{-1}(1+\phi)\delta_i^\mu. \quad (3.2)$$

For completeness, we list the connection coefficients for the metric in Eq. (3.1) correct to first-order in metric perturbations:

$$\begin{aligned} \Gamma_{00}^0 &= \mathcal{H} + \dot{\psi} \\ \Gamma_{0i}^0 &= \partial_i \psi \\ \Gamma_{00}^i &= \delta^{ij} \partial_j \psi \\ \Gamma_{ij}^0 &= \mathcal{H} \delta_{ij} - [\dot{\phi} + 2\mathcal{H}(\phi + \psi)] \delta_{ij} \\ \Gamma_{j0}^i &= \mathcal{H} \delta_j^i - \dot{\phi} \delta_j^i \\ \Gamma_{jk}^i &= -2\delta_{(j}^i \partial_{k)} \phi + \delta_{jk} \delta^{il} \partial_l \phi. \end{aligned} \quad (3.3)$$

3.1 $d \ln \epsilon / d\eta$ for scalar perturbations

Photons move on geodesics of the perturbed metric so

$$\frac{dp^\mu}{d\lambda} + \Gamma_{\nu\rho}^\mu p^\nu p^\rho = 0, \quad (3.4)$$

where λ is an affine parameter such that $p^\mu = dx^\mu/d\lambda$. The parameterisation of Eq. (2.28) for p^μ reduces to

$$p^\mu = a^{-2}\epsilon[1-\psi, (1+\phi)\mathbf{e}] \quad (3.5)$$

for scalar perturbations. It follows that

$$\begin{aligned}\frac{d\eta}{d\lambda} &= \frac{\epsilon}{a^2}(1 - \psi) \\ \frac{dx^i}{d\eta} &= (1 + \phi + \psi)e^i\end{aligned}\quad (3.6)$$

at linear order. The geodesic equation in conformal time is

$$(1 - \psi)\frac{\epsilon}{a^2}\frac{dp^\mu}{d\eta} + \Gamma_{\nu\rho}^\mu p^\nu p^\rho = 0. \quad (3.7)$$

The 0-component of this is

$$\begin{aligned}(1 - \psi)\frac{\epsilon}{a^2}\frac{d}{d\eta}\left(\frac{\epsilon}{a^2}(1 - \psi)\right) + \frac{\epsilon^2}{a^4}[\Gamma_{00}^0(1 - 2\psi) + 2\Gamma_{0i}^0e^i + \Gamma_{ij}^0(1 + 2\phi)e^i e^j] &= 0 \\ \Rightarrow \frac{1}{\epsilon}\frac{d\epsilon}{d\eta} - 2\mathcal{H} - \frac{d\psi}{d\eta} + \mathcal{H} + \dot{\psi} + 2e^i\partial_i\psi \\ + \delta_{ij}e^i e^j [\mathcal{H} - \dot{\phi} - 2\mathcal{H}(\phi + \psi)] [1 + 2(\phi + \psi)] &= 0,\end{aligned}\quad (3.8)$$

where we have substituted for the perturbed connection coefficients. The derivative $d\psi/d\eta$ is along the path of the photon so

$$d\psi/d\eta = \dot{\psi} + \mathbf{e} \cdot \nabla\psi \quad (3.9)$$

to first order. Adding and subtracting twice this term to Eq. (3.8) and simplifying gives

$$\frac{d\ln\epsilon}{d\eta} = -\frac{d\psi}{d\eta} + (\dot{\phi} + \dot{\psi}). \quad (3.10)$$

This equation tells us how the comoving energy evolves along the photon path in the presence of the metric perturbations. In the background, ϵ is constant, but this is modified by variation of ψ along the path (the first term on the right) and by time evolution of the potentials (second term on the right). For the CMB, the latter is important at late times, once accelerated expansion starts (again), and also around the time of last scattering due to the universe not being fully matter dominated at that time.

The i th component of the geodesic equation is left as an exercise.

Exercise: Show that the direction of propagation evolves along the photon path according to the first-order equation

$$\frac{d\mathbf{e}}{d\eta} = -(\nabla - \mathbf{e}\mathbf{e} \cdot \nabla)(\phi + \psi). \quad (3.11)$$

Equation (3.11) describes *gravitational lensing* whereby the photon direction is disturbed by the gradient of the potential $\phi + \psi$ perpendicular to the line of sight. In the background model, the direction is constant and, recall, this is all we require to compute the CMB temperature anisotropies since the dependence of the temperature on direction is already first order.

3.2 Temperature anisotropies

The contribution of $d \ln \epsilon / d\eta$ to the source term S in the Boltzmann equation, written in the form of Eq. (2.48), is

$$\begin{aligned} e^{-\tau} \frac{d \ln \epsilon}{d\eta} &= -e^{-\tau} \frac{d\psi}{d\eta} + e^{-\tau} (\dot{\phi} + \dot{\psi}) \\ &= -\frac{d}{d\eta} (e^{-\tau} \psi) + g\psi + e^{-\tau} (\dot{\phi} + \dot{\psi}) . \end{aligned} \quad (3.12)$$

It is convenient to rewrite the Boltzmann equation in the form

$$\frac{d}{d\eta} [e^{-\tau} (\Theta + \psi)] = S_{\text{scal}} , \quad (3.13)$$

where the modified (scalar) source term is

$$S_{\text{scal}} = e^{-\tau} (\dot{\phi} + \dot{\psi}) + g\psi + \frac{3}{16\pi} g \int d\hat{\mathbf{m}} \Theta(\hat{\mathbf{m}}) [1 + (\mathbf{e} \cdot \hat{\mathbf{m}})^2] + g\mathbf{e} \cdot \mathbf{v}_b . \quad (3.14)$$

Integrating along the (zero-order) line of sight gives

$$\Theta(\eta_0, \mathbf{x}_0, \mathbf{e}) + \psi(\eta_0, \mathbf{x}_0) = \int_0^{\eta_0} d\eta' S_{\text{scal}}(\eta', \mathbf{x}_0 - (\eta_0 - \eta')\mathbf{e}, \mathbf{e}) . \quad (3.15)$$

The second term on the left is a monopole and we shall generally omit this term since it does not contribute to anisotropy.

We gain useful insight into the physics of anisotropy formation from scalar perturbations by approximating last scattering as sharp (which is harmless on large angular scales that probe large linear scales) and ignoring reionization. In this case, $g(\eta) \approx \delta(\eta - \eta_*)$ and $e^{-\tau}$ is unity for $\eta > \eta_*$ and zero for $\eta < \eta_*$. Noting that

$$1 + (\mathbf{e} \cdot \hat{\mathbf{m}})^2 = \frac{4}{3} + e^{\langle i} e^{\rangle j} \hat{m}_{\langle i} \hat{m}_{\rangle j} , \quad (3.16)$$

so that

$$\int d\hat{\mathbf{m}} \Theta(\hat{\mathbf{m}}) [1 + (\mathbf{e} \cdot \hat{\mathbf{m}})^2] = \underbrace{\frac{16\pi}{3} \int \frac{d\hat{\mathbf{m}}}{4\pi} \Theta(\hat{\mathbf{m}})}_{\delta_\gamma/4} + \underbrace{4\pi e^{\langle i} e^{\rangle j} \int \frac{d\hat{\mathbf{m}}}{4\pi} \Theta(\hat{\mathbf{m}}) \hat{m}_{\langle i} \hat{m}_{\rangle j}}_{\propto \Pi_{\gamma, ij}} , \quad (3.17)$$

if we further neglect the quadrupole CMB anisotropy (i.e., the photon anisotropic stress) at last scattering⁹ we find that Eq. (3.15) reduces to

$$\Theta(\eta_0, \mathbf{x}_0, \mathbf{e}) + \psi(\eta_0, \mathbf{x}_0) = \Theta_0 + \psi + \mathbf{e} \cdot \mathbf{v}_b + \int_{\eta_*}^{\eta_0} d\eta' (\dot{\psi} + \dot{\phi}). \quad (3.18)$$

Here, we have defined $\Theta_0 = \delta_\gamma/4$, the monopole of the temperature perturbation. All non-integral terms on the right of Eq. (3.18) are evaluated at the time of last scattering, η_* , and position $\mathbf{x}_0 - \chi_* \mathbf{e}$, where $\chi_* \equiv \eta_0 - \eta_*$. Therefore we observe a spherical projection of the density, velocity, and gravitational potential at the time of last scattering over a sphere of radius χ_* centred on our location \mathbf{x}_0 . This sphere is the *last-scattering surface*.

The various terms in Eq. (3.18) have simple physical interpretations, as follows.

- Part of the temperature received from direction \mathbf{e} is sourced by the isotropic temperature of the CMB at the last-scattering event on the line-of-sight, Θ_0 .
- This isotropic temperature contribution is corrected for the gravitational redshift due to the difference in potential between the “emission” and reception events. The combination of the isotropic temperature and the gravitational redshift is often called the Sachs–Wolfe contribution, after the authors of the first paper to calculate the effect of inhomogeneities on the CMB.
- The Doppler shift $\mathbf{e} \cdot \mathbf{v}_b$ resulting from scattering off moving electrons further enhances the observed temperature when the scattering electron is moving towards the observer (i.e., $\mathbf{e} \cdot \mathbf{v}_b > 0$).
- Finally, there is an additional gravitational redshift contribution arising from evolution of the gravitational potentials in time (often called the integrated Sachs–Wolfe, or ISW, contribution).

The contributions to the angular power spectrum of the temperature anisotropies (to be discussed further below) from these various physical effects are illustrated in Fig. 6.

3.3 Evolution of plasma fluctuations

To evaluate the temperature anisotropies with Eq. (3.18) we need to know Θ_0 , \mathbf{v}_b , and ψ on the last-scattering surface (and the later evolution of the gravitational potentials). We shall see how to do this properly with the Boltzmann equation in Sec. 4, but to gain

⁹The neglect of the quadrupole anisotropy at last scattering is a good approximation on large scales, since a significant quadrupole only develops once the radiation starts to free-stream.

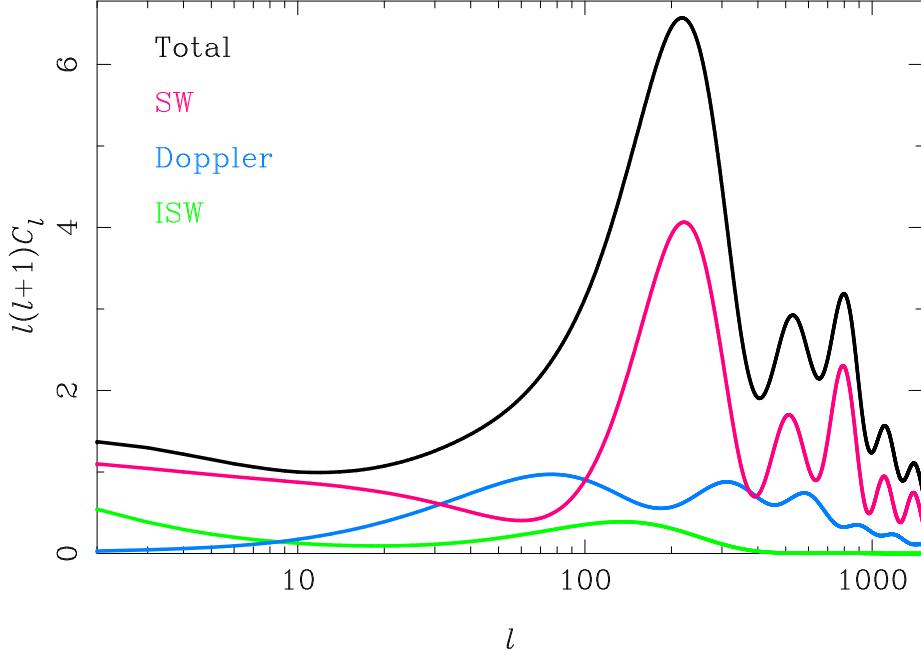


Figure 6: Contributions from the various terms in Eq. (3.18) to the angular power spectrum of the CMB temperature anisotropies. At high l , the contributions are (from top to bottom): total power (black); $\Theta_0 + \psi$ (magenta, denoted SW for Sachs–Wolfe); Doppler contribution from \mathbf{v}_b ; and integrated Sachs–Wolfe (ISW) effect from evolution of the gravitational potentials (green). For the latter, the late-time evolution generates power on large angular scales, while the evolution around recombination, due to the universe not being fully matter dominated there, generates power near the first acoustic peak. Note that this “early-time” ISW effect adds *coherently* with the $\Theta_0 + \psi$ contribution at the first peak, boosting the height of the first peak (and shifting its location a little) further than would be suggested from the sum of the individual power spectra.

some intuition for the physics of the CMB we shall first introduce a simpler, fluid-based approximate approach based on the idea that photons and baryons are tightly-coupled prior to recombination.

The coupling is, of course, not perfect and the mean-free path of the photons grows as the number density of electrons falls due to expansion and, later, recombination. For any given Fourier mode of the perturbations, there are three comoving time/length scales relevant to the dynamics of the pre-recombination plasma:

- the photon mean-free path $l_p = -\dot{\tau}^{-1}$;
- the expansion time $1/\mathcal{H}$;
- the perturbation wavelength k^{-1} .

The *tight-coupling approximation* holds when the mean-free path is much shorter than the other two timescales. For reference, the mean free path of CMB photons prior to recombination is $l_p \approx 4.9 \times 10^4 (\Omega_b h^2)^{-1} (1+z)^{-2}$ Mpc; just before recombination this is around 2 Mpc. In the tight-coupling limit, scattering is so efficient that the photons and baryons behave like a single *photon–baryon fluid*. The photon distribution function is forced to be isotropic in the rest-frame of the baryons, and so the photon bulk velocity \mathbf{v}_γ equals the baryon velocity \mathbf{v}_b .

Form of the distribution function in the tight-coupling limit. We write the Boltzmann equation for scalar perturbations in the form

$$\frac{d}{d\eta} (\Theta + \psi) - (\dot{\phi} + \dot{\psi}) = -a\bar{n}_e \sigma_T \Theta + \frac{3a\bar{n}_e \sigma_T}{16\pi} \int d\hat{\mathbf{m}} \Theta(\hat{\mathbf{m}}) [1 + (\mathbf{e} \cdot \hat{\mathbf{m}})^2] + a\bar{n}_e \sigma_T \mathbf{e} \cdot \mathbf{v}_b. \quad (3.19)$$

When the scattering rate $a\bar{n}_e \sigma_T$ is much larger than the wavenumber k and the Hubble rate \dot{a}/a , if the right-hand side of the Boltzmann equation is to equal the left we must have

$$\Theta(\mathbf{e}) - \mathbf{e} \cdot \mathbf{v}_b - \frac{3}{16\pi} \int d\hat{\mathbf{m}} \Theta(\hat{\mathbf{m}}) [1 + (\mathbf{e} \cdot \hat{\mathbf{m}})^2] \approx 0 \quad (\text{tight coupling}). \quad (3.20)$$

The integral term here has a monopole part, which is just Θ_0 and exactly cancels the monopole of $\Theta(\mathbf{e})$, and a quadrupole part that does not cancel with the quadrupole of $\Theta(\mathbf{e})$. It follows that

$$\Theta(\mathbf{e}) \approx \Theta_0 + \mathbf{e} \cdot \mathbf{v}_b \quad (\text{tight coupling}). \quad (3.21)$$

Extracting the photon bulk velocity with Eq. (2.65) gives $\mathbf{v}_\gamma \approx \mathbf{v}_b$.

The total stress–energy tensor of the photon–baryon fluid is conserved, $\nabla_\mu T^{\mu\nu} = 0$, since it only has gravitational interactions with other matter. As shown in the Part-III *Cosmology* course, for scalar perturbations this leads to the relativistic continuity equation

$$\dot{\delta} + \left(1 + \frac{\bar{P}}{\bar{\rho}}\right) (\nabla \cdot \mathbf{v} - 3\dot{\phi}) + 3\mathcal{H} \left(\frac{\delta P}{\bar{\rho}} - \frac{\bar{P}}{\bar{\rho}}\delta\right) = 0, \quad (3.22)$$

where $\mathcal{H} = \dot{a}/a$ is the conformal Hubble parameter, and the Euler equation for the momentum density $\mathbf{q} = (\bar{\rho} + \bar{P})\mathbf{v}$:

$$\dot{\mathbf{q}} + 4\mathcal{H}\mathbf{q} + \nabla \delta P + (\bar{\rho} + \bar{P})\nabla \psi = 0. \quad (3.23)$$

We have dropped the anisotropic stress here since for the photon–baryon fluid we can ignore it. Generally though, the gradient of the anisotropic stress adds with the gradients of the isotropic pressure and the gravitational potential to determine the force accelerating the fluid.

To apply the Euler equation (3.23) to the photon–baryon fluid, we note that baryons enhance the momentum density through their inertia, while adding negligible pressure support. We have

$$\mathbf{q} = (\bar{\rho} + \bar{P})\mathbf{v} = (\bar{\rho}_\gamma + \bar{P}_\gamma)\mathbf{v}_\gamma + \bar{\rho}_b\mathbf{v}_b \approx \frac{4}{3}(1+R)\bar{\rho}_\gamma\mathbf{v}_\gamma, \quad (3.24)$$

where $R \equiv \bar{\rho}_b/(\bar{\rho}_\gamma + \bar{P}_\gamma)$ and is proportional to the scale-factor a . The Euler equation is therefore

$$\dot{\mathbf{v}}_\gamma + \frac{\mathcal{H}R}{1+R}\mathbf{v}_\gamma + \frac{1}{4(1+R)}\nabla\delta_\gamma + \nabla\psi = 0. \quad (3.25)$$

The second term on the left arises from the effect of expansion on the baryon velocity – if there were no scattering, expansion would cause \mathbf{v}_b to fall as $1/a$, and the effect of this on the tightly-coupled fluid is dependent on the fraction $R/(1+R)$ of the fluid’s momentum density coming from the baryons. The third term on the left is the pressure gradient of the photons suppressed by the baryon inertia. The final term is gravitational infall that, by the equivalence principle, affects the photons and baryons in the same way.

To evolve the photon density, we note that there is no energy exchange between the photons and baryons due to Thomson scattering in linear perturbation theory¹⁰. This follows directly from the monopole component of the Boltzmann equation, as we shall show in Sec. 4. The continuity equation for the photons is therefore the same as for a non-interacting fluid, Eq. (3.22), so that

$$\dot{\delta}_\gamma + \frac{4}{3}\nabla \cdot \mathbf{v}_\gamma - 4\dot{\phi} = 0. \quad (3.26)$$

Combining with the Euler equation, we find

$$\ddot{\delta}_\gamma + \frac{\mathcal{H}R}{1+R}\dot{\delta}_\gamma - \frac{1}{3(1+R)}\nabla^2\delta_\gamma = 4\ddot{\phi} + \frac{4\mathcal{H}R}{1+R}\dot{\phi} + \frac{4}{3}\nabla^2\psi. \quad (3.27)$$

This is of the form of a damped, driven wave equation with (time-dependent) sound speed

$$c_s \equiv \frac{1}{\sqrt{3(1+R)}}. \quad (3.28)$$

¹⁰Generally, there is no energy transfer in Thomson scattering off static electrons. This is not the case for moving electrons; however, this transfer is second order in perturbations. To see this, note that the force on the scattering electron is proportional to the CMB dipole in the electron rest-frame, and the rate of working by this force in a frame in which the electron velocity is \mathbf{v}_b therefore involves the product of the rest-frame dipole and the electron velocity.

This differs from the value for a photon fluid ($c_s^2 = 1/3$) because of baryon inertia. The damping term (i.e., the second term on the left of Eq. 3.27) arises from the free decay of \mathbf{v}_b in an expanding universe, while photon pressure provides the restoring force which is weakly suppressed by the additional inertia of the baryons. The system is gravitationally driven (the terms on the right-hand side). This equation underlies much of the phenomenology of the acoustic peaks in the angular power spectrum of the temperature anisotropies.

3.3.1 Acoustic oscillations

In Fourier space, Eq. (3.27) becomes

$$\ddot{\delta}_\gamma + \frac{\mathcal{H}R}{1+R}\dot{\delta}_\gamma + \frac{1}{3(1+R)}k^2\delta_\gamma = 4\ddot{\phi} + \frac{4\mathcal{H}R}{1+R}\dot{\phi} - \frac{4}{3}k^2\psi. \quad (3.29)$$

On scales $kc_s \gg \mathcal{H}$, this is a lightly-damped oscillator driven by the source-term on the right-hand side. The solutions of the free equation, in the limit of no baryons ($R \rightarrow 0$; we shall discuss the impact of non-zero R shortly), are simply sine and cosine oscillations,

$$\delta_\gamma \propto \sin kr_s \quad \text{or} \quad \delta_\gamma \propto \cos kr_s, \quad (3.30)$$

where the *sound horizon*

$$r_s(\eta) \equiv \int_0^\eta c_s(\eta') d\eta', \quad (3.31)$$

which is $r_s = \eta/\sqrt{3}$ for $R = 0$.

Now consider the effect of the source term in Eq. (3.29). The evolution of the gravitational potential follows from the trace part of the perturbed ij Einstein equations:

$$\ddot{\phi} + \frac{1}{3}\nabla^2(\psi - \phi) + (2\dot{\mathcal{H}} + \mathcal{H}^2)\psi + \mathcal{H}\dot{\psi} + 2\mathcal{H}\dot{\phi} = 4\pi Ga^2\delta P. \quad (3.32)$$

(This was derived in the *Cosmology* course, but only for the case that $\phi = \psi$, as appropriate if the anisotropic stress vanishes.) During matter domination, $\phi = \psi$ and the pressure perturbation is negligible. Since $a \propto \eta^2$, we have solutions $\phi = \text{const.}$ and $\phi \propto \eta^{-5}$, so the fastest growing mode is constant on all scales. For constant potentials, the source term in Eq. (3.29) is constant, and the midpoint of the oscillation is shifted to $\delta_\gamma = -4\psi$. This means that at the midpoint photons are overdense in potential wells ($\psi < 0$), with their pressure support balancing gravity. During radiation domination, things are more complicated since the potentials will generally evolve. Our strategy will therefore be to determine initial conditions for the oscillator equation at the onset of matter domination, after which the source term becomes constant in time.

Initial conditions

The matter era is preceded by the radiation era, so we need to evolve the primordial fluctuations through the radiation era to get initial conditions for the oscillator equation (3.29) at the start of matter domination.

During the radiation era, Eq. (3.29) is not very convenient, since the fluctuations in δ_γ source the gravitational potentials and hence the source term on the right-hand side of the equation. It is simpler to return to the perturbed Einstein equations. In doing so, we make the following approximations.

- We ignore neutrinos. More carefully, we ignore the free-streaming of neutrinos, which otherwise generates anisotropic stress on sub-Hubble scales after neutrino decoupling. Ignoring anisotropic stress allows us to set $\phi = \psi$ (from the trace-free ij Einstein equation).
- We only consider *adiabatic initial conditions*, as appropriate for simple inflation models. Recall that for adiabatic initial conditions, the perturbations in the relative densities of the various components are the same as if one applied a position-dependent time shift to the background universe (note that this is a gauge-independent statement). It follows that the non-adiabatic pressure perturbation $\delta P_{\text{nad}} = \delta P - (\dot{\bar{P}}/\dot{\bar{\rho}})\delta\rho$ vanishes for adiabatic initial conditions. An important consequence is that the fractional density contrasts in photons and neutrinos are equal and comparable to those in the cold dark matter and baryons. The total density contrast is therefore dominated by the major component in the background universe, and during the radiation era is approximately $\delta \approx \delta_\gamma = \delta_\nu$.

With these approximations, we can treat the fluctuations in the radiation era as those in an ideal fluid with equation of state $P = \rho/3$. The δP term in Eq. (3.32) can be replaced by $\delta\rho/3$, which can be eliminated in favour of the potential using the perturbed 00 Einstein equation:

$$\nabla^2\phi = 3\mathcal{H}(\dot{\phi} + \mathcal{H}\psi) + 4\pi Ga^2\delta\rho. \quad (3.33)$$

Putting these pieces together, and using $a \propto \eta$ for radiation domination, we find the following closed equation for the gravitational potential:

$$\ddot{\phi} + \frac{4}{\eta}\dot{\phi} - \frac{1}{3}\nabla^2\phi = 0. \quad (3.34)$$

This is a damped wave equation with sound speed $1/\sqrt{3}$, as expected for a radiation fluid. The solutions in Fourier space are

$$\phi(\eta, \mathbf{k}) = A(\mathbf{k})\frac{j_1(k\eta/\sqrt{3})}{k\eta/\sqrt{3}} + B(\mathbf{k})\frac{n_1(k\eta/\sqrt{3})}{k\eta/\sqrt{3}}, \quad (3.35)$$

where $A(\mathbf{k})$ and $B(\mathbf{k})$ are integration constants, and j_1 and n_1 are the spherical Bessel and Neumann functions, respectively:

$$j_1(x) = \frac{\sin x}{x^2} - \frac{\cos x}{x} = \frac{x}{3} + O(x^3) \quad (3.36)$$

$$n_1(x) = -\frac{\cos x}{x^2} - \frac{\sin x}{x} = -\frac{1}{x^2} + O(x^0). \quad (3.37)$$

Recall that adiabatic initial conditions are best parameterised in terms of the primordial curvature perturbation $\mathcal{R}(\mathbf{k})$, which is constant for adiabatic fluctuations on super-Hubble scales irrespective of the equation of state of matter¹¹. Expressed in terms of the potentials,

$$\mathcal{R} = -\phi - \frac{\mathcal{H}(\dot{\phi} + \mathcal{H}\psi)}{4\pi G a^2(\bar{\rho} + \bar{P})}. \quad (3.38)$$

Deep in the radiation era, the regular solution for ϕ in Eq. (3.35) will dominate and is constant on super-Hubble scales. By Eq. (3.38), this must be related to the primordial curvature perturbation as

$$\mathcal{R}(\mathbf{k}) = -3\phi(\eta, \mathbf{k})/2 \quad (k\eta \ll 1). \quad (3.39)$$

It follows that the solution for ϕ deep in the radiation era is

$$\phi(\eta, \mathbf{k}) = -2\mathcal{R}(\mathbf{k}) \frac{j_1(k\eta/\sqrt{3})}{(k\eta/\sqrt{3})}. \quad (3.40)$$

In the small-scale limit, $k\eta \gg 1$, this reduces to

$$\phi(\eta, \mathbf{k}) \sim 6\mathcal{R}(\mathbf{k}) \frac{\cos(k\eta/\sqrt{3})}{(k\eta)^2} \quad (k\eta \gg 1), \quad (3.41)$$

where we have used the asymptotic form of the spherical Bessel functions:

$$j_l(x) \sim \frac{1}{x} \sin\left(x - \frac{l\pi}{2}\right). \quad (3.42)$$

Notice how the amplitude of the acoustic oscillations in ϕ decay in time as $1/(k\eta)^2$. Defining η_{eq} as the conformal time at matter–radiation equality, and $k_{\text{eq}} = 1/\eta_{\text{eq}}$ the wavenumber of a mode that just enters the Hubble radius at equality, we see that the potential at the start of the matter-dominated era is suppressed for $k \gg k_{\text{eq}}$ by a factor $(k_{\text{eq}}/k)^2$.

The photon density contrast follows from Eq. (3.33). For radiation domination this reduces to

$$\delta_\gamma = -\frac{2}{3}(k\eta)^2\phi - 2\eta\dot{\phi} - 2\phi. \quad (3.43)$$

¹¹The variable $\mathcal{R}(\mathbf{k})$ is the comoving-gauge curvature perturbation introduced in the *Cosmology* course.

On large scales, $k\eta \ll 1$, Eq. (3.40) gives $\eta\dot{\phi} \ll \phi$ and so

$$\delta_\gamma(\eta, \mathbf{k}) \approx -2\phi(\eta, \mathbf{k}) = 4\mathcal{R}(\mathbf{k})/3 \quad (k\eta \ll 1), \quad (3.44)$$

and is constant. On small scales, $k\eta \gg 1$, ϕ oscillates and so $\dot{\phi} \sim k\phi$ giving

$$\delta_\gamma(\eta, \mathbf{k}) \approx -\frac{2}{3}(k\eta)^2\phi(\eta, \mathbf{k}) = -4\mathcal{R}(\mathbf{k}) \cos(k\eta/\sqrt{3}) \quad (k\eta \gg 1). \quad (3.45)$$

We see that well inside the sound horizon during radiation domination, δ_γ oscillates with constant amplitude.

These acoustic oscillations appear as standing waves in real space. All modes of a given k , irrespective of the direction of the wavevector \mathbf{k} , oscillate in phase. This *phase coherence* gives rise to well-defined acoustic peaks in the CMB temperature power spectrum (see Fig. 6). The observation of such acoustic peaks is strong evidence that the fluctuations were laid down in the early universe (as happens for inflation) and have passively evolved since. This contrasts with alternative models for structure formation, such as those based on topological defects that form at phase transitions. Such a network of defects continually sources fluctuations in the matter and radiation and there is no phase coherence between different Fourier modes. Such models were popular before the discovery of acoustic peaks in the CMB, which started to emerge from observations around the mid 1990s. Defects are now limited to play only a sub-dominant role in structure formation.

Acoustic oscillations during matter domination

Consider the oscillator equation (3.29) during matter domination. For now, we shall treat the effect of baryons only crudely by approximating R as a constant, whereas in reality $R \propto a$. Our approximate treatment can be refined with the WKB approximation (see below), but taking $R = \text{const.}$ is enough to illustrate the key effect of baryon inertia on the acoustic oscillations. The general solution of Eq. (3.29) with constant $\phi = \psi$ is

$$\delta_\gamma(\eta, \mathbf{k}) = C(\mathbf{k}) \cos kr_s + D(\mathbf{k}) \sin kr_s - 4(1+R)\psi. \quad (3.46)$$

We want to match this onto the solutions derived above in radiation domination.

There is a subtlety associated with the matter–radiation transition on large scales that we need to account for: on super-Hubble scales with adiabatic initial conditions, the gravitational potential decays by 1/10 through the transition. This follows from constancy of the curvature perturbation and Eq. (3.38), which gives

$$\mathcal{R}(\mathbf{k}) = -5\phi(\eta, \mathbf{k})/3 \quad (\text{matter domination; } k \ll k_{\text{eq}}), \quad (3.47)$$

compared to $\mathcal{R}(\mathbf{k}) = -3\phi(\eta, \mathbf{k})/2$ during radiation domination. The density contrast in the radiation is also modified by the decay in ϕ through the continuity equation (3.26).

On large scales, the Euler equation (3.25) implies that $\mathcal{H}|\mathbf{v}_\gamma| \sim k\psi$, so that $\nabla \cdot \mathbf{v}_\gamma \sim k^2\psi/\mathcal{H} \ll \mathcal{H}\psi$. It follows that the velocity divergence can be neglected in Eq. (3.26) and so

$$\delta_\gamma - 4\phi = \text{const.} \quad (\text{large scales}). \quad (3.48)$$

The constant can be determined from the super-Hubble solutions during radiation domination; it equals $4\mathcal{R}(\mathbf{k})$. Accounting for the decay in ϕ through the matter–radiation transition, we therefore have the initial conditions for the matter era on large scales

$$\delta_\gamma(\eta_{\text{eq}}, \mathbf{k}) = \frac{8}{5}\mathcal{R}(\mathbf{k}) \quad \text{and} \quad \dot{\delta}_\gamma(\eta_{\text{eq}}, \mathbf{k}) = 0 \quad (k \ll k_{\text{eq}}), \quad (3.49)$$

where η_{eq} is after the transition (and decay) is complete.

We can now match onto the general solution, Eq. (3.46), for large-scale modes. Using $\psi(\eta, \mathbf{k}) = -3\mathcal{R}(\mathbf{k})/5$ at all times in the matter era for $k \ll k_{\text{eq}}$, we find $C(\mathbf{k}) = -4\mathcal{R}(\mathbf{k})(1+3R)/5$ and $D(\mathbf{k}) = 0$, so that

$$\begin{aligned} \delta_\gamma(\eta, \mathbf{k}) &= -\frac{4}{5}\mathcal{R}(\mathbf{k})(1+3R)\cos kr_s - 4(1+R)\left[-\frac{3}{5}\mathcal{R}(\mathbf{k})\right] \\ &= -\frac{4}{5}\mathcal{R}(\mathbf{k})[(1+3R)\cos kr_s - 3(1+R)]. \end{aligned} \quad (3.50)$$

This describes an oscillation at frequency kc_s with a midpoint shifted by gravity. At the midpoint, the pressure gradient in the photon–baryon fluid balances gravitational attraction. Since baryons contribute inertia but no pressure, δ_γ at the midpoint increases in magnitude if the amount of baryons is increased (i.e., R increased).

The Sachs–Wolfe contribution to the temperature anisotropies involves the combination $\Theta_0 + \psi$ on the last-scattering surface. Since $\Theta_0 = \delta_\gamma/4$, we have

$$\Theta_0(\eta, \mathbf{k}) + \psi(\eta, \mathbf{k}) = -\frac{1}{5}\mathcal{R}(\mathbf{k})[(1+3R)\cos kr_s - 3R]. \quad (3.51)$$

For $kr_s \ll 1$, we have $\Theta_0 + \psi \approx -\mathcal{R}(\mathbf{k})/5$. Although the photons are initially overdense in potential wells, the Sachs–Wolfe source term is negative there due to the contribution from ψ . As the mode enters the sound horizon, the photons compress further in potential wells reaching a maximum with $\Theta_0 + \psi \approx (1+6R)\mathcal{R}(\mathbf{k})/5$. Pressure support then reverses the compression and at maximum rarefaction $\Theta_0 + \psi \approx -\mathcal{R}(\mathbf{k})/5$ again. Note how at maximum compression the amplitude is enhanced by $1+6R$ relative to maximum rarefaction. The acoustic oscillations in $\Theta_0 + \psi$ are illustrated in Fig. 7. These are based on a full numerical solution, but retain the basic features predicted by our approximate analytic treatment here.

The solution in Eq. (3.50) is valid for $k \ll k_{\text{eq}}$. On smaller scales we have to account for the gravitational driving of the acoustic oscillations during the radiation era. A proper treatment requires a numerical solution, but we can easily obtain the asymptotic

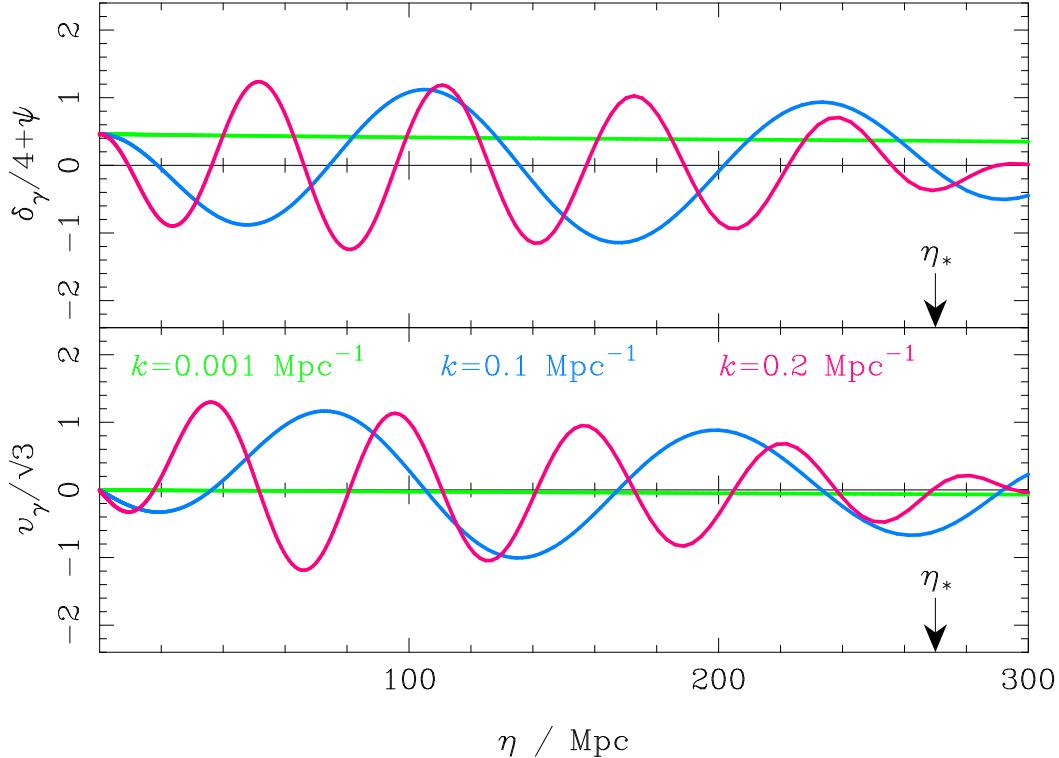


Figure 7: Evolution of the combination $\delta_\gamma/4 + \psi$ (top) and the photon velocity $|\mathbf{v}_\gamma|$ (bottom), which determine the temperature anisotropies produced at last scattering (denoted by the arrow at η_*). Three modes are shown with wavenumbers $k = 0.001$, 0.1 , and 0.2 Mpc^{-1} , and the initial conditions are adiabatic.

solution for $k \gg k_{\text{eq}}$ by matching Eq. (3.46) onto the small-scale solution in radiation domination, Eq. (3.45), and using the fact that the gravitational potential is much smaller than δ_γ since the former decays during the acoustic oscillations in the radiation era. The final result is

$$\delta_\gamma(\eta, \mathbf{k}) \sim -4\mathcal{R}(\mathbf{k}) \cos kr_s \quad (k \gg k_{\text{eq}}). \quad (3.52)$$

We see that modes with $k > k_{\text{eq}}$ oscillate with larger amplitude during matter domination due to the effect of radiation driving during the oscillations in the radiation era.

The photon bulk velocity follows from the continuity equation (3.26). For scalar perturbations, we can write \mathbf{v}_γ as the gradient of a scalar potential, so we can take

$$\mathbf{v}_\gamma(\eta, \mathbf{x}) = \int \frac{d^3\mathbf{k}}{(2\pi)^{3/2}} i\hat{\mathbf{k}} v_\gamma(\eta, \mathbf{k}) e^{i\mathbf{k}\cdot\mathbf{x}}. \quad (3.53)$$

During matter domination, the continuity equation reduces to $v_\gamma = 3\dot{\delta}_\gamma/(4k)$ so that

$$v_\gamma(\eta, \mathbf{k}) = \begin{cases} \frac{3}{5}\mathcal{R}(\mathbf{k})(1+3R)c_s \sin kr_s & (k \ll k_{\text{eq}}) \\ 3\mathcal{R}(\mathbf{k})c_s \sin kr_s & (k \gg k_{\text{eq}}). \end{cases} \quad (3.54)$$

The acoustic oscillations in v_γ are illustrated in Fig. 7, based on a full numerical solution.

Finally, we consider how to refine the crude approximation of treating R as a constant. This can be done by using the WKB approximation to improve the homogeneous solutions of the oscillator equation. As we shall use the WKB approximation in our treatment of diffusion damping in Sec. 4, as a warm-up we derive here the WKB solutions of the homogeneous part of Eq. (3.29) on scales smaller than the sound horizon.

Example: WKB solution of the damped oscillator equation. The homogeneous part of Eq. (3.29) is of the form

$$\ddot{u} + \gamma(\eta)\dot{u} + \omega^2(\eta)u = 0, \quad (3.55)$$

where $\gamma > 0$. There are two timescales in the problem: the expansion timescale and the oscillation period $\omega^{-1} \sim k^{-1}$. We are interested in the adiabatic limit where the expansion time is long compared to the oscillation period. The expansion rate controls the magnitude of the damping γ , its time derivative (since $\dot{\gamma} = O(\mathcal{H}^2) = \gamma^2$), and the time derivative of ω (with $\dot{\omega} \sim \mathcal{H}\omega$ so that $\dot{\omega}/\omega^2 \sim \mathcal{H}/k \ll 1$). The expansion damping is light on sub-horizon scales since $\gamma/\omega \ll 1$.

We first remove the effect of the damping by a field redefinition, writing $u = fv$ for some function f to be determined. In terms of v , the original equation is

$$\ddot{v} + \left(\frac{2\dot{f}}{f} + \gamma \right) \dot{v} + \left(\frac{\ddot{f}}{f} + \frac{\gamma\dot{f}}{f} + \omega^2 \right) v = 0. \quad (3.56)$$

We can eliminate the \dot{v} term by choosing

$$2\dot{f}/f + \gamma = 0 \quad \Rightarrow \quad f = \exp \left(-\frac{1}{2} \int^\eta \gamma d\eta' \right), \quad (3.57)$$

and Eq. (3.56) becomes

$$\ddot{v} + \left(\omega^2 - \frac{1}{4} (2\dot{\gamma} + \gamma^2) \right) v = 0. \quad (3.58)$$

For Eq. (3.27), we have

$$\gamma = \frac{\mathcal{H}R}{1+R} = \frac{d}{d\eta} \ln(1+R), \quad (3.59)$$

so that $f = (1+R)^{-1/2}$. With the restriction $\dot{\gamma} \ll \omega^2$ and $\gamma \ll \omega$, we can ignore the small correction to the frequency of the redefined oscillator in Eq. (3.58).

We now have to solve

$$\ddot{v} + \omega^2 v = 0, \quad (3.60)$$

with $\dot{\omega} \ll \omega^2$. We can do this with the standard WKB approximation that may be familiar from, for example, tunnelling problems in quantum mechanics. We attempt a solution $v = e^{i\delta}$ with $|\ddot{\delta}| \ll \dot{\delta}^2$. Equation (3.60) requires

$$i\ddot{\delta} - \dot{\delta}^2 + \omega^2 = 0. \quad (3.61)$$

If we first neglect $\ddot{\delta}$, we have $\dot{\delta} = \pm\omega$ and so

$$\delta = \pm \int^\eta \omega d\eta'. \quad (3.62)$$

We can use this expression for δ to approximate the $\ddot{\delta}$ term in Eq. (3.61), to obtain a refined solution for δ that satisfies

$$\begin{aligned} \dot{\delta} &= \pm\omega (1 \pm i\dot{\omega}/\omega^2)^{1/2} \\ &\approx \pm\omega + i\dot{\omega}/(2\omega). \end{aligned} \quad (3.63)$$

Integrating this equation gives

$$\begin{aligned} \delta &= \pm \int^\eta \omega d\eta' + i \ln \omega^{1/2} \\ \Rightarrow \quad v &\propto \omega^{-1/2} \exp\left(\pm i \int^\eta \omega d\eta'\right). \end{aligned} \quad (3.64)$$

Putting these pieces together, and noting that $\omega \propto (1+R)^{-1/2}$, we have

$$\delta_\gamma = (1+R)^{-1/4} \cos kr_s, \quad \text{and} \quad \delta_\gamma = (1+R)^{-1/4} \sin kr_s. \quad (3.65)$$

Note that the slow variation of the amplitude of the oscillations derives both from the damping term, which gives a factor $(1+R)^{-1/2}$, and the slow variation of ω^2 , which gives a factor $(1+R)^{1/4}$.

3.4 Large-scale temperature anisotropies

On large angular scales, we can use Eq. (3.51) to determine the Sachs–Wolfe source term $\Theta_0 + \psi$ over the last-scattering surface. In particular, in the limit $kr_s(\eta_*) \ll 1$ the wavelength is so large that modes have not had chance to oscillate before last scattering and $\Theta_0 + \psi \approx -\mathcal{R}/5$ on the last-scattering surface. For such modes, the velocity is suppressed by a factor of $kr_s(\eta_*)$ and makes a negligible contribution to the

temperature anisotropies. If we also neglect the ISW contribution¹², Eq. (3.18) reduces to

$$\Theta(\eta_0, \mathbf{x}_0, \mathbf{e}) + \psi(\eta_0, \mathbf{x}_0) \approx -\frac{1}{5}\mathcal{R}(\mathbf{x}_0 - \chi_*\mathbf{e}). \quad (3.66)$$

This is simply the spherical projection of the primordial curvature perturbation over the last-scattering surface. These large scales provide the most direct view of the primordial fluctuations since there are no evolution effects involved.

The angular power spectrum on large angular scales follows from the general result for the projection of a 3D (scalar) random field given in Eq. (1.33); here we find

$$C_l = \frac{4\pi}{25} \int d\ln k \mathcal{P}_{\mathcal{R}}(k) j_l^2(k\chi_*). \quad (3.67)$$

If we make use of the standard integral

$$\int_0^\infty d\ln x j_l^2(x) = \frac{1}{2l(l+1)} \quad (3.68)$$

for $l > 0$, we see that a scale-invariant primordial spectrum, for which $\mathcal{P}_{\mathcal{R}}(k) = A_s$ is a constant, gives a scale-invariant angular power spectrum

$$\frac{l(l+1)C_l}{2\pi} = \frac{1}{25}A_s. \quad (3.69)$$

More generally, a power-law primordial spectrum

$$\mathcal{P}_{\mathcal{R}}(k) = A_s(k/k_0)^{n_s-1}, \quad (3.70)$$

where k_0 is some pivot scale, gives an angular power spectrum going like

$$C_l \sim \frac{\Gamma(l+n_s/2-1/2)}{\Gamma(l-n_s/2+5/2)} A_s, \quad (3.71)$$

where $\Gamma(x)$ is the Gamma function. We see that the CMB power spectrum on large scales is directly related to the amplitude and slope of the primordial power spectrum (ignoring the late-time ISW effect). For $n_s \approx 1$, we expect a plateau in $l(l+1)C_l$ at small l (see Fig. 6). The latest measurements from Planck indicate that $\mathcal{P}_{\mathcal{R}}(k)$ is close to but not exactly scale-invariant:

$$A_s = (2.198 \pm 0.08) \times 10^{-9} \quad \text{and} \quad n_s = 0.9655 \pm 0.006. \quad (3.72)$$

The departure of n_s from unity is consistent with the prediction of simple inflation models, as discussed in the *Cosmology* course.

¹²The ISW effect can be significant on large angular scales after the potentials start to decay during the late-time accelerated expansion. This adds power incoherently to the Sachs–Wolfe contribution (see Fig. 6).

3.5 Intermediate-scale temperature anisotropies

On smaller angular scales, we probe modes that have had time to oscillate by last scattering. The combination $\Theta_0 + \psi$ on the last-scattering surface can be determined from Eq. (3.50) for $k \ll k_{\text{eq}}$ and Eq. (3.52) for $k \gg k_{\text{eq}}$. A numerical solution is needed to interpolate properly between these regimes, and the result of such a calculation is shown in Fig. 8. This figure does not include the effects of photon diffusion, which, as we shall see in Sec. 4, are actually very important in determining the small-scale power spectrum, to be directly comparable to our approximate analytic solutions that use the tight-coupling approximation. The qualitative features in Fig. 8 do follow our approximate analytic solutions, as follows.

- The source term $\Theta_0 + \psi$ has maximum power at last scattering for those modes that are at extrema of their acoustic oscillations at that time, i.e., those with $kr_s(\eta_*) = n\pi$. Since the anisotropy power at multipole l is dominated by linear modes at last scattering with $k\chi_* \approx l$, we obtain a series of acoustic peaks in the angular power spectrum at $l \approx n\pi\chi_*/r_s(\eta_*)$ in a flat universe. The first acoustic peak, observed to be around $l \approx 220$, corresponds to modes that are just first maximally compressed at last scattering. More generally, the peak locations depend on the angular scale subtended by the sound horizon at last scattering, i.e., the ratio of $r_s(\eta_*)$ to the angular-diameter distance to last scattering, d_A . With the sound horizon calibrated by the determination of the baryon and matter densities from the morphology of the CMB peaks (see below), the peak locations allow one to measure d_A . In spatially-flat Λ CDM models, this translates to a measurement of the Hubble constant, H_0 .
- With baryons, the oscillations in $\Theta_0 + \psi$ are offset from zero by an amount proportional to R for scales where the potential has not completely damped away by oscillations in the radiation era. This produces a baryon-dependent alternation in the peak heights of the angular power spectrum on intermediate scales. In particular, the first (compressional) peak height relative to the second is a sensitive probe of the baryon density.
- Gravitational driving during the radiation era produces a monotonic increase in $\Theta_0 + \psi$ as a function of k at last scattering, once the baryon offset is removed. Modes with larger k start to oscillate earlier in the radiation era and feel a fuller impact of gravitational driving. Since most of the driving effect occurs around horizon entry, modes with $k \gg k_{\text{eq}}$ were fully driven giving a plateau in the amplitude of $\Theta_0 + \psi$ at large k . This imprints in the CMB power spectrum the angular scale of the sound horizon at the time of matter–radiation equality, which allows determination of the matter density $\Omega_m h^2$ from the CMB.

If we neglect the ISW effect (and the quadrupole at last-scattering), the temperature

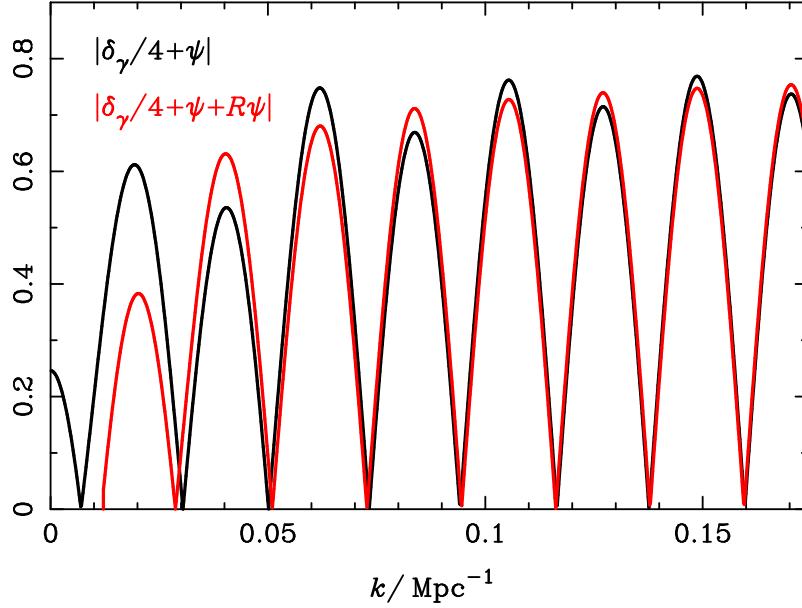


Figure 8: Source term $\Theta_0 + \psi$ (recall, $\delta_\gamma = 4\Theta_0$), evaluated at last scattering, as a function of wavenumber. The effects of diffusion damping have been removed by enforcing the tight-coupling approximation. The black curve shows the modulation of the peak heights due to the baryon offset $-R\psi$. Subtracting the offset (red) reveals the monotonic increase of $\Theta_0 + \psi + R\psi$ with k at last scattering due to gravitational driving during the radiation-dominated era. Since the driving effect is most important through horizon entry, modes that are well inside the horizon at the time of matter–radiation equality are maximally driven and this gives rise to the plateau in the amplitude at large k .

anisotropies are a spherical projection of $\Theta_0 + \psi$ and $\mathbf{e} \cdot \mathbf{v}_\gamma$ on the last-scattering surface of radius χ_* . Using results from Question 1 on Examples II, the angular power spectrum evaluates to

$$C_l \approx 4\pi \int d \ln k \mathcal{P}_{\mathcal{R}}(k) \left[\frac{(\Theta_0 + \psi)(\eta_*, \mathbf{k})}{\mathcal{R}(\mathbf{k})} j_l(k\chi_*) - \frac{v_\gamma(\eta_*, \mathbf{k})}{\mathcal{R}(\mathbf{k})} j'_l(k\chi_*) \right]^2. \quad (3.73)$$

The term in square brackets here is an example of a *transfer function*, linearly relating the multipoles of the temperature anisotropies to the primordial curvature perturbation. Such transfer functions depend on k , due to the scale-dependence of the physics that processes the primordial perturbations, but not on the direction of \mathbf{k} since the physics is isotropic. The Doppler contribution from $v_b \approx v_\gamma$ projects with $j'_l(k\chi_*)$, which does not have the same sharp peak at $l = k\chi_*$ as the spherical Bessel function itself (see Fig. 9). This means that the projection from wavenumber k to multipole l is less sharp for the Doppler term than the Sachs–Wolfe term¹³. The power from

¹³Physically, this arises because for those lines of sight perpendicular to \mathbf{k} (i.e., those for which

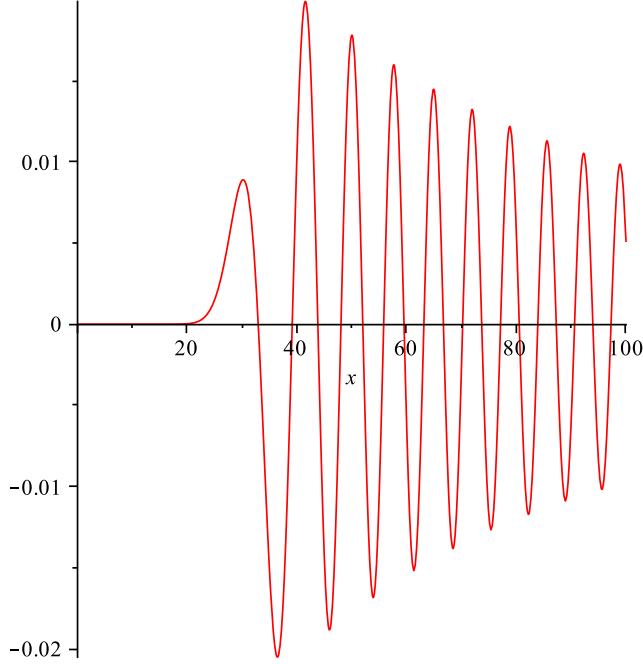


Figure 9: Derivative of the spherical Bessel function, $j'_l(x)$, for $l = 30$.

the Doppler term adds almost incoherently to the Sachs–Wolfe part, but is generally subdominant except at the acoustic troughs where it fills in for the small Sachs–Wolfe contribution (see Fig. 6).

If we drop the Doppler term from Eq. (3.73), we have

$$C_l \approx 4\pi \int d \ln k \mathcal{P}_{\mathcal{R}}(k) \left[\frac{(\Theta_0 + \psi)(\eta_*, \mathbf{k})}{\mathcal{R}(\mathbf{k})} \right]^2 j_l^2(k\chi_*) . \quad (3.74)$$

Roughly, $(\Theta_0 + \psi)(\eta_*, \mathbf{k}) \sim \mathcal{R}(\mathbf{k}) \cos[kr_s(\eta_*)]$. This varies slowly compared to $j_l^2(k\chi_*)$, since $r_s(\eta_*) \approx 145$ Mpc is much smaller than $\chi_* \approx 13890$ Mpc. We can therefore approximate the source function by its value at $k = l/\chi_*$, and perform the remaining integral with Eq. (3.68), to find

$$\frac{l(l+1)C_l}{2\pi} \approx \mathcal{P}_{\mathcal{R}}(k) \left[\frac{(\Theta_0 + \psi)(\eta_*, \mathbf{k})}{\mathcal{R}(\mathbf{k})} \right]^2 \quad (k = l/\chi_*) , \quad (3.75)$$

where the right-hand side is evaluated at $k = l/\chi_*$. We see that the information on the primordial power spectrum $\mathcal{P}_{\mathcal{R}}(k)$ is contained in the broadband shape of the CMB

spatial structure with wavenumber k projects to angular structure with multipole $l = k\chi_*$), the radial projection of the velocity vanishes since $\mathbf{k} \cdot \mathbf{e}$ then equals zero. The velocity source is maximal for lines of sight parallel to \mathbf{k} , but in this case there is no spatial variation of the source in the vicinity of the line of sight since the plane of projection is parallel to a wavefront of the perturbation.

angular power spectrum, but on top of this is imprinted the effects of the acoustic processing of the primordial fluctuations.

4 Scalar perturbations on small scales: diffusion damping and small-scale anisotropies

Equation (3.18) is a good approximation for the temperature anisotropies from perturbations on scales larger than the width of the last-scattering surface (few $\times 10$ Mpc), but even then ignores effects such as re-scattering at reionization. On smaller scales, we have to worry about the non-zero duration of last-scattering since this will tend to blend the contribution from perturbations through the last-scattering surface to the observed temperature anisotropies. An accurate calculation of the CMB anisotropy on all scales where linear perturbation theory is valid requires a full (numerical) solution of the Boltzmann equation.

There is a further issue, related to the breakdown of the tight-coupling approximation on small scales. As we have seen in Sec. 3, on scales large compared to the photon mean-free path, we can treat the baryons and CMB as a tightly-coupled fluid. However, on scales smaller than the mean-free path, photon diffusion becomes important and a self-consistent treatment requires one to work with the Boltzmann equation directly. For these reasons, we begin this section by introducing some useful machinery for working with the Boltzmann equation.

4.1 Machinery for an accurate calculation

We shall be working with the Boltzmann equation, which for scalar perturbations we can write as

$$\frac{\partial \Theta}{\partial \eta} + \mathbf{e} \cdot \nabla \Theta = \dot{\phi} - \mathbf{e} \cdot \nabla \psi + \dot{\tau}(\Theta - \Theta_0) - \frac{3}{4} \dot{\tau} e^{\langle i e^j \rangle} \int \frac{d\hat{\mathbf{m}}}{4\pi} \Theta(\hat{\mathbf{m}}) \hat{m}_{\langle i} \hat{m}_{j \rangle} - \dot{\tau} \mathbf{e} \cdot \mathbf{v}_b, \quad (4.1)$$

where we have used Eq. (3.17). The homogeneity of the background FRW model makes a Fourier expansion very convenient, since the Fourier modes evolve independently, but here we have the added complexity of the angular dependence of Θ . A natural way to deal with the angular dependence is to expand in spherical harmonics. However, for scalar perturbations we can go further than this and note that the perturbations from a single Fourier mode are necessarily *axisymmetric*, and the same will be true of the

temperature anisotropies. We can therefore write

$$\Theta(\eta, \mathbf{x}, \mathbf{e}) = \int \frac{d^3\mathbf{k}}{(2\pi)^{3/2}} \Theta(\eta, \mathbf{k}, \mathbf{e}) e^{i\mathbf{k}\cdot\mathbf{x}} \quad \text{with} \quad \Theta(\eta, \mathbf{k}, \mathbf{e}) = \sum_{l \geq 0} (-i)^l \Theta_l(\eta, \mathbf{k}) P_l(\hat{\mathbf{k}} \cdot \mathbf{e}), \quad (4.2)$$

where $P_l(x)$ are Legendre polynomials. Note that

$$\int \frac{d\mathbf{e}}{4\pi} \Theta(\eta, \mathbf{x}, \mathbf{e}) = \int \frac{d^3\mathbf{k}}{(2\pi)^{3/2}} \Theta_0(\eta, \mathbf{k}) e^{i\mathbf{k}\cdot\mathbf{x}}, \quad (4.3)$$

so that $\Theta_0(\eta, \mathbf{k})$ is simply the Fourier transform of the real-space monopole $\Theta_0(\eta, \mathbf{x})$ that we introduced earlier. More generally, the $\Theta_l(\eta, \mathbf{k})$ are directly related to the spherical multipoles of the Fourier transform of Θ . Writing

$$\Theta(\eta, \mathbf{k}, \mathbf{e}) = \sum_{lm} \Theta_{lm}(\eta, \mathbf{k}) Y_{lm}(\mathbf{e}), \quad (4.4)$$

we have

$$\Theta_{lm}(\eta, \mathbf{k}) = (-i)^l \frac{4\pi}{2l+1} \Theta_l(\eta, \mathbf{k}) Y_{lm}^*(\hat{\mathbf{k}}), \quad (4.5)$$

where we have used the addition theorem

$$P_l(\hat{\mathbf{k}} \cdot \mathbf{e}) = \sum_{|m| \leq l} \frac{4\pi}{2l+1} Y_{lm}^*(\hat{\mathbf{k}}) Y_{lm}(\mathbf{e}). \quad (4.6)$$

We should now check that the assumed axisymmetry of Θ is consistent with the Boltzmann equation.

Consistency of the axisymmetric assumption. Axisymmetry is clearly consistent for terms like $\mathbf{e} \cdot \nabla \Theta$ and $\mathbf{e} \cdot \mathbf{v}_b$ in the Boltzmann equation since, in Fourier space, the latter is proportional to $\mathbf{e} \cdot \hat{\mathbf{k}} v_b(\eta, \mathbf{k})$, where we use the Fourier convention

$$\mathbf{v}_b(\eta, \mathbf{x}) = \int \frac{d^3\mathbf{k}}{(2\pi)^{3/2}} i\hat{\mathbf{k}} v_b(\eta, \mathbf{k}) e^{i\mathbf{k}\cdot\mathbf{x}}, \quad (4.7)$$

for velocities. For the quadrupole contribution to the collision term, we note that in Fourier space this involves

$$\begin{aligned} \int \frac{d\hat{\mathbf{m}}}{4\pi} \Theta(\eta, \mathbf{k}, \hat{\mathbf{m}}) e^{i\hat{e}^j} \hat{m}_{(i} \hat{m}_{j)} &= \int \frac{d\hat{\mathbf{m}}}{4\pi} \Theta(\eta, \mathbf{k}, \hat{\mathbf{m}}) \left((\mathbf{e} \cdot \hat{\mathbf{m}})^2 - \frac{1}{3} \right) \\ &= \frac{2}{3} \int \frac{d\hat{\mathbf{m}}}{4\pi} \Theta(\eta, \mathbf{k}, \hat{\mathbf{m}}) P_2(\mathbf{e} \cdot \hat{\mathbf{m}}) \\ &= \frac{8\pi}{15} \sum_{|m| \leq 2} \int \frac{d\hat{\mathbf{m}}}{4\pi} \Theta(\eta, \mathbf{k}, \hat{\mathbf{m}}) Y_{2m}^*(\hat{\mathbf{m}}) Y_{2m}(\mathbf{e}) \\ &= \frac{2}{15} \sum_{|m| \leq 2} \Theta_{2m}(\eta, \mathbf{k}) Y_{2m}(\mathbf{e}). \end{aligned} \quad (4.8)$$

This result holds generally, irrespective of any assumption about axisymmetry of Θ about the wavevector (i.e., it is not restricted to scalar perturbations). For scalar perturbations, we can substitute for $\Theta_{2m}(\eta, \mathbf{k})$ from Eq. (4.5) to reduce the quadrupole collision term to

$$\begin{aligned} -\frac{3}{4}\dot{\tau}e^{\langle i}e^{j\rangle}\int\frac{d\hat{\mathbf{m}}}{4\pi}\Theta(\eta, \mathbf{k}, \hat{\mathbf{m}})\hat{m}_{\langle i}\hat{m}_{j\rangle} &= \frac{1}{10}\dot{\tau}\Theta_2(\eta, \mathbf{k})\sum_{|m|\leq 2}\frac{4\pi}{5}Y_{2m}^*(\hat{\mathbf{k}})Y_{2m}(\mathbf{e}) \\ &= \frac{1}{10}\dot{\tau}\Theta_2(\eta, \mathbf{k})P_2(\hat{\mathbf{k}} \cdot \mathbf{e}). \end{aligned} \quad (4.9)$$

This is axisymmetric about $\hat{\mathbf{k}}$ as required.

4.2 Boltzmann hierarchy

The Boltzmann equation (4.1) can now be recast as an infinite set of ordinary differential equations for the $\Theta_l(\eta, \mathbf{k})$. In Fourier space, the Boltzmann equation becomes

$$\sum_l(-i)^lP_l(\mu)\left(\dot{\Theta}_l+ik\mu\Theta_l\right)=\sum_l(-i)^lP_l(\mu)\left[\delta_{l0}\dot{\phi}+\delta_{l1}k\psi+\dot{\tau}(1-\delta_{l0})\Theta_l-\frac{1}{10}\dot{\tau}\delta_{l2}\Theta_2+\dot{\tau}\delta_{l1}v_b\right], \quad (4.10)$$

where $\mu \equiv \hat{\mathbf{k}} \cdot \mathbf{e}$ and we have used $P_0(\mu) = 1$ and $P_1(\mu) = \mu$. To avoid clutter, we have left the arguments η and \mathbf{k} of all perturbed variables implicit. If we now use the recursion relation

$$(2l+1)\mu P_l(\mu)=(l+1)P_{l+1}(\mu)+lP_{l-1}(\mu), \quad (4.11)$$

and the orthogonality of the Legendre polynomials, we obtain the *Boltzmann hierarchy* for the Θ_l :

$$\begin{aligned} \dot{\Theta}_l+k\left(\frac{l+1}{2l+3}\Theta_{l+1}-\frac{l}{2l-1}\Theta_{l-1}\right) &= -\dot{\tau}\left[(\delta_{l0}-1)\Theta_l-\delta_{l1}v_b+\frac{1}{10}\delta_{l2}\Theta_2\right] \\ &\quad +\delta_{l0}\dot{\phi}+\delta_{l1}k\psi. \end{aligned} \quad (4.12)$$

This system of ordinary differential equations can be integrated directly with the linearised Einstein equations for the metric perturbations, and the fluid equations governing perturbations in the other matter components, with the stress-energy tensor of the CMB expressed in terms of the $\Theta_l(\eta, \mathbf{k})$ (see below). For *adiabatic* initial conditions, all perturbations in linear perturbation theory are linear in the primordial $\mathcal{R}(\mathbf{k})$, with coefficients that are independent of the direction of \mathbf{k} (because of the isotropy of

the background Robertson–Walker geometry). Therefore, we only need to solve the Boltzmann hierarchy for an appropriate range of k values to determine any observable. For example, writing

$$\begin{aligned}\Theta_{lm}(\eta_0, \mathbf{x}_0) &= \int \frac{d^3\mathbf{k}}{(2\pi)^{3/2}} \Theta_{lm}(\eta_0, \mathbf{k}) e^{i\mathbf{k}\cdot\mathbf{x}_0} \\ &= (-i)^l \frac{4\pi}{2l+1} \int \frac{d^3\mathbf{k}}{(2\pi)^{3/2}} \left[\frac{\Theta_l(\eta_0, \mathbf{k})}{\mathcal{R}(\mathbf{k})} \right] \mathcal{R}(\mathbf{k}) Y_{lm}^*(\hat{\mathbf{k}}) e^{i\mathbf{k}\cdot\mathbf{x}_0},\end{aligned}\quad (4.13)$$

where the ratio of $\Theta(\eta_0, \mathbf{k})$ to $\mathcal{R}(\mathbf{k})$ (an example of a *transfer function*) is real and does not depend on the direction of $\hat{\mathbf{k}}$, the two-point correlation at η_0 and \mathbf{x}_0 becomes

$$\begin{aligned}\langle \Theta_{lm} \Theta_{l'm'}^* \rangle &= (-i)^{l-l'} \frac{(4\pi)^2}{(2l+1)(2l'+1)} \int \frac{d^3\mathbf{k}}{(2\pi)^{3/2}} \frac{d^3\mathbf{k}'}{(2\pi)^{3/2}} \left[\frac{\Theta_l(\eta_0, \mathbf{k})}{\mathcal{R}(\mathbf{k})} \right] \left[\frac{\Theta_{l'}(\eta_0, \mathbf{k}')}{\mathcal{R}(\mathbf{k}')} \right] \\ &\quad \times \underbrace{\langle \mathcal{R}(\mathbf{k}) \mathcal{R}^*(\mathbf{k}') \rangle}_{\frac{2\pi^2}{k^3} \mathcal{P}_{\mathcal{R}}(k) \delta^{(3)}(\mathbf{k}-\mathbf{k}')} Y_{lm}^*(\hat{\mathbf{k}}) Y_{l'm'}(\hat{\mathbf{k}'}) e^{i(\mathbf{k}-\mathbf{k}')\cdot\mathbf{x}} \\ &= (-i)^{l-l'} \frac{4\pi}{(2l+1)(2l'+1)} \int d \ln k \left[\frac{\Theta_l(\eta_0, \mathbf{k})}{\mathcal{R}(\mathbf{k})} \right] \left[\frac{\Theta_{l'}(\eta_0, \mathbf{k}')}{\mathcal{R}(\mathbf{k}')} \right] \mathcal{P}_{\mathcal{R}}(k) \\ &\quad \times \int d\hat{\mathbf{k}} Y_{lm}^*(\hat{\mathbf{k}}) Y_{l'm'}(\hat{\mathbf{k}'}) \\ &= \delta_{ll'} \delta_{mm'} \frac{4\pi}{(2l+1)^2} \int d \ln k \left[\frac{\Theta_l(\eta_0, \mathbf{k})}{\mathcal{R}(\mathbf{k})} \right]^2 \mathcal{P}_{\mathcal{R}}(k).\end{aligned}\quad (4.14)$$

We recover statistically-isotropic correlations in the CMB with angular power spectrum

$$C_l = \frac{4\pi}{(2l+1)^2} \int d \ln k \left[\frac{\Theta_l(\eta_0, \mathbf{k})}{\mathcal{R}(\mathbf{k})} \right]^2 \mathcal{P}_{\mathcal{R}}(k).\quad (4.15)$$

4.3 Line-of-sight solution

The direct solution of the Boltzmann hierarchy was originally used in numerical codes such as the publicly-available COSMICS¹⁴. Accurately solving the required large number of differential equations makes such an approach rather slow. A major step forward was made by Seljak & Zaldarriaga in 1996 with their introduction of the line-of-sight approach, which we discuss below, and associated code CMBFAST. More recent codes (such as CAMB¹⁵) are now very efficient and can compute the power spectrum to multipoles $l_{\max} \sim 2500$ in a matter of seconds on modern computers.

¹⁴<http://web.mit.edu/edbert/>

¹⁵<http://camb.info/>

The line-of-sight approach starts from the scalar source term $S_{\text{scal.}}$ in Eq. (3.14). The Fourier transform of this at time η is

$$S_{\text{scal.}}(\eta, \mathbf{k}, \mathbf{e}) = e^{-\tau(\eta)} \left[\dot{\phi} + \dot{\psi} - \dot{\tau} \left(\Theta_0 + \psi + iv_b P_1(\hat{\mathbf{k}} \cdot \mathbf{e}) - \frac{1}{10} \Theta_2 P_2(\hat{\mathbf{k}} \cdot \mathbf{e}) \right) \right], \quad (4.16)$$

and the temperature anisotropy is, from Eq. (3.15),

$$\Theta(\eta_0, \mathbf{x}_0, \mathbf{e}) + \psi(\eta_0, \mathbf{x}_0) = \int \frac{d^3 \mathbf{k}}{(2\pi)^{3/2}} \int_0^{\eta_0} d\eta' S_{\text{scal.}}(\eta', \mathbf{k}, \mathbf{e}) e^{i\mathbf{k} \cdot (\mathbf{x}_0 - \chi \mathbf{e})}, \quad (4.17)$$

where $\chi \equiv \eta_0 - \eta'$ is the conformal distance back along the line-of-sight. The $e^{-i\chi \mathbf{k} \cdot \mathbf{e}}$ can be expanded with the Rayleigh plane-wave expansion,

$$e^{-i\chi \mathbf{k} \cdot \mathbf{e}} = \sum_l (-i)^l (2l+1) j_l(k\chi) P_l(\hat{\mathbf{k}} \cdot \mathbf{e}), \quad (4.18)$$

and powers of $\hat{\mathbf{k}} \cdot \mathbf{e}$ in the source function can be replaced with derivatives of the spherical Bessel functions, for example

$$\begin{aligned} i(\hat{\mathbf{k}} \cdot \mathbf{e}) e^{-i\chi \mathbf{k} \cdot \mathbf{e}} &= -\frac{d}{d(k\chi)} e^{-i\chi \mathbf{k} \cdot \mathbf{e}} \\ &= -\sum_l (-i)^l (2l+1) j'_l(k\chi) P_l(\hat{\mathbf{k}} \cdot \mathbf{e}), \end{aligned} \quad (4.19)$$

where the prime denotes a derivative with respect to the argument of the spherical Bessel function. Putting these pieces together, we find

$$\begin{aligned} \Theta_l(\eta_0, \mathbf{k}) + \delta_{l0} \psi(\eta_0, \mathbf{k}) &= (2l+1) \int_0^{\eta_0} d\eta' e^{-\tau} \left[(\dot{\phi} + \dot{\psi}) j_l(k\chi) - \dot{\tau} (\Theta_0 + \psi) j_l(k\chi) \right. \\ &\quad \left. + \dot{\tau} v_b j'_l(k\chi) - \frac{1}{20} \dot{\tau} \Theta_2 (3j''_l + j_l)(k\chi) \right]. \end{aligned} \quad (4.20)$$

If all source terms in the integrand on the right-hand side of Eq. (4.20) were known, integrating products of these with the spherical Bessel functions and their derivatives would determine the temperature anisotropy at the observation point. Of course, the source terms are not known and these must be found by solving the Boltzmann equation along with the Einstein equations (and fluid equations). However, only the $l \leq 2$ moments $\Theta_l(\eta, \mathbf{k})$ enter the source term directly, and it is also only these moments that determine the stress-energy tensor of the CMB and hence are needed to solve the Einstein equation. The low- l moments can be obtained accurately by solving the Boltzmann hierarchy truncated at some suitably low l_{\max} ¹⁶, and these can then be used to form the stress-energy tensor and as input sources to the line-of-sight integral (4.20). Moreover, the line-of-sight solution allows the angular power spectrum (which is a smooth function of l) to be evaluated at only selected multipoles, with the full C_l being obtained with a suitable interpolation scheme. This procedure is much quicker than evolving the Boltzmann hierarchy directly to some large l_{\max} .

¹⁶Care must be taken with how the hierarchy is truncated to avoid spurious “reflections”.

4.4 Projection and the Boltzmann hierarchy

It is worth pausing to think physically about why the spherical Bessel functions appear in the line-of-sight integral, and the relation of these to the advection terms on the left-hand side of the Boltzmann hierarchy (4.12) that couple Θ_l to $\Theta_{l\pm 1}$. Consider a plane-wave perturbation $\Theta(\eta, \mathbf{k}, \mathbf{e})$ to the distribution function. Assume the radiation propagates freely for some time interval $\Delta\eta = \chi$. The observed anisotropy at \mathbf{x} at time $\eta + \chi$ is then

$$\Theta(\eta + \chi, \mathbf{x}, \mathbf{e}) = \Theta(\eta, \mathbf{k}, \mathbf{e}) e^{i\mathbf{k}\cdot\mathbf{x}} e^{-i\chi\mathbf{k}\cdot\mathbf{e}}, \quad (4.21)$$

so that

$$\Theta(\eta + \chi, \mathbf{k}, \mathbf{e}) = \sum_l (-i)^l \Theta_l(\eta, \mathbf{k}) P_l(\hat{\mathbf{k}} \cdot \mathbf{e}) e^{-i\chi\mathbf{k}\cdot\mathbf{e}}. \quad (4.22)$$

For a small time interval such that $k\chi \ll 1$ (i.e., the radiation streams only a small fraction of a perturbation wavelength), we can expand the exponential to first-order in $k\chi$ to get

$$\Theta(\eta + \chi, \mathbf{k}, \mathbf{e}) = \sum_l (-i)^l \Theta_l(\eta, \mathbf{k}) P_l(\hat{\mathbf{k}} \cdot \mathbf{e}) (1 - ik\chi \hat{\mathbf{k}} \cdot \mathbf{e}). \quad (4.23)$$

Since $\mu P_l(\mu)$ is a sum of $P_{l\pm 1}(\mu)$, we see that a fraction $O(k\chi)$ of the anisotropy at multipole l free streams to adjacent multipoles in propagating a small fraction $k\chi$ of a wavelength. This is the origin of the advection terms on the left-hand side of the Boltzmann hierarchy (4.12).

Integrating this free-streaming effect over a finite time interval, we instead use the Rayleigh plane-wave expansion for the exponential to get

$$\begin{aligned} \Theta(\eta + \chi, \mathbf{k}, \mathbf{e}) &= \sum_l (-i)^l \Theta_l(\eta, \mathbf{k}) P_l(\hat{\mathbf{k}} \cdot \mathbf{e}) \sum_{l'} (-i)^{l'} (2l' + 1) j_{l'}(k\chi) P_{l'}(\hat{\mathbf{k}} \cdot \mathbf{e}) \\ &= \sum_L (-i)^L P_L(\hat{\mathbf{k}} \cdot \mathbf{e}) \sum_{l'} \left[(-i)^{l+l'-L} (2L+1)(2l'+1) \begin{pmatrix} l & l' & L \\ 0 & 0 & 0 \end{pmatrix}^2 \right. \\ &\quad \left. \times j_{l'}(k\chi) \Theta_l(\eta, \mathbf{k}) \right], \end{aligned} \quad (4.24)$$

where we used Eq. (1.51) to express the product of two Legendre polynomials as a sum over terms involving a $3j$ symbol and a single Legendre polynomial. It follows that in time χ anisotropy at multipole l free-streams to anisotropy at multipole L with

$$\Theta_L(\eta + \chi, \mathbf{k}) \supset \sum_{l'} (-i)^{l+l'-L} (2L+1)(2l'+1) \begin{pmatrix} l & l' & L \\ 0 & 0 & 0 \end{pmatrix}^2 j_{l'}(k\chi) \Theta_l(\eta, \mathbf{k}). \quad (4.25)$$

For the special case of an $l = 0$ source, the sum is restricted to $l' = L$ and we have

$$\Theta_L(\eta + \chi, \mathbf{k}) = (2L+1) j_L(k\chi) \Theta_0(\eta, \mathbf{k}), \quad (4.26)$$

which is equivalent to the projection of the Θ_0 source term in the line-of-sight solution of Eq. (4.20), taking the visibility function to be a delta-function. For an $l = 1$ source (such as the Doppler term in the temperature anisotropies from the baryon velocity), the sum is restricted to $l' = L \pm 1$ by the $3j$ symbol (by the triangle conditions and parity). Using the explicit form of the relevant $3j$ symbols, we have

$$\begin{aligned}\Theta_L(\eta + \chi, \mathbf{k}) &= [L j_{L-1}(k\chi) - (L+1) j_{L+1}(k\chi)] \Theta_1(\eta, \mathbf{k}) \\ &= (2L+1) j'_L(k\chi) \Theta_1(\eta, \mathbf{k}),\end{aligned}\quad (4.27)$$

where we used a recursion relation for spherical Bessel functions.

4.5 Stress–energy tensor

The elements of the stress–energy tensor, δ_γ , δP_γ , \mathbf{v}_γ and Π_γ^{ij} (we include a subscript γ here to distinguish photons from other components) were expressed generally in terms of angular moments of the temperature perturbation in Eqs. (2.64) and (2.65). Here we develop these expressions further for scalar perturbations using the expansion (4.2).

For δ_γ , taking the Fourier transform of Eq. (2.64), we have

$$\begin{aligned}\delta_\gamma(\eta, \mathbf{k}) &= 4 \int \frac{d\mathbf{e}}{4\pi} \sum_l (-i)^l \Theta_l(\eta, \mathbf{k}) P_l(\hat{\mathbf{k}} \cdot \mathbf{e}) \\ &= 4\Theta_0(\eta, \mathbf{k}).\end{aligned}\quad (4.28)$$

It also follows that $\delta P_\gamma = 4\bar{\rho}_\gamma \Theta_0/3$.

For \mathbf{v}_γ , we have

$$\mathbf{v}_\gamma = 3 \int \frac{d\mathbf{e}}{4\pi} \sum_l (-i)^l \Theta_l(\eta, \mathbf{k}) P_l(\hat{\mathbf{k}} \cdot \mathbf{e}) \mathbf{e}. \quad (4.29)$$

The integral over \mathbf{e} must be proportional to $\hat{\mathbf{k}}$ by symmetry, so that

$$\begin{aligned}A_l \hat{\mathbf{k}} &= \int \frac{d\mathbf{e}}{4\pi} P_l(\hat{\mathbf{k}} \cdot \mathbf{e}) \mathbf{e} \\ \Rightarrow A_l &= \int \frac{d\mathbf{e}}{4\pi} P_l(\hat{\mathbf{k}} \cdot \mathbf{e}) \hat{\mathbf{k}} \cdot \mathbf{e} \\ &= \frac{1}{2} \int_{-1}^1 d\mu \mu P_l(\mu) \\ &= \frac{1}{3} \delta_{l1}.\end{aligned}\quad (4.30)$$

It follows that $\mathbf{v}_\gamma(\eta, \mathbf{k}) = -i\hat{\mathbf{k}}\Theta_1(\eta, \mathbf{k})$ and so, following the Fourier convention of Eq. (4.7) for scalar velocity perturbations,

$$v_\gamma(\eta, \mathbf{k}) = -\Theta_1(\eta, \mathbf{k}). \quad (4.31)$$

Finally, for the anisotropic stress we have

$$\Pi^{\hat{i}\hat{j}}(\eta, \mathbf{k}) = -4\bar{\rho}_\gamma \int \frac{d\mathbf{e}}{4\pi} \sum_l (-i)^l \Theta_l(\eta, \mathbf{k}) P_l(\hat{\mathbf{k}} \cdot \mathbf{e}) e^{\langle \hat{i} e^j \rangle}. \quad (4.32)$$

The integral over \mathbf{e} must be a symmetric-trace-free 3-tensor and $\hat{\mathbf{k}}$ is the only preferred direction in the integrand. It follows that the integral must be proportional to $\hat{k}^{\langle i} \hat{k}^{j \rangle}$ so that

$$\begin{aligned} B_l \hat{k}^{\langle i} \hat{k}^{j \rangle} &= \int \frac{d\mathbf{e}}{4\pi} P_l(\hat{\mathbf{k}} \cdot \mathbf{e}) e^{\langle \hat{i} e^j \rangle} \\ \Rightarrow \quad \frac{2B_l}{3} &= \int \frac{d\mathbf{e}}{4\pi} P_l(\hat{\mathbf{k}} \cdot \mathbf{e}) \left((\hat{\mathbf{k}} \cdot \mathbf{e})^2 - \frac{1}{3} \right) \\ &= \frac{1}{3} \int_{-1}^1 d\mu P_l(\mu) P_2(\mu) \\ &= \frac{2}{15} \delta_{l2}. \end{aligned} \quad (4.33)$$

It follows that

$$\Pi_\gamma^{\hat{i}\hat{j}}(\eta, \mathbf{k}) = \frac{4}{5} \bar{\rho}_\gamma \Theta_2(\eta, \mathbf{k}) \hat{k}^{\langle i} \hat{k}^{j \rangle}. \quad (4.34)$$

Adopting the Fourier convention for scalar perturbations that

$$\Pi^{\hat{i}\hat{j}}(\eta, \mathbf{x}) = -(\bar{\rho} + \bar{P}) \int \frac{d^3\mathbf{k}}{(2\pi)^{3/2}} \hat{k}^{\langle i} \hat{k}^{j \rangle} \Pi(\eta, \mathbf{k}) e^{i\mathbf{k} \cdot \mathbf{x}}, \quad (4.35)$$

we have

$$\Pi_\gamma(\eta, \mathbf{k}) = -\frac{3}{5} \Theta_2(\eta, \mathbf{k}). \quad (4.36)$$

For freely-propagating radiation, the stress-energy tensor is conserved, $\nabla_\mu T^{\mu\nu} = 0$, leading to the relativistic continuity and Euler equations. In the presence of scattering, energy and momentum exchange will generally alter these equations. We must instead use the moments of the Boltzmann equation to derive the evolution of δ and \mathbf{v} . We expect to obtain equations like the continuity and Euler equations but with additional scattering terms.

We start with the density. Since $\delta_\gamma(\eta, \mathbf{k}) = 4\Theta_0(\eta, \mathbf{k})$, we use the $l = 0$ moment of the Boltzmann hierarchy [Eq. (4.12)] to find

$$\begin{aligned} \dot{\Theta}_0 + \frac{1}{3} k \Theta_1 - \dot{\phi} &= 0 \\ \Rightarrow \quad \dot{\delta}_\gamma + \frac{4}{3} (-kv_\gamma - 3\dot{\phi}) &= 0, \end{aligned} \quad (4.37)$$

which is simply the Fourier-space version of the continuity equation (3.26) that we used earlier. There, we justified the form of this equation by noting there is no energy

exchange in linear theory due to Thomson scattering off cold electrons. Using the $l = 0$ moment of the Boltzmann equation proves this assertion.

For the evolution of v_γ , we use the $l = 1$ moment of the Boltzmann equation to find

$$\begin{aligned} \dot{\Theta}_1 + k \left(\frac{2}{5} \Theta_2 - \Theta_0 \right) - k\psi &= \dot{\tau}(\Theta_1 + v_b) \\ \Rightarrow \quad \dot{v}_\gamma + \frac{1}{4}k\delta_\gamma + \frac{2}{3}k\Pi_\gamma + k\psi &= \dot{\tau}(v_\gamma - v_b). \end{aligned} \quad (4.38)$$

The left-hand side is exactly the Fourier space version of the usual Euler equation (including the effect of anisotropic stress, which was dropped in Eq. 3.23), but there is an additional scattering term on the right. Since $\dot{\tau} < 0$, this is a drag term that depends on the difference of the bulk velocities of the radiation and the electrons. We see that scattering tries to make the radiation comove with the electrons or, equivalently, force the radiation to have no dipole moment in the electron rest-frame.

4.6 Acoustic oscillations revisited

We can use the machinery developed above to derive rigorously the behaviour of the photons and baryons prior to recombination. While scattering is efficient, we shall see that we recover the photon–baryon fluid of Sec. 3.3. However, the comoving mean-free path $1/|\dot{\tau}|$ scales as a^2 (one factor of a^3 from the electron density and one factor of $1/a$ converting to a comoving length), so small-scale perturbations may reach a time prior to recombination where the mean-free path is comparable to wavelength of the perturbation. The tight-coupling approximation breaks down at this point and photon diffusion damps out the acoustic oscillations. We need the full apparatus of the Boltzmann equation to describe self-consistently both regimes.

4.6.1 Baryon dynamics

We start by considering the evolution of the baryons. These are coupled to the photons by Thomson scattering, but the stress–energy tensor of the combined system of photons and baryons is conserved. Recalling $R \equiv 3\bar{\rho}_b/(4\bar{\rho}_\gamma)$, the total system has an energy density perturbation

$$\delta\rho = \bar{\rho}_\gamma \left(\delta_\gamma + \frac{4}{3}R\delta_b \right), \quad (4.39)$$

and, ignoring baryon pressure, momentum density

$$\mathbf{q} = \frac{4}{3}\bar{\rho}_\gamma(\mathbf{v}_\gamma + R\mathbf{v}_b). \quad (4.40)$$

Using these in the continuity and Euler¹⁷ equations, (3.22) and (3.23), and eliminating time derivatives of δ_γ and \mathbf{v}_γ with Eqs. (4.37) and (4.38), we find

$$\dot{\delta}_b + \nabla \cdot \mathbf{v}_b - 3\dot{\phi} = 0, \quad (4.41)$$

$$\dot{\mathbf{v}}_b + \mathcal{H}\mathbf{v}_b + \nabla\psi = \frac{\dot{\tau}}{R}(\mathbf{v}_b - \mathbf{v}_\gamma). \quad (4.42)$$

The first of these is just the usual continuity equation for pressure-free matter. This is expected since we argued above that Thomson scattering does not lead to energy exchange in linear theory. The left-hand side of the second equation is the usual Euler equation for pressure-free matter, but there is now a scattering term on the right-hand side. This is a drag term that tries to equalise the baryon and photon bulk velocities. Note the presence of $1/R$ in the drag term for the baryon velocity, which is not present in the drag term for the radiation. This arises because, for a given bulk velocity, the momentum density of the baryons is larger than the photons by a factor of R but the momentum exchange due to scattering is equal in magnitude.

4.6.2 Tight-coupling approximation

In Sec. 3.3 we used the lowest-order form of the tight-coupling approximation in which we ignore all $l > 1$ moments of the distribution function of the CMB. More generally, the tight-coupling approximation is a systematic expansion of the collisional Boltzmann equation in the ratio of the scattering time (or photon mean-free path) to the smaller of the perturbation wavelength or the expansion time. The tight-coupling approximation holds on sufficiently large scales that this ratio is small, making the $l > 1$ moments of the distribution function strongly suppressed. This makes physical sense since the efficient scattering wants to make the radiation isotropic in the rest frame of the electrons.

We can see just how suppressed the higher moments of the distribution function are by considering radiation initially with only a monopole and dipole free streaming over one mean-free path. After propagating for l_p , the anisotropy is related to the initial monopole and dipole moments by

$$\Theta_l \sim j_l(kl_p)\Theta_0 + j'_l(kl_p)\Theta_1. \quad (4.43)$$

For $kl_p \ll 1$ (i.e., well in the tight-coupling limit), the arguments of the spherical Bessel functions are small and we can use $j_l(x) = O(x^l)$ for small x to argue that

$$\Theta_l = O(kl_p)^{l-1} \quad (\text{tight-coupling}) \quad (4.44)$$

for $l > 1$.

¹⁷An additional term $-\partial_j \Pi^j_i$ has to be added to the left-hand side of Eq. (3.23) to account for anisotropic stress.

Moreover, the photon and baryon bulk velocities are almost equal when tight coupling holds. To see this explicitly, consider the *slip velocity* $\mathbf{v}_{\gamma-b} \equiv \mathbf{v}_\gamma - \mathbf{v}_b$. The Euler equations for the baryons and photons imply that

$$\dot{v}_{\gamma-b} = \left(\frac{1+R}{R} \right) \dot{\tau} v_{\gamma-b} - \frac{1}{4} k \delta_\gamma - \frac{2}{3} k \Pi_\gamma + \mathcal{H} v_b \quad (4.45)$$

in Fourier space. By considering the order of magnitude of the various terms, and noting that the timescale for variation of the slip velocity is the shorter of k^{-1} and \mathcal{H}^{-1} , we see that

$$\left(\frac{1+R}{R} \right) v_{\gamma-b} \approx l_p \left(\mathcal{H} v_b - \frac{1}{4} k \delta_\gamma \right) \quad (4.46)$$

in the tightly-coupled limit, and so the relative velocity is suppressed by one factor of l_p .

To lowest order in tight-coupling, we take $\mathbf{v}_\gamma = \mathbf{v}_b$ and ignore the photon anisotropic stress. Adding R times the Euler equation (4.42) for the baryons to the Euler equation (4.38) for the photons, recovers the Euler equation (3.25) for the photon–baryon fluid that we used in our earlier treatment of acoustic oscillations.

4.6.3 Beyond tight-coupling: diffusion damping

On small scales the tight-coupling approximation breaks down. A new effect – photon diffusion – then becomes important. We can think of a photon as random walking from collision to collision, with a typical step length given by the mean free path. In a short interval $\Delta\eta$, over which we can treat the mean free path as constant, the photon will random walk a mean-squared distance $l_p^2 N$ where the number of collisions is $N = \Delta\eta/l_p$. Integrating up from $\eta = 0$ gives the squared diffusion length

$$k_D^{-2} \sim \int_0^\eta |\dot{\tau}|^{-1} d\eta', \quad (4.47)$$

where we have used $l_p = -\dot{\tau}^{-1}$. The diffusion length is roughly the geometric mean of the conformal age of the universe and the mean free path. Approximating the universe as matter-dominated before recombination, the squared diffusion length scales as $a^{5/2}$ with value

$$k_D^{-2} \sim \frac{2l_p(\eta_*)}{5\mathcal{H}(\eta_*)} \sim 20 \text{ Mpc}^2 \quad (4.48)$$

at last scattering. As we shall see, the effect of diffusion is to damp the photon (and baryon) acoustic oscillations exponentially by the time of last scattering for $k > k_D$.

To describe diffusion damping more quantitatively, we develop an expansion in the dimensionless parameters kl_p and $\mathcal{H}l_p$. In the leading-order tight-coupling approximation used in Sec. 3.3, we work to zero order in these parameters. We now extend this to first order.

Tight-coupling only breaks down before recombination for modes that are sub-Hubble, so we can ignore the effects of expansion (i.e., $\mathcal{H}v_b \ll \dot{v}_b$) and gravity (since $-k^2\phi \sim \mathcal{H}^2\delta$ and so infall terms $k\psi$ are negligible compared to the photon pressure gradients and \dot{v} for the photons and baryons). For sub-Hubble modes, the Euler equations for the photons and baryons reduce to

$$\dot{v}_\gamma + \frac{1}{4}k\delta_\gamma + \frac{2}{3}k\Pi_\gamma = \dot{\tau}v_{\gamma-b} \quad (4.49)$$

$$R\dot{v}_b = -\dot{\tau}v_{\gamma-b}, \quad (4.50)$$

and Eq. (4.46) for the slip velocity $v_{\gamma-b}$ becomes

$$(1 + R^{-1})v_{\gamma-b} = \frac{1}{4}k\dot{\tau}^{-1}\delta_\gamma. \quad (4.51)$$

This last equation is correct to first-order in the tight-coupling parameters. The slip velocity describes the bulk velocity of the photons in the rest-frame of the baryons, and hence the energy flux in that frame. Since $\rho_\gamma \propto T_\gamma^4$, we see that the right-hand side of Eq. (4.51) is of the form $-l_p \nabla T_\gamma / \bar{T}_\gamma$. This describes the generation of an energy flux by gradients in the photon temperature, i.e., *thermal conduction*.

We can add Eqs (4.49) and (4.50) to eliminate the scattering terms, giving

$$R\dot{v}_b + \dot{v}_\gamma + \frac{1}{4}k\delta_\gamma + \frac{2}{3}k\Pi_\gamma = 0. \quad (4.52)$$

We now combine this with $\dot{v}_b = \dot{v}_\gamma - \dot{v}_{\gamma-b}$, and eliminate $\dot{v}_{\gamma-b}$ with the time derivative of the conduction equation (4.51), to find¹⁸

$$(1 + R)\dot{v}_\gamma = -\frac{1}{4}k\delta_\gamma - \frac{R^2}{4(1 + R)}k|\dot{\tau}^{-1}|\dot{\delta}_\gamma - \frac{2}{3}k\Pi_\gamma. \quad (4.53)$$

The second and third term on the right are first order in the tight-coupling parameters. We can find an explicit expression for the photon anisotropic stress to first order in tight coupling by using the $l = 2$ moment of the Boltzmann equation, which we write in the form

$$\Theta_2 = \frac{10}{9}\dot{\tau}^{-1} \left(\dot{\Theta}_2 + \frac{3}{7}k\Theta_3 - \frac{2}{3}k\Theta_1 \right). \quad (4.54)$$

To first order in the tight-coupling parameters, we therefore have¹⁹

$$\Theta_2 = -\frac{20}{27}k\dot{\tau}^{-1}\Theta_1 \quad \Rightarrow \quad \Pi_\gamma = -\frac{4}{9}k\dot{\tau}^{-1}v_\gamma. \quad (4.55)$$

¹⁸In taking the derivative of Eq. (4.51), we can ignore the variation of R and $\dot{\tau}$ on sub-Hubble scales well away from recombination.

¹⁹There are 20% corrections to this result from polarization that we are ignoring here.

This describes the generation of photon anisotropic stress from spatial derivatives of the bulk velocity, i.e., an effective photon viscosity. Using Eq. (4.55) in Eq. (4.53), and noting that $\dot{\delta}_\gamma = 4kv_\gamma/3$ (ignoring the gravitational redshifting effect), we have

$$(1+R)\dot{v}_\gamma = -\frac{1}{4}k\delta_\gamma - \frac{1}{4}\left(\frac{R^2}{1+R} + \frac{8}{9}\right)k|\dot{\tau}^{-1}|\dot{\delta}_\gamma. \quad (4.56)$$

At zero order in tight-coupling, this is just the small-scale limit of Eq. (3.25). The corrections at first order in tight-coupling are damping terms arising from photon viscosity and thermal conduction.

Finally, we differentiate the continuity equation $\dot{\delta}_\gamma \approx 4kv_\gamma/3$, and eliminate \dot{v}_γ with Eq. (4.56), to find the oscillator equation to first order in tight coupling on small scales:

$$\ddot{\delta}_\gamma + \frac{k^2|\dot{\tau}^{-1}|}{3(1+R)}\left(\frac{8}{9} + \frac{R^2}{1+R}\right)\dot{\delta}_\gamma + \frac{k^2}{3(1+R)}\delta_\gamma = 0. \quad (4.57)$$

We can solve this with the WKB approximation (see Sec. 3.3). We eliminate $\dot{\delta}_\gamma$ with

$$f = e^{-k^2/k_D^2} \quad \text{where} \quad \frac{1}{k_D^2} = \frac{1}{6} \int_0^\eta d\eta' \frac{|\dot{\tau}^{-1}|}{1+R} \left(\frac{8}{9} + \frac{R^2}{1+R}\right), \quad (4.58)$$

so that the lowest-order WKB solution (ignoring adiabatic decay of the amplitude due to expansion) is

$$\delta_\gamma \propto e^{\pm ikr_s(\eta)}e^{-k^2/k_D^2}. \quad (4.59)$$

The diffusion damping scale here is close to the rough estimate of Eq. (4.47).

We see that the effect of diffusion is to introduce an exponential suppression of the amplitude of the acoustic oscillations for $k \gg k_D(\eta)$. This behaviour is illustrated in Fig. 10. The amplitude of the source term $\Theta_0 + \psi$ on the last-scattering surface is Gaussian-suppressed in k , and this leads to a strong suppression of the angular power spectrum on small scales with a characteristic multipole $l_D = k_D\chi_* \approx 2000$. This *damping tail* can be seen in Fig. 4, and is illustrated further in Fig. 11, which compares the angular power spectrum from a full calculation with that from making the assumption that tight-coupling holds all the way to last scattering.

The finite mean-free path of CMB photons around last scattering has an additional effect on the temperature anisotropies. The visibility function $-\dot{\tau}e^{-\tau}$ has a finite width of a few $\times 10$ Mpc and so along a given line of sight photons will be last scattered over this interval. Averaging over scattering events will tend to wash out the anisotropy from wavelengths short compared to the width of the visibility function. This effect is described mathematically in the line-of-sight solution (Eq. 4.20) by integrating the oscillations in the spherical Bessel functions against the product of the visibility function and the (damped) perturbations.

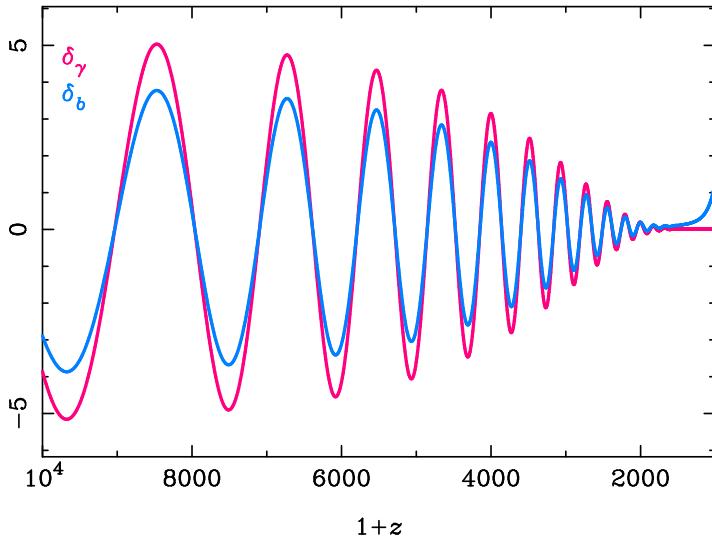


Figure 10: Evolution of the photon (magenta) and baryon (blue) overdensities showing the effect of diffusion damping. Note that the baryons track the photons until around recombination, at which point they decouple and fall into the potential wells to track the cold dark matter.

Boltzmann codes such as CMBFAST and CAMB use the tight-coupling approximation at early times to avoid the numerical problems associated with integrating the stiff Euler equations in their original forms.

4.7 Physics from the CMB damping tail (non-examinable)

We end our discussion of diffusion damping by noting that the damping tail can give strong constraints on some extensions to the Λ CDM model. We focus here on the question of whether there is evidence for additional relativistic degrees of freedom at recombination beyond the photons and three families of neutrinos expected from the Standard Model. A concrete example is models with *sterile neutrinos* (e.g., right-handed neutrinos) that do not take part in weak interactions.

The effective number of (non-photonic) relativistic degrees of freedom is usually described in terms of a parameter N_{eff} defined so that their energy density is related to the photon energy density by

$$\bar{\rho}_\nu = N_{\text{eff}}(7/8)(4/11)^{4/3}\bar{\rho}_\gamma. \quad (4.60)$$

Note that we use the subscript ν here, although any additional degrees of freedom need

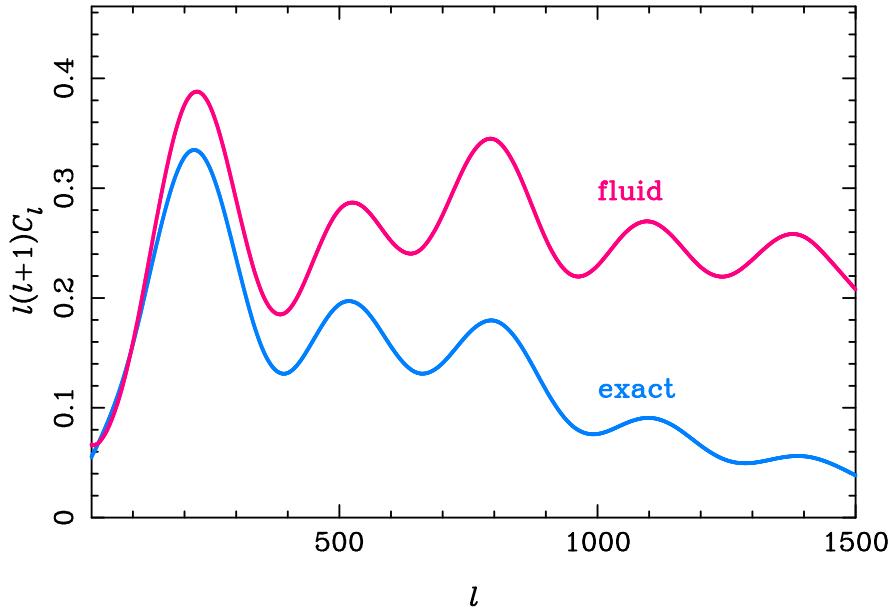


Figure 11: Comparison of the angular power spectrum from a full numerical calculation (“exact”) and that from a numerical calculation employing the tight-coupling approximation until last scattering (“fluid”).

not be neutrinos. The point is that, irrespective of what they are, they act *dynamically* like additional massless neutrinos. In models with only standard neutrinos, $N_{\text{eff}} = 3$, with the factor of $7/8$ in Eq. (4.60) arising from the difference in Fermi and Bose statistics and the factor of $(4/11)^{4/3}$ from the $(4/11)^{1/3}$ temperature difference between neutrinos and photons due to electron–positron annihilation. In practice, neutrinos are not fully decoupled at electron–positron annihilation and are reheated a little giving $N_{\text{eff}} = 3.046$ for standard neutrinos.

We now consider the effect of increasing N_{eff} on the CMB power spectrum. In particular, we are interested in whether there is any unique signature of N_{eff} that cannot be offset in the spectrum by varying other parameters. The main effects of increasing N_{eff} are as follows.

- It increases the expansion rate in the radiation-dominated era and hence reduces the time to recombination and so $r_s(\eta_*)$. This shifts the acoustic peaks to smaller angular scales, but this can be offset (in flat models) by reducing the distance to last scattering by increasing H_0 (or, equivalently, increasing the amount of dark energy).
- Matter–radiation equality occurs at lower redshift. Recall from the *Cosmology* course that the constancy of the comoving-gauge curvature perturbation \mathcal{R} for adiabatic perturbations on super-Hubble scales implies that the gravitational po-

tentials decay (by a factor of 9/10) through the matter–radiation transition. This decay is not fully complete by recombination and the evolution of the potentials leads to a contribution from the last term in Eq. (4.6.2) around recombination (the *early-time integrated Sachs–Wolfe effect*) that contributes coherently with the $\Theta_0 + \psi$ source around the first peak. (Look back at Fig. 6 to see the integrated Sachs–Wolfe contribution. The effect on large scales is from the decay of the potentials once dark energy dominates, but the effect around the first peak is from the decay around recombination.) Reducing the redshift of matter–radiation equality increases the early-time integrated Sachs–Wolfe effect, but this can be offset by an increase in the CDM density to preserve the redshift of matter–radiation equality.

- If we only change N_{eff} , the diffusion distance is reduced since there is less time to recombination. Increasing the CDM density to preserve the redshift of matter–radiation equality further reduces the diffusion distance.

We might expect that the reduction in the damping scale would imply an increase in the CMB power spectrum in the damping tail. However, this does not take account of the change in H_0 to preserve the peak locations. What really matters is the angle θ_D that the diffusion scale subtends at last scattering. If this increases, a given multipole l is more damped and the power spectrum is reduced. Since we are fixing the angular scale of the acoustic peaks (usually denoted θ_*), we consider the ratio of θ_D to θ_* :

$$\frac{\theta_D}{\theta_*} = \frac{1}{r_s(\eta_*) k_D(\eta_*)}. \quad (4.61)$$

Note how the distance to last scattering cancels in this ratio. Since the effect of increasing N_{eff} (and $\Omega_c h^2$) is to increase the expansion rate before recombination, let us work in terms of the (proper) expansion rate H_* at last scattering. The *squared* diffusion distance increases linearly with the available time to diffuse, hence $k_D(\eta_*) \propto H_*^{1/2}$, but the sound horizon increases linearly with time so $r_s(\eta_*) \propto H_*^{-1}$. Combining these gives²⁰

$$\frac{\theta_D}{\theta_*} \propto \frac{1}{H_*^{-1} H_*^{1/2}} \propto H_*^{1/2}. \quad (4.62)$$

It follows that increasing N_{eff} while preserving the peak locations increases θ_D and reduces the power in the damping tail. This unique signature of extra relativistic degrees of freedom is illustrated in Fig. 12.

²⁰A similar argument applies to the angular size θ_{eq} the horizon at the matter–radiation transition subtends at last scattering, which determines the characteristic angular scale associated with resonant driving of the acoustic oscillations during radiation domination. With fixed redshift of matter–radiation equality, both H_{eq}^{-1} and $r_s(\eta_*)$ respond proportionately to changes in the radiation density and so the ratio $\theta_{\text{eq}}/\theta_*$ is fixed.

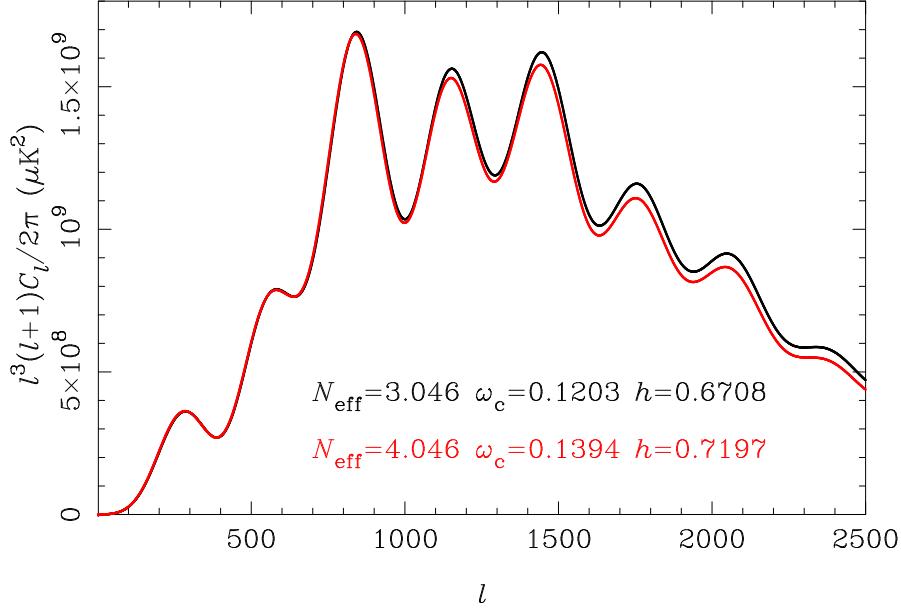


Figure 12: Effect of increasing N_{eff} on the CMB damping tail. The model in black is the best-fit Λ CDM model to the Planck data with the standard value $N_{\text{eff}} = 3.046$. The model in red has $N_{\text{eff}} = 4.046$, and larger H_0 (here parameterised as $H_0 = 100h \text{ km s}^{-1} \text{ Mpc}^{-1}$) to preserve the peak locations and larger $\Omega_c h^2 = \omega_c$ to preserve the redshift of matter–radiation equality. These models can be distinguished by their differing power in the damping tail. Note that we plot $l^3(l+1)C_l/(2\pi)$ here to emphasise the high- l part of the spectrum.

The latest result (from Planck) on N_{eff} in Λ CDM + N_{eff} models is

$$N_{\text{eff}} = 3.13^{+0.30}_{-0.34}, \quad (4.63)$$

where the errors are the 68% confidence interval. This is clearly consistent with the standard value. The constraints on N_{eff} are positively correlated with constraints on the spectral index n_s , since the reduction in small-scale power from an increase in N_{eff} can be mitigated somewhat with an increase in n_s .

5 CMB anisotropies from gravitational waves

It was noted briefly in the *Cosmology* course that a spectrum of primordial gravitational waves is arguably the cleanest probe of inflation. In this section we shall see how gravitational waves affect the CMB temperature anisotropies, and how this can be used to constrain the primordial power spectrum of gravitational waves.

Recall that gravitational waves are tensor perturbations to the metric:

$$ds^2 = a^2(\eta) [-d\eta^2 + (\delta_{ij} + h_{ij}) dx^i dx^j] , \quad (5.1)$$

with $h_i^i = 0$ and $\partial_i h_j^i = 0$. We prefer to use h_{ij} for the tensor metric perturbation than the notation $2\hat{E}_{ij}$ (or $2E_{ij}^T$) used in the *Cosmology* course. In linear theory, h_{ij} is gauge-invariant.

The coordinate system in Eq. (5.1) is synchronous and so observers following worldlines of constant x^i are geodesic (i.e., free-falling) and their 4-velocity is normal to surfaces of constant time. The physical separation between these geodesic observers grows as the scale factor in the absence of perturbations. Factoring out this effect, the physical separation in a perturbed model is determined by the metric perturbation h_{ij} , and variation of this separation in time is the observable manifestation of gravitational waves. Since h_{ij} is trace-free, distortions in the separation of free-falling observers are volume preserving and so the induced motions are pure shear.

The conditions $h_i^i = 0$ and $\partial_i h_j^i = 0$ mean there are two degrees of freedom associated with gravitational waves, i.e., two polarization states. Often the $+$ and \times polarization states are used, in which the shear is along the x and y -directions, or at 45° , for a wave propagating along the z -direction. However, here it will be more convenient to use *circularly-polarized* modes. This means we decompose h_{ij} in Fourier modes as follows

$$h_{ij} = \sum_{\pm} \int \frac{d^3 \mathbf{k}}{(2\pi)^{3/2}} h_{ij}^{(\pm 2)}(\eta, \mathbf{k}) e^{i\mathbf{k}\cdot\mathbf{x}} , \quad \text{where} \quad h_{ij}^{(\pm 2)}(\eta, \mathbf{k}) = \frac{1}{\sqrt{2}} m_{ij}^{(\pm 2)}(\hat{\mathbf{k}}) h^{(\pm 2)}(\eta, \mathbf{k}) . \quad (5.2)$$

For \mathbf{k} along $\hat{\mathbf{z}}$, the circularly-polarized basis tensors are

$$m^{(\pm 2)}(\hat{\mathbf{z}}) = \frac{1}{2} (\hat{\mathbf{x}} \pm i\hat{\mathbf{y}}) \otimes (\hat{\mathbf{x}} \pm i\hat{\mathbf{y}}) , \quad (5.3)$$

while for a general \mathbf{k} with polar coordinates $\theta_{\mathbf{k}}$ and $\phi_{\mathbf{k}}$, we rotate by $D(\phi_{\mathbf{k}}, \theta_{\mathbf{k}}, 0)$ (so that $\hat{\mathbf{x}}$ is along the $\theta_{\mathbf{k}}$ -direction and $\hat{\mathbf{y}}$ is along $\phi_{\mathbf{k}}$). The basis tensors satisfy the following properties

$$m_{ij}^{(p)}(\hat{\mathbf{k}}) \left[m^{(p')ij}(\hat{\mathbf{k}}) \right]^* = \delta^{pp'} , \quad (5.4)$$

and

$$\left[m_{ij}^{(\pm)}(\hat{\mathbf{k}}) \right]^* = m_{ij}^{(\mp)}(\hat{\mathbf{k}}) = m_{ij}^{(\pm)}(-\hat{\mathbf{k}}) , \quad (5.5)$$

so reality of $h_{ij}(\eta, \mathbf{x})$ implies that $[h^{(\pm 2)}(\eta, \mathbf{k})]^* = h^{(\pm 2)}(\eta, -\mathbf{k})$.

Under active translations and rotations, the $h^{(\pm 2)}(\eta, \mathbf{k})$ transform as (exercise!)

$$\begin{aligned} h^{(\pm 2)}(\eta, \mathbf{k}) &\rightarrow h^{(\pm 2)}(\eta, \mathbf{k}) e^{-i\mathbf{k}\cdot\mathbf{a}} && \text{(translation by } \mathbf{a} \text{)} \\ h^{(\pm 2)}(\eta, \mathbf{k}) &\rightarrow e^{\mp 2i\psi} h^{(\pm 2)}(\eta, \mathbf{R}^{-1}\mathbf{k}) && \text{(rotation by } \mathbf{R} \text{),} \end{aligned} \quad (5.6)$$

where \mathbf{R} is a 3×3 rotation matrix and the angle ψ depends on \mathbf{k} and the particular rotation²¹. Assuming statistical invariance under these transformations (and parity also), the 2-point correlator takes the form

$$\langle h^{(p)}(\mathbf{k})[h^{(p')}(\mathbf{k}')]^* \rangle = \frac{2\pi^2}{k^3} \mathcal{P}_h(k) \delta^{(3)}(\mathbf{k} - \mathbf{k}') \delta^{pp'}, \quad (5.7)$$

where we have left the time dependence implicit. The dimensionless power spectrum $\mathcal{P}_h(k)$ is the same for both polarization states by parity invariance. The variance of h_{ij} in real-space is

$$\langle h_{ik}(\mathbf{x})h^{ij}(\mathbf{x}) \rangle = \int d\ln k \mathcal{P}_h(k). \quad (5.8)$$

5.1 Gravitational waves from inflation

The calculation of gravitational waves from inflation was sketched in the *Cosmology* course. The quadratic action for tensor perturbations can be shown to be²²

$$S^{(2)} = \frac{M_{\text{Pl}}^2}{8} \int d\eta d^3\mathbf{x} a^2 \left(\dot{h}_{ij} \dot{h}^{ij} - \partial_i h_{jk} \partial^i h^{jk} \right), \quad (5.9)$$

where $M_{\text{Pl}} = 1/\sqrt{8\pi G}$ is the reduced Planck mass. If we use Eq. (5.2) to expand h_{ij} , the action becomes

$$S^{(2)} = \frac{M_{\text{Pl}}^2}{8} \frac{1}{2} \sum_p \int d\eta d^3\mathbf{k} a^2 \left[(\dot{h}^{(p)})^2 + k^2 (h^{(p)})^2 \right]. \quad (5.10)$$

If we compare with the action for a massless scalar field $\delta\phi$,

$$\begin{aligned} S^{(2)} &= \frac{1}{2} \int d\eta d^3\mathbf{x} a^2 \left[(\partial_\eta \delta\phi)^2 - (\nabla \delta\phi)^2 \right] \\ &= \frac{1}{2} \int d\eta d^3\mathbf{k} a^2 \left[(\partial_\eta \delta\phi)^2 + k^2 (\delta\phi)^2 \right], \end{aligned}$$

we see that each polarization of gravitational waves behaves like an independent massless scalar field with $\delta\phi^{(p)} = M_{\text{Pl}} h^{(p)}/\sqrt{8}$. Each polarization therefore develops independent quantum fluctuations that freeze out on super-Hubble scales with (primordial) power spectra

$$\mathcal{P}_h(k) = \frac{8}{M_{\text{Pl}}^2} \left(\frac{H_k}{2\pi} \right)^2, \quad (5.11)$$

²¹To see the need for the phase factor $e^{\mp 2i\psi}$, consider rotating about the z -axis a Fourier mode with \mathbf{k} along the z -direction.

²²If you want to verify this using the metric in the form of Eq. (5.1), you need to be careful to include the contribution that is quadratic in h_{ij} to the metric determinant in the action for the matter sector. With the Friedmann equations, this contribution cancels with part of the $\sqrt{-g}R$ (i.e., Einstein–Hilbert) term in the action leaving Eq. (5.9) up to a total derivative.

where H_k is the Hubble parameter during inflation when the mode of wavenumber k exits the horizon (i.e., when $k = aH$).

The production of gravitational waves is a very robust prediction of inflation. The primordial power spectrum depends only on the (square of the) expansion rate. Introducing the *energy scale of inflation* E_{inf} , via the Friedmann equation²³

$$H^2 = \frac{E_{\text{inf}}^4}{3M_{\text{Pl}}^2}, \quad (5.12)$$

we see that

$$\mathcal{P}_h(k) = \frac{128}{3} \left(\frac{E_{\text{inf}}}{\sqrt{8\pi} M_{\text{Pl}}} \right)^4 = 1.93 \times 10^{-11} \left(\frac{E_{\text{inf}}}{10^{16} \text{ GeV}} \right)^4. \quad (5.13)$$

Observational constraints on $\mathcal{P}_h(k)$ are usually quoted in terms of the *tensor-to-scalar ratio* r , defined as the ratio of \mathcal{P}_h to the power spectrum of curvature perturbations, $\mathcal{P}_{\mathcal{R}}$, at some pivot scale k_* that is typically taken to be $k_* = 0.002 \text{ Mpc}^{-1}$. Recalling the result (derived in the *Cosmology* course) for canonical, single-field inflation

$$\mathcal{P}_{\mathcal{R}}(k) = \frac{1}{2M_{\text{Pl}}^2 \epsilon} \left(\frac{H_k}{2\pi} \right)^2, \quad \text{where } \epsilon \equiv -\frac{1}{H^2} \frac{dH}{dt}, \quad (5.14)$$

we see that

$$r = 16\epsilon, \quad (5.15)$$

and so r depends on the extent to which inflation departs from exact de Sitter dynamics.

The spectrum $\mathcal{P}_h(k)$ is almost scale-invariant for the quasi-de Sitter expansion of slow-roll inflation. The tensor spectral index is

$$n_t \equiv \frac{d \ln \mathcal{P}_h(k)}{d \ln k} \approx -2\epsilon, \quad (5.16)$$

to first order in ϵ , and so is close to zero. Since the Hubble parameter necessarily decreases during inflation, $\epsilon > 0$ and $n_t < 0$. For canonical, single-field inflation, we see that r and n_t are related by the consistency relation

$$r = -8n_t. \quad (5.17)$$

Verifying this relation would be a remarkable vindication of the simplest inflation models, but this is extremely challenging observationally.

²³The energy scale is the 1/4 power of the energy density. In slow-roll inflation it changes little during the time interval in which cosmologically-relevant scales exit the horizon.

5.2 Cosmological evolution of gravitational waves

The cosmological evolution of gravitational waves was an exercise on the Examples Sheets for the *Cosmology* course. The trace-free part of the ij Einstein equation gives

$$\ddot{h}_{ij} + 2\mathcal{H}\dot{h}_{ij} - \nabla^2 h_{ij} = -16\pi G a^2 \Pi_{ij}^T, \quad (5.18)$$

where Π_{ij}^T is the tensor part of the anisotropic stress. If we ignore the anisotropic stress, then in Fourier space

$$\ddot{h}^{(\pm 2)} + 2\mathcal{H}\dot{h}^{(\pm 2)} + k^2 h^{(\pm 2)} = 0. \quad (5.19)$$

Exact solutions can be found in terms of spherical Bessel functions during matter and radiation domination. However, here we shall content ourselves with some asymptotic features of the solution for general $a(\eta)$.

First, consider modes that are well outside the Hubble radius so $k \ll \mathcal{H}$. Equation (5.19) is then an over-damped oscillator. One solution is $h^{(\pm 2)} = \text{const.}$, and the other decays in time (like a^{-1} in radiation domination and $a^{-3/2}$ in matter domination). We see that the most rapidly growing mode outside the Hubble radius has constant amplitude. This result does not depend on any specific form for $a(\eta)$. This is important since it allows us to relate the super-Hubble amplitude of gravitational waves at late(ish) times to that generated in inflation without needing to know the detailed expansion history of the post-inflation universe. (The comoving curvature perturbation \mathcal{R} plays a similar role for adiabatic scalar perturbations.)

For modes well inside the Hubble radius ($k \gg \mathcal{H}$) we expect oscillatory solutions since Eq. (5.19) is a lightly-damped oscillator. We can eliminate the (expansion) damping with a field redefinition to find

$$\frac{\partial^2}{\partial \eta^2}(ah^{(\pm 2)}) + \left(k^2 - \frac{\ddot{a}}{a}\right)(ah^{(\pm 2)}) = 0, \quad (5.20)$$

and for $k \gg \mathcal{H}$ we can neglect the \ddot{a}/a term relative to k^2 . It follows that the solutions are

$$h^{(\pm 2)} \propto e^{\pm ik\eta}/a, \quad (5.21)$$

i.e., oscillations at (comoving) frequency k with adiabatic decay of the amplitude as $1/a$ due to expansion. We see that well inside the Hubble radius we recover the expected (flat-space) behaviour for gravitational waves but with an additional slow decay in amplitude. The decay in amplitude can be understood by thinking of the short-wavelength gravitational waves as a gas of (massless) gravitons. Like photons, the energy density of the graviton gas falls as $1/a^4$. Since the (effective) energy density of gravitational waves goes like the proper time derivative of h_{ij} squared, we require $a^{-2}\langle \dot{h}_{ij}\dot{h}^{ij} \rangle_k \sim a^{-4}$, where the angle brackets denote an average over many oscillations. This condition is consistent with the adiabatic decay in amplitude, $h_{ij} \propto a^{-1}$, in Eq. (5.21).

5.3 Gravity waves and the CMB

We now calculate the effect of gravitational waves on the temperature anisotropies of the CMB. We express the photon momentum in an orthonormal tetrad; for tensor perturbations, Eq. (2.3) reduces to

$$(E_0)^\mu = a^{-1} \delta_0^\mu \quad \text{and} \quad (E_i)^\mu = a^{-1} \left(\delta_i^\mu - \frac{1}{2} h_i^j \delta_j^\mu \right). \quad (5.22)$$

The photon 4-momentum expressed in terms of comoving energy and direction cosines e^i relative to the $(E_i)^\mu$ is

$$p^\mu = \frac{\epsilon}{a} [(E_0)^\mu + e^i (E_i)^\mu] \quad \Rightarrow \quad p^\mu = \frac{\epsilon}{a^2} \left[1, e^i - \frac{1}{2} h^i_j e^j \right]. \quad (5.23)$$

The evolution of the comoving energy, ϵ , follows from the geodesic equation (see below):

$$\frac{1}{\epsilon} \frac{d\epsilon}{d\eta} + \frac{1}{2} \dot{h}_{ij} e^i e^j = 0. \quad (5.24)$$

d ln $\epsilon/d\eta$ for tensor perturbations. The non-zero connection coefficients for the metric in Eq. (5.1) are (exercise!)

$$\begin{aligned} \Gamma_{00}^0 &= \mathcal{H} \\ \Gamma_{ij}^0 &= \mathcal{H} \delta_{ij} + \mathcal{H} h_{ij} + \frac{1}{2} \dot{h}_{ij} \\ \Gamma_{j0}^i &= \mathcal{H} \delta_j^i + \frac{1}{2} \dot{h}_j^i \\ \Gamma_{jk}^i &= \partial_{(j} h^i_{k)} - \frac{1}{2} \partial^i h_{jk}, \end{aligned} \quad (5.25)$$

correct to first order in perturbations. Here round brackets around indices denote symmetrisation. With $p^\mu = dx^\mu/d\lambda$, it follows that

$$\begin{aligned} \frac{d\eta}{d\lambda} &= \frac{\epsilon}{a^2} \\ \frac{dx^i}{d\eta} &= e^i - \frac{1}{2} h^i_j e^j, \end{aligned} \quad (5.26)$$

at linear order. The geodesic equation in conformal time is

$$\frac{\epsilon}{a^2} \frac{dp^\mu}{d\eta} + \Gamma_{\nu\rho}^\mu p^\nu p^\rho = 0, \quad (5.27)$$

and the 0-component is

$$\begin{aligned} &\frac{\epsilon}{a^2} \frac{d}{d\eta} \left(\frac{\epsilon}{a^2} \right) + \frac{\epsilon^2}{a^4} \left[\Gamma_{00}^0 + \Gamma_{ij}^0 \left(e^i - \frac{1}{2} h^i_k e^k \right) \left(e^j - \frac{1}{2} h^j_l e^l \right) \right] = 0 \\ \Rightarrow \quad &\frac{d \ln \epsilon}{d\eta} - 2\mathcal{H} + \mathcal{H} + \left(\mathcal{H} \delta_{ij} + \mathcal{H} h_{ij} + \frac{1}{2} \dot{h}_{ij} \right) \left(e^i - \frac{1}{2} h^i_k e^k \right) \left(e^j - \frac{1}{2} h^j_l e^l \right) = 0, \end{aligned} \quad (5.28)$$

where we have substituted for the perturbed connection coefficients. Expanding to first order in h_{ij} establishes Eq. (5.24).

The baryon bulk velocity vanishes for tensor perturbations so the Boltzmann equation (2.47) for the temperature anisotropies reduces to

$$\begin{aligned} \frac{\partial \Theta}{\partial \eta} + \mathbf{e} \cdot \nabla \Theta &= -a n_e \sigma_T \Theta + \frac{3 a n_e \sigma_T}{16\pi} \int d\hat{\mathbf{m}} \Theta(\hat{\mathbf{m}}) [1 + (\mathbf{e} \cdot \hat{\mathbf{m}})^2] \\ &\quad - \frac{1}{2} \dot{h}_{ij} e^i e^j. \end{aligned} \quad (5.29)$$

All perturbed scalar and vector quantities vanish for tensor perturbations so $\Theta(\mathbf{e})$ has only $l \geq 2$ moments. Neglecting the anisotropic nature of Thomson scattering, or, equivalently, the temperature quadrupole at last scattering, the solution of Eq. (5.29) is an integral along the unperturbed line of sight:

$$\Theta(\eta_0, \mathbf{x}_0, \mathbf{e}) \approx -\frac{1}{2} \int_0^{\eta_0} d\eta' e^{-\tau} \dot{h}_{ij} e^i e^j. \quad (5.30)$$

The approximation is good on large scales (where the temperature quadrupole has not had time to grow by last scattering) and, as we shall see, it is only on large scales where gravitational waves can contribute significantly to the temperature anisotropies. The physics behind Eq. (5.30) is as follows. The time derivative \dot{h}_{ij} is the shear induced by the gravitational waves. This quadrupole perturbation to the expansion rate produces an anisotropic redshifting of the CMB photons and an associated temperature anisotropy.

The shear source term for the tensor anisotropies, $\dot{h}_{ij} e^i e^j$, is locally a quadrupole since \dot{h}_{ij} is trace-free. This means that shear locally produces quadrupole anisotropies. However, anisotropies at smaller angular scales are generated by free-streaming due to the spatial variation of the source. To see this in detail, consider a Fourier mode of h_{ij} with \mathbf{k} along the z -axis. We then have the shear source

$$\begin{aligned} \dot{h}_{ij}^{(\pm 2)}(\eta, k\hat{\mathbf{z}}) e^i e^j &= \frac{1}{2\sqrt{2}} \dot{h}^{(\pm 2)}(\eta, k\hat{\mathbf{z}}) [(\hat{\mathbf{x}} \pm i\hat{\mathbf{y}}) \cdot \mathbf{e}]^2 \\ &= \frac{1}{2\sqrt{2}} \dot{h}^{(\pm 2)}(\eta, k\hat{\mathbf{z}}) \sin^2 \theta e^{\pm 2i\phi} \\ &= \sqrt{\frac{4\pi}{15}} \dot{h}^{(\pm 2)}(\eta, k\hat{\mathbf{z}}) Y_{2\pm 2}(\mathbf{e}), \end{aligned} \quad (5.31)$$

where θ and ϕ are the polar coordinates associated with the direction \mathbf{e} . The contribution of this Fourier mode at time $\eta_0 - \chi$ to the integral in Eq. (5.30) for the observed

temperature anisotropies is

$$\begin{aligned} \sqrt{\frac{4\pi}{15}} \dot{h}^{(\pm 2)}(\eta, k\hat{\mathbf{z}}) Y_{2\pm 2}(\mathbf{e}) e^{-ik\chi \cos \theta} &= \frac{4\pi}{\sqrt{15}} \dot{h}^{(\pm 2)}(\eta, k\hat{\mathbf{z}}) \\ &\times \sum_{L \geq 0} (-i)^L \sqrt{2L+1} j_L(k\chi) Y_{2\pm 2}(\mathbf{e}) Y_{L0}(\mathbf{e}), \end{aligned} \quad (5.32)$$

where we used the Rayleigh plane-wave expansion and expressed $P_l(\cos \theta)$ in terms of $Y_{L0}(\mathbf{e})$. The azimuthal dependence of this contribution to the anisotropies has $m = \pm 2$. We want the spherical multipoles of Eq. (5.32), which we can extract by expanding the product of spherical harmonics on the right with Eq. (1.51):

$$\begin{aligned} \sqrt{\frac{4\pi}{15}} \dot{h}^{(\pm 2)}(\eta, k\hat{\mathbf{z}}) Y_{2\pm 2}(\mathbf{e}) e^{-ik\chi \cos \theta} &= \sqrt{\frac{4\pi}{3}} \dot{h}^{(\pm 2)}(\eta, k\hat{\mathbf{z}}) \sum_{L \geq 0} \left[(-i)^L (2L+1) j_L(k\chi) \right. \\ &\times \left. \sum_{l \geq 2} \sqrt{2l+1} \begin{pmatrix} 2 & L & l \\ \mp 2 & 0 & \pm 2 \end{pmatrix} \begin{pmatrix} 2 & L & l \\ 0 & 0 & 0 \end{pmatrix} Y_{l\pm 2}(\mathbf{e}) \right]. \end{aligned} \quad (5.33)$$

Since the anisotropies have $m = \pm 2$, their expansion in spherical harmonics necessarily has $l \geq 2$ (as advertised earlier). The second $3j$ symbols forces $L + l$ to be even, so the l th multipoles of the projection involve the $L = l$ and $L = l \pm 2$ multipoles of the expansion of the plane wave $\exp(-ik\chi \cos \theta)$.

We can simplify further by writing out the $3j$ symbols explicitly and using the recursion relations for Bessel functions to express $j_{l\pm 2}$ in terms of j_l . The final result is remarkably compact:

$$\dot{h}_{ij}^{(\pm 2)}(\eta, k\hat{\mathbf{z}}) e^i e^j e^{-ik\chi \cos \theta} = -\sqrt{\frac{\pi}{2}} \dot{h}^{(\pm 2)}(\eta, k\hat{\mathbf{z}}) \sum_l (-i)^l \sqrt{2l+1} \sqrt{\frac{(l+2)!}{(l-2)!}} \frac{j_l(k\chi)}{(k\chi)^2} Y_{l\pm 2}(\mathbf{e}). \quad (5.34)$$

We see that for gravitational waves, the spatial-to-angular projection for the shear source is controlled by $j_l(k\chi)/(k\chi)^2$. As for scalar perturbations, this is concentrated on multipoles $l \sim k\chi$. However, the shear source is not generally localised at the last-scattering surface, and we have to integrate contributions all along the lightcone.

Exercise: derive Eq. (5.34). You will need to use the following explicit expressions for products of the $3j$ symbols:

$$\begin{aligned} \left(\begin{array}{ccc} 2 & l+2 & l \\ 2 & 0 & -2 \end{array} \right) \left(\begin{array}{ccc} 2 & l+2 & l \\ 0 & 0 & 0 \end{array} \right) &= \sqrt{\frac{3}{8}} \sqrt{\frac{(l+2)!}{(l-2)!}} \frac{1}{(2l+5)(2l+3)(2l+1)} \\ \left(\begin{array}{ccc} 2 & l-2 & l \\ 2 & 0 & -2 \end{array} \right) \left(\begin{array}{ccc} 2 & l-2 & l \\ 0 & 0 & 0 \end{array} \right) &= \sqrt{\frac{3}{8}} \sqrt{\frac{(l+2)!}{(l-2)!}} \frac{1}{(2l+1)(2l-1)(2l-3)} \\ \left(\begin{array}{ccc} 2 & l & l \\ 2 & 0 & -2 \end{array} \right) \left(\begin{array}{ccc} 2 & l & l \\ 0 & 0 & 0 \end{array} \right) &= -\sqrt{\frac{3}{2}} \sqrt{\frac{(l+2)!}{(l-2)!}} \frac{1}{(2l+3)(2l+1)(2l-1)}, \end{aligned} \quad (5.35)$$

and the recursion relation $(2l+1)j_l(x) = xj_{l+1}(x) + xj_{l-1}(x)$.

So far we have only considered Fourier modes with \mathbf{k} along the z -direction. We can obtain the contribution to the temperature anisotropy from a general wavenumber by actively rotating the result in Eq. (5.34); this gives

$$\begin{aligned} \dot{h}_{ij}^{(\pm 2)}(\eta, \mathbf{k}) e^i e^j e^{-ik\chi \hat{\mathbf{k}} \cdot \mathbf{e}} &= -\sqrt{\frac{\pi}{2}} \dot{h}^{(\pm 2)}(\eta, \mathbf{k}) \\ &\times \sum_{lm} (-i)^l \sqrt{2l+1} \sqrt{\frac{(l+2)!}{(l-2)!}} \frac{j_l(k\chi)}{(k\chi)^2} D_{m\pm 2}^l(\phi_{\mathbf{k}}, \theta_{\mathbf{k}}, 0) Y_{lm}(\mathbf{e}). \end{aligned} \quad (5.36)$$

Finally, we can write down the spherical multipoles of Θ_{lm} in the form

$$\Theta_{lm}(\eta_0, \mathbf{x}_0) = \frac{1}{\sqrt{2}} \sum_{\pm 2} \int \frac{d^3 \mathbf{k}}{(2\pi)^{3/2}} (-i)^l \Theta_l^{(\pm 2)}(\eta_0, \mathbf{k}) \sqrt{\frac{4\pi}{2l+1}} D_{m\pm 2}^l(\phi_{\mathbf{k}}, \theta_{\mathbf{k}}, 0) e^{i\mathbf{k} \cdot \mathbf{x}_0}, \quad (5.37)$$

where the normal-mode coefficient

$$\Theta_l^{(\pm 2)}(\eta_0, \mathbf{k}) = \frac{2l+1}{4} \int_0^{\eta_0} d\eta' e^{-\tau} \dot{h}^{(\pm 2)}(\eta', \mathbf{k}) \sqrt{\frac{(l+2)!}{(l-2)!}} \frac{j_l(k\chi)}{(k\chi)^2}. \quad (5.38)$$

Equation (5.37) is the generalisation of the result in Eq. (4.5) for scalar perturbations. The latter generate only $m = 0$ anisotropies for \mathbf{k} along the z -direction. Recalling the result, $D_{m0}^l(\phi_{\mathbf{k}}, \theta_{\mathbf{k}}, 0) = \sqrt{4\pi/(2l+1)} Y_{lm}^*(\hat{\mathbf{k}})$, we see that replacing ± 2 by 0 in Eq. (5.37) recovers Eq. (4.5) (up to factors of $1/\sqrt{2}$).

The calculation of the 2-point function follows that for scalar perturbations. For gravitational waves, the $\Theta^{(\pm 2)}(\eta, \mathbf{k})$ are linear in the primordial $h^{(\pm 2)}(\mathbf{k})$, with a coefficient

$\Theta^{(\pm 2)}(\eta, \mathbf{k})/h^{(\pm 2)}(\mathbf{k})$ that is independent of the polarization state and direction of \mathbf{k} . We therefore have

$$\begin{aligned}
\langle \Theta_{lm} \Theta_{l'm'}^* \rangle &= (-i)^{l-l'} \frac{4\pi}{\sqrt{(2l+1)(2l'+1)}} \\
&\times \frac{1}{2} \sum_{pp'} \int \frac{d^3\mathbf{k}}{(2\pi)^{3/2}} \frac{d^3\mathbf{k}'}{(2\pi)^{3/2}} \left\{ \left[\frac{\Theta_l^{(p)}(\eta_0, \mathbf{k})}{h^{(p)}(\mathbf{k})} \right] \left[\frac{\Theta_l^{(p')}\!(\eta_0, \mathbf{k}')}{h^{(p')}\!(\mathbf{k}')} \right] \right. \\
&\times \underbrace{\langle h^{(p)}(\mathbf{k}) h^{(p')*}(\mathbf{k}') \rangle}_{\frac{2\pi^2}{k^3} \mathcal{P}_h(k) \delta^{(3)}(\mathbf{k}-\mathbf{k}') \delta^{pp'}} D_{mp}^l(\phi_{\mathbf{k}}, \theta_{\mathbf{k}}, 0) D_{m'p'}^{l'}(\phi_{\mathbf{k}'}, \theta_{\mathbf{k}'}, 0) e^{i(\mathbf{k}-\mathbf{k}') \cdot \mathbf{x}_0} \Big\} \\
&= (-i)^{l-l'} \frac{1}{\sqrt{(2l+1)(2l'+1)}} \int d \ln k \left[\frac{\Theta_l^{(p)}(\eta_0, \mathbf{k})}{h^{(p)}(\mathbf{k})} \right]^2 \mathcal{P}_h(k) \\
&\times \int d\hat{\mathbf{k}} D_{mp}^l(\phi_{\mathbf{k}}, \theta_{\mathbf{k}}, 0) D_{m'p'}^{l'}(\phi_{\mathbf{k}}, \theta_{\mathbf{k}}, 0) \\
&= \delta_{ll'} \delta_{mm'} \frac{4\pi}{(2l+1)^2} \int d \ln k \left[\frac{\Theta_l^{(p)}(\eta_0, \mathbf{k})}{h^{(p)}(\mathbf{k})} \right]^2 \mathcal{P}_h(k), \tag{5.39}
\end{aligned}$$

where we have used

$$\int_0^{2\pi} d\alpha \int_0^\pi \sin \beta d\beta D_{mn}^l(\alpha, \beta, 0) D_{m'n}^l(\alpha, \beta, 0) = \frac{4\pi}{2l+1} \delta_{ll'} \delta_{mm'}, \tag{5.40}$$

which follows from orthogonality of the $D_{mm'}^l$ over the group manifold²⁴. Finally, we see that the angular power spectrum

$$C_l = \frac{4\pi}{(2l+1)^2} \int d \ln k \left[\frac{\Theta_l^{(p)}(\eta_0, \mathbf{k})}{h^{(p)}(\mathbf{k})} \right]^2 \mathcal{P}_h(k). \tag{5.41}$$

The angular power spectrum of the temperature anisotropies from gravitational waves is shown in Fig. 13 for the tensor-to-scalar ratio $r = 1$. This is obtained from numerical evaluation of Eqs (5.38) and (5.41). The main features to note from this plot are as follows.

- The angular power spectrum is approximately scale-invariant [$[l(l+1)C_l \approx \text{const.}]$ on large scales for a near scale-invariant primordial power spectrum.

²⁴The following orthogonality relation,

$$\int_0^{2\pi} d\alpha \int_0^\pi \sin \beta d\beta \int_0^{2\pi} d\gamma D_{mn}^l(\alpha, \beta, \gamma) D_{m'n'}^l(\alpha, \beta, \gamma) = \frac{8\pi^2}{2l+1} \delta_{ll'} \delta_{mm'} \delta_{nn'},$$

is the continuous version of the result from representation theory concerning the orthogonality over the group of the matrix elements of inequivalent irreducible representations.

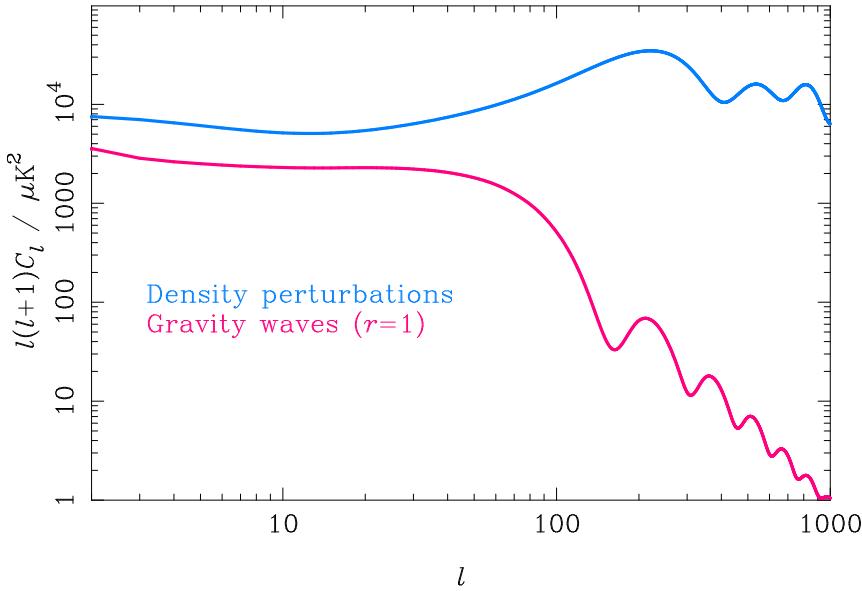


Figure 13: Angular power spectra of the temperature anisotropies for scalar (density) perturbations (blue) and gravitational waves (magenta) with tensor-to-scalar ratio $r = 1$.

- The power spectrum falls rapidly beyond $l \approx 60$. This corresponds to the angular scale subtended by the horizon at last scattering. For larger multipoles, there are oscillations in the spectrum. These are analogous to the acoustic oscillations in the angular power spectrum from scalar perturbations. In the case of gravitational waves, the oscillations occur at multipoles $l = k\chi_*$ where the wavenumbers k correspond to modes that reach extrema of their sub-horizon oscillations at last scattering (see Sec. 5.2).

5.4 Current constraints on gravitational waves from CMB temperature anisotropies

Joint constraints on the tensor-to-scalar ratio r and the spectral index of the power spectrum of primordial curvature perturbations,

$$n_s \equiv 1 + d \ln \mathcal{P}_R(k) / d \ln k, \quad (5.42)$$

are shown in Fig. 14. These are from the temperature anisotropies measured by Planck. You should take away the following points from this plot.

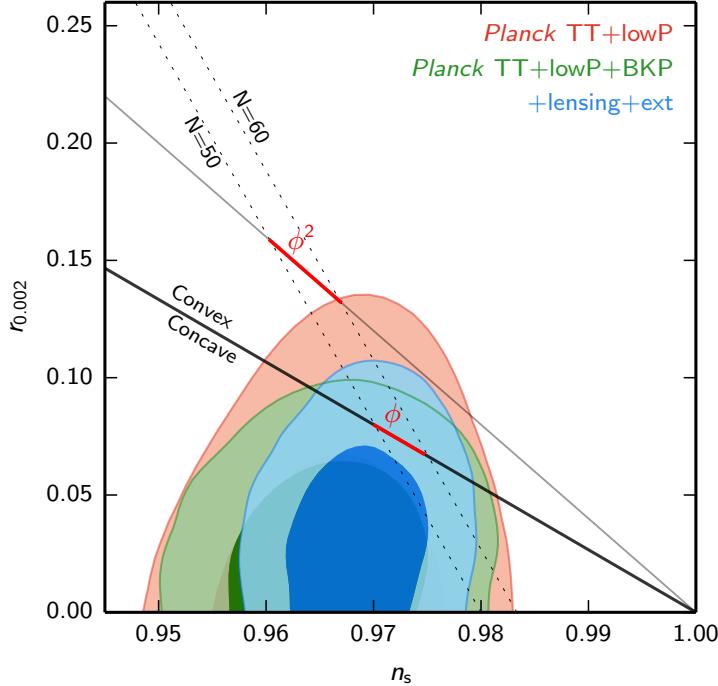


Figure 14: Joint constraints (68% and 95%) on the inflation observables n_s and r in models with power-law spectra. Also shown are the predictions for the simple $m^2\phi^2$ model in which the inflaton is a massive, non-interacting scalar field, and models with a linear potential. Models with monomial potentials lie on the dashed lines for $N = 50$ or $N = 60$ e -folds of inflation from the time that cosmological scales exited the horizon to the end of inflation. Red contours use the temperature anisotropies measured by Planck. Green contours further include measurements of B -mode polarization (see Sec. 6.6) from BICEP2 and Keck Array, with Galactic dust cleaned using Planck polarization data at 353 GHz (where dust emission dominates the CMB). Blue contours include other astrophysical data and CMB lensing data from Planck.

- There is now strong observational evidence for a lack of scale-invariance of the power spectrum of the primordial curvature perturbation, i.e., $n_s < 1$. This is evidence for a small, but non-zero, departure from de Sitter expansion during inflation.
- There is no detection of gravitational waves (i.e., r is consistent with zero). The marginalised constraint in models with power-law primordial spectra is $r < 0.11$ (95% confidence).

The marginalised constraint on r from the Planck temperature measurements are about as good as one could ever do with measurements of the temperature anisotropies. To see this, suppose that r is the only unknown parameter and that we have noise-free measurements of the temperature anisotropies on large scales where gravitational waves

can make a significant contribution to the anisotropies. We can form a set of power spectrum estimates \hat{C}_l , and from each of these we can estimate r as the excess power over the power C_l from curvature perturbations, i.e.,

$$\hat{r}_l = \frac{\hat{C}_l - C_l}{C_l^{\text{gw}}(r=1)}, \quad (5.43)$$

where $C_l^{\text{gw}}(r=1)$ is the angular power spectrum from gravitational waves for $r=1$. Due to cosmic variance, there is independent uncertainty in each of these estimates. Combining the \hat{r}_l with inverse-variance weighting (the optimal thing to do), we forecast an error on r of $\sigma(r)$ where, in the null hypothesis ($r=0$),

$$\frac{1}{\sigma^2(r)} = \sum_l \frac{2l+1}{2} \left(\frac{C_l^{\text{gw}}(r=1)}{C_l} \right)^2. \quad (5.44)$$

We can make a rough estimate of $\sigma(r)$ by approximating $C_l^{\text{gw}}(r=1)/C_l \approx 0.4$ as constant for $l < 60$, and zero otherwise (see Fig. 13). It follows that

$$\frac{1}{\sigma^2(r)} \sim \frac{1}{2} \times 60^2 \times (0.4)^2 \quad \Rightarrow \quad \sigma(r) \sim 0.06. \quad (5.45)$$

A more careful calculation using the actual spectra in Fig. 13 gives $\sigma(r) = 0.08$, close to the rough value obtained here. The 95% upper limit from Planck is crudely a 2σ limit (it would be this if the marginalised probability density for r were a Gaussian centred on $r=0$) and is comparable (actually rather better!) to what our ideal forecast would suggest.

Note that what is limiting our ability to constrain r further from the temperature anisotropies is the cosmic variance of the dominant signal from scalar perturbations. To improve the limit on gravitational waves further with the CMB, we need an observable that circumvents this variance. Fortunately there is one such CMB observable – the B -mode polarization, which we now discuss.

6 CMB polarization

So far we have not discussed the *polarization* of the CMB. Generally, scattering of anisotropic unpolarized radiation generates polarization (e.g., the scattered light from the overhead sky is linearly polarized). For the CMB, Thomson scattering of the small anisotropies around recombination induces linear polarization. Polarization is an important observable of the CMB, providing complementary information to the temperature anisotropies. In particular, polarization is potentially a much more sensitive probe of gravitational waves than the temperature anisotropies.

6.1 Stokes parameters

We begin by introducing the *Stokes parameters* as a convenient way of parameterising (partially-)polarized radiation. Classically, for a quasi-monochromatic plane wave propagating along the z -direction, the Hermitian correlation tensor of the (complex) components of the electric field can be written in terms of four real Stokes parameters as

$$\begin{pmatrix} \langle E_x E_x^* \rangle & \langle E_x E_y^* \rangle \\ \langle E_y E_x^* \rangle & \langle E_y E_y^* \rangle \end{pmatrix} \equiv \frac{1}{2} \begin{pmatrix} I + Q & U + iV \\ U - iV & I - Q \end{pmatrix}. \quad (6.1)$$

Here, the angle brackets denote time averages over intervals that are long compared to the wave period, but short compared to any variation in the intensity. The trace is the total intensity I and is the sum of the intensities transmitted by two orthogonal linear polarizing grids. Linear polarization is described by Q and U , with Q giving the difference in intensity received through two polarizing grids aligned with the x and y -directions respectively. Similarly, U is the difference in intensity after rotating the grids by 45° . The Stokes parameters Q and U clearly depend on the choice of x and y -axes (see below). Finally, V describes circular polarization. As we shall discuss in Sec. 6.4.1, Thomson scattering does not excite circular polarization so we shall not consider it further here.

Exercise: show that $I^2 \geq Q^2 + U^2 + V^2$. The equality only holds for *fully-polarized* radiation, for which the complex field amplitudes E_x and E_y are independent of time. Otherwise, the radiation is said to be *partially polarized*.

With the CMB, we are dealing with diffuse radiation fields. We can easily generalise the definition of the Stokes parameters by considering specific intensities, i.e., the power per unit area (orthogonal to the propagation direction of interest) per energy/frequency interval per solid angle. (An equivalent viewpoint, generalising kinetic theory to polarized radiation, is given below.) In linear theory, the frequency dependence of the Stokes parameters Q and U (which vanish at zero order in perturbations by symmetry) is inherited from the anisotropies in the total intensity I . We express these as *equivalent temperatures* by dividing out the derivative of the Planck function, as in Eq. (2.33). We shall denote these equivalent temperatures by $Q(\mathbf{e})$ and $U(\mathbf{e})$.

Stokes parameters in relativistic kinetic theory. To understand how to generalise relativistic kinetic theory to account for polarized radiation, we must first consider the geometric-optics limit of classical electrodynamics. In this limit, the electromagnetic 4-vector potential A^μ in

spacetime is decomposed as

$$A^\mu = \mathcal{A}\epsilon^\mu e^{iS}, \quad (6.2)$$

where the phase S varies rapidly compared to the amplitude \mathcal{A} and the (complex) *polarization 4-vector* ϵ^μ . The latter is spacelike and satisfies

$$\epsilon^\mu \epsilon_\mu^* = 1. \quad (6.3)$$

In the Lorenz gauge, ϵ^μ is orthogonal to the wavenumber $k^\mu = \nabla^\mu S$. Applying the WKB approximation to Maxwell's equations shows that $k^\mu \nabla_\mu k^\nu = 0$ and that ϵ^μ is parallel-transported along the integral curves of k^μ ("light rays").

In a quantum picture, we talk of photons with 4-momentum $p^\mu \propto k^\mu$ and polarization ϵ^μ . A gas of photons is described by a Hermitian-tensor-valued one-particle distribution function $f_{\mu\nu}(x^\rho, p^\rho)$ that is a function of the spacetime position x^ρ and four-momentum p^ρ of the photon. It is defined such that the expected number of photons contained in a proper phase-space element $d^3\mathbf{x}d^3\mathbf{p}$ with measured polarization state ϵ^μ is $\epsilon^{\mu*} f_{\mu\nu} \epsilon^\nu d^3\mathbf{x}d^3\mathbf{p}$.

An observer with 4-velocity u^μ sees a photon as having energy $E = -p^\mu u_\mu$ and direction e^μ where

$$p^\mu = E(u^\mu + e^\mu), \quad (6.4)$$

with $e^\mu u_\mu = 0$. For a photon in a pure polarization state ϵ^μ , the direction of polarization measured by the observer is $\mathcal{H}^\mu{}_\nu \epsilon^\nu$. Here,

$$\mathcal{H}_{\mu\nu} \equiv g_{\mu\nu} + u_\mu u_\nu - e_\mu e_\nu \quad (6.5)$$

is the observer's *screen projection tensor* that projects into the space perpendicular to u^μ and e^μ . The (Lorenz-gauge) polarization 4-vector is only unique up to constant multiples of p^μ , reflecting the remaining electromagnetic gauge freedom, but the *observed* polarization vector $\mathcal{H}^\mu{}_\nu \epsilon^\nu$ is unique. As the residual gauge freedom also affects the distribution function $f_{\mu\nu}$, it is sometimes convenient to work directly with the polarization tensor observed with respect to u^μ . This is obtained by screen projecting on both indices and is (electromagnetic) gauge-invariant; it does, however, depend on the choice of observer 4-velocity.

The projected polarization tensor has four real degrees of freedom. It can be decomposed on the spatial elements of an orthonormal tetrad (with $u^\mu = (E_0)^\mu$) as

$$(E_i)^\mu (E_j)^\nu \mathcal{H}_\mu{}^\rho \mathcal{H}_\nu{}^\tau f_{\rho\tau} = \frac{1}{2} f (\delta_{ij} - e_i e_j) + \mathcal{P}_{ij} + \frac{1}{2} i f_V \epsilon_{ijk} e^k. \quad (6.6)$$

The quantity f is the polarization-averaged distribution function that we have been (implicitly) using up to now, while f_V describes circular polarization. The 2D tensor \mathcal{P}_{ij} is orthogonal to e_i and is symmetric and trace-free. It describes linear polarization. On an orthonormal diad of vectors, forming a right-handed orthonormal basis with \mathbf{e} , the diagonal elements of \mathcal{P}_{ij} can be written as $f_Q/2$ and $-f_Q/2$, while the off-diagonal elements are $f_U/2$ and $-f_U/2$. The quantities f , f_Q , f_U , and f_V are related to the intensity Stokes parameters in Eq. (6.1) by, for example, $f \propto I/E^3$. The factor of E^3 arises from relating a phase-space density to specific intensity.

We have already seen that the phase-space volume element $d^3\mathbf{x}d^3\mathbf{p}$ is Lorentz-invariant and is conserved along the photon path through phase space. It follows that the distribution function $f_{\mu\nu}$ is also Lorentz-invariant, and is parallel transported (like the polarization vectors ϵ^ν) in the absence of scattering. Furthermore, the decomposition in Eq. (6.6) can be shown to be irreducible with respect to Lorentz transformations. This means that f and f_V are Lorentz-invariant and conserved, while f_Q and f_U are so on a basis that is obtained by screen projection of a parallel-propagated polarization basis.

In linear perturbation theory, we can write f_Q and f_U in terms of equivalent temperatures. For example,

$$f_Q(\eta, \mathbf{x}, \epsilon, \mathbf{e}) = -Q(\eta, \mathbf{x}, \mathbf{e}) \frac{d\bar{f}}{d\ln\epsilon}. \quad (6.7)$$

The linear Stokes parameters depend on the choice of basis. For a propagation direction \mathbf{e} , we define the Stokes parameters on a local x - y basis defined by $\hat{\theta}, \hat{\phi}$ of the spherical-polar basis at \mathbf{e} . Note that $\hat{\theta}, \hat{\phi}$ and \mathbf{e} form a right-handed basis. It will be convenient to use the complex null basis

$$\mathbf{m}_\pm \equiv \hat{\theta} \pm i\hat{\phi}. \quad (6.8)$$

Under a right-handed rotation about the propagation direction through an angle ψ ,

$$\begin{aligned} \mathbf{m}_\pm &\rightarrow e^{\mp i\psi} \mathbf{m}_\pm \\ Q \pm iU &\rightarrow e^{\mp 2i\psi} (Q \pm iU). \end{aligned} \quad (6.9)$$

Any object on the 2-sphere, ${}_s\eta(\mathbf{e})$ that transforms as

$${}_s\eta \rightarrow e^{is\psi} {}_s\eta \quad \text{for} \quad \mathbf{m}_\pm \rightarrow e^{\pm i\psi} \mathbf{m}_\pm \quad (6.10)$$

is said to be *spin s*. A simple example of a spin 1 quantity is $\mathbf{V} \cdot \mathbf{m}_+$, where \mathbf{V} is a vector. The complex linear polarization $Q \pm iU$ is spin ± 2 .

In plots, linear polarization is often represented by a headless vector at an angle α to the polarization basis vector that defines positive Q (with the angle defined in a right-handed sense with respect to the propagation direction). The angle α is defined by

$$\frac{Q}{\sqrt{Q^2 + U^2}} = \cos 2\alpha \quad \text{and} \quad \frac{U}{\sqrt{Q^2 + U^2}} = \sin 2\alpha. \quad (6.11)$$

For fully-polarized radiation, the headless vector is along the (unsigned) direction of the electric field. The length of the vector usually represents $\sqrt{Q^2 + U^2}$.

6.2 *E* and *B*-mode polarization and spin-weighted spherical harmonics

The Stokes parameters Q and U are a very convenient local description of the polarization of the CMB, but they have the awkward property of depending on the choice of polarization basis. Fundamentally, this is because Q and U are the components of the linear-polarization tensor \mathcal{P}_{ij} . A more convenient alternative is to express the linear polarization in terms of (spherical) derivatives of scalar potentials E and B .

As a warm-up, we first consider a 2D vector field on the sphere V^a .

Tensor calculus on the 2-sphere. In this section, we shall use Roman letters for (abstract) tensor indices on the 2-sphere, and the metric on the 2-sphere will be denoted g_{ab} and the alternating tensor ϵ_{ab} . The covariant derivative is denoted by ∇_a . The metric can be written in terms of our complex null vectors (which we now denote as m_\pm^a) as

$$g_{ab} = \frac{1}{2}(m_{+a}m_{-b} + m_{-a}m_{+b}) . \quad (6.12)$$

The alternating tensor can be written as

$$\epsilon_{ab} = \frac{i}{2}(m_{+a}m_{-b} - m_{-a}m_{+b}) , \quad (6.13)$$

and the action on the null vectors is

$$\epsilon_a{}^b m_{\pm b} = \pm im_{\pm a} , \quad (6.14)$$

i.e., rotation through $\pi/2$.

If we use as a coordinate chart the usual spherical polar coordinates, the null vectors can be expressed in terms of the coordinate basis vectors $(\partial_\theta)^a$ and $(\partial_\phi)^a$ as

$$m_\pm^a = (\partial_\theta)^a \pm i \operatorname{cosec} \theta (\partial_\phi)^a . \quad (6.15)$$

The non-zero connection coefficients are

$$\Gamma_{\phi\phi}^\theta = -\sin \theta \cos \theta \quad \text{and} \quad \Gamma_{\theta\phi}^\phi = \cot \theta , \quad (6.16)$$

for this chart.

The directional derivatives of the null basis vectors are

$$m_+^a \nabla_a m_\pm^b = \pm \cot \theta m_\pm^b \quad \text{and} \quad m_-^a \nabla_a m_\pm^b = \mp \cot \theta m_\pm^b . \quad (6.17)$$

We can always write V_a as the sum of a gradient of a scalar $\nabla_a V_E$ and a divergence-free vector. In 2D, the latter can be written as $\epsilon^b{}_a \nabla_b V_B$, so that

$$V_a = \nabla_a V_E + \epsilon^b{}_a \nabla_b V_B. \quad (6.18)$$

We can form spin ± 1 quantities by taking the inner product with m_\pm^a . Using $\epsilon^b{}_a m_\pm^a = \pm i m_\pm^b$ [Eq. (6.14)], we have

$$\begin{aligned} m_\pm^a V_a &= m_\pm^a \nabla_a (V_E \pm i V_B) \\ &= (\partial_\theta \pm i \operatorname{cosec} \theta \partial_\phi)(V_E \pm i V_B). \end{aligned} \quad (6.19)$$

It is convenient to introduce *spin-raising* and *spin-lowering* operators, \mathfrak{D} and $\bar{\mathfrak{D}}$, whose action on a spin s quantity ${}_s\eta$ is

$$\begin{aligned} \mathfrak{D}_s \eta &= -\sin^s \theta (\partial_\theta + i \operatorname{cosec} \theta \partial_\phi) (\sin^{-s} \theta {}_s\eta) \\ \bar{\mathfrak{D}}_s \eta &= -\sin^{-s} \theta (\partial_\theta - i \operatorname{cosec} \theta \partial_\phi) (\sin^s \theta {}_s\eta). \end{aligned} \quad (6.20)$$

(The minus signs are conventional.) We see that the spin ± 1 null components of V^a can be written in terms of these operators acting on the spin 0 (scalar) $V_E \pm i V_B$:

$$\begin{aligned} m_+^a V_a &= -\mathfrak{D}(V_E + i V_B) \\ m_-^a V_a &= -\bar{\mathfrak{D}}(V_E - i V_B). \end{aligned} \quad (6.21)$$

We now return to linear polarization. As we have seen, this can be described by a symmetric trace-free tensor on the 2-sphere, \mathcal{P}_{ab} . This has two real degrees of freedom and can be expanded in terms of two real scalar potentials P_E and P_B :

$$\mathcal{P}_{ab} = \nabla_{(a} \nabla_{b)} P_E + \epsilon^c{}_{(a} \nabla_{b)} \nabla_c P_B. \quad (6.22)$$

This is like the decomposition of a 3D symmetric trace-free tensor into scalar, vector and tensor parts (as in cosmological perturbation theory). However, in 2D there is no tensor part and the divergence-free vector whose symmetrised derivative gives the vector part can be written as $\epsilon^c{}_b \nabla_c P_B$. The complex combinations of Stokes parameters are the null components of \mathcal{P}_{ab} :

$$Q \pm iU = m_\pm^a m_\pm^b \mathcal{P}_{ab}, \quad (6.23)$$

and the inverse relation is

$$\mathcal{P}_{ab} = \frac{1}{4} [(Q + iU)m_{-a}m_{-b} + (Q - iU)m_{+a}m_{+b}]. \quad (6.24)$$

Note that in 2D, a symmetric rank- n tensor that is trace-free on all indices has only two degrees of freedom, and a basis is provided by $m_+^{a_1} \dots m_+^{a_n}$ and $m_-^{a_1} \dots m_-^{a_n}$. Using

Eq. (6.22), we have

$$\begin{aligned}
Q \pm iU &= m_{\pm}^a m_{\pm}^b (\nabla_a \nabla_b P_E + \epsilon^c_a \nabla_b \nabla_c P_B) \\
&= m_{\pm}^a m_{\pm}^b \nabla_a \nabla_b (P_E \pm iP_B) \\
&= (m_{\pm}^a \nabla_a)^2 (P_E \pm iP_B) - \underbrace{(m_{\pm}^a \nabla_a m_{\pm}^b)}_{\cot \theta m_{\pm}^b} \nabla_b (P_E \pm iP_B) \\
&= (\partial_{\theta} \pm i \operatorname{cosec} \theta \partial_{\phi})^2 (P_E \pm iP_B) - \cot \theta (\partial_{\theta} \pm i \operatorname{cosec} \theta \partial_{\phi}) (P_E \pm iP_B) \\
&= \sin \theta (\partial_{\theta} \pm i \operatorname{cosec} \theta \partial_{\phi}) [(\sin \theta)^{-1} (\partial_{\theta} \pm i \operatorname{cosec} \theta \partial_{\phi}) (P_E \pm iP_B)] . \quad (6.25)
\end{aligned}$$

Note how the connection term from $m_{\pm}^a \nabla_a m_{\pm}^b$ gives rise to the $(\sin \theta)^{-1}$ in the final bracket. Recalling the definitions of the spin-raising and lowering operators, we can write

$$Q + iU = \bar{\mathfrak{D}}\bar{\mathfrak{D}}(P_E + iP_B) \quad \text{and} \quad Q - iU = \bar{\mathfrak{D}}\bar{\mathfrak{D}}(P_E - iP_B) . \quad (6.26)$$

In these expressions, the action of the first $\bar{\mathfrak{D}}$ raises the spin of $P_E + iP_B$ to +1, and the second $\bar{\mathfrak{D}}$ raises it further to +2.

To see quite generally that $\bar{\mathfrak{D}}$ and $\bar{\mathfrak{D}}$ are spin-raising and lowering operators, respectively, consider a spin s quantity (for simplicity, consider $s > 0$). We can form a symmetric trace-free tensor $\eta_{a_1 \dots a_s} \equiv {}_s \eta m_{-a_1} \dots m_{-a_s}$ from this. Taking the covariant derivative along m_{+}^a returns another symmetric trace-free rank- s tensor, and the contraction of this with $m_{+}^{a_1} \dots m_{+}^{a_s}$ gives a quantity with spin $s + 1$. We can express this spin $s + 1$ quantity in terms of $\bar{\mathfrak{D}}_s \eta$ since

$$\begin{aligned}
(m_{+}^b \nabla_b \eta_{a_1 \dots a_s}) m_{+}^{a_1} \dots m_{+}^{a_s} &= 2^s m_{+}^b \nabla_{bs} \eta + s 2^{s-1} {}_s \eta \underbrace{(m_{+}^b \nabla_b m_{-a})}_{-\cot \theta m_{-a}} m_{+}^a \\
&= 2^s [(\partial_{\theta} \pm i \operatorname{cosec} \theta \partial_{\phi}) {}_s \eta - s \cot \theta {}_s \eta] \\
&= 2^s \sin^s \theta (\partial_{\theta} \pm i \operatorname{cosec} \theta \partial_{\phi}) (\sin^{-s} \theta {}_s \eta) \\
&= -2^s \bar{\mathfrak{D}}_s \eta , \quad (6.27)
\end{aligned}$$

so that $\bar{\mathfrak{D}}_s \eta$ does indeed have spin $s + 1$.

The real spin 0 fields P_E and P_B may be expanded in spherical harmonics:

$$\begin{aligned}
P_E(\mathbf{e}) &= \sum_{lm} \sqrt{\frac{(l-2)!}{(l+2)!}} E_{lm} Y_{lm}(\mathbf{e}) \\
P_B(\mathbf{e}) &= \sum_{lm} \sqrt{\frac{(l-2)!}{(l+2)!}} B_{lm} Y_{lm}(\mathbf{e}) . \quad (6.28)
\end{aligned}$$

The pre-factors here are convenient, as they effectively undo two factors of l that arise from taking two derivatives of P_E and P_B in forming $Q \pm iU$. This ensures that the variance of the E_{lm} and B_{lm} more directly reflect the scale dependence of Q and U .

Note that the summations in Eq. (6.28) begin at $l = 2$: modes with $l = 0$ or $l = 1$ modes would be annihilated by the spin-raising and lowering operators and cannot contribute to $Q \pm iU$. We can now write

$$(Q \pm iU)(\mathbf{e}) = \sum_{lm} (E_{lm} \pm iB_{lm})_{\pm 2} Y_{lm}(\mathbf{e}). \quad (6.29)$$

Here, the $\pm 2 Y_{lm}(\mathbf{e})$ are spin ± 2 spherical harmonics. More generally, spin s spherical harmonics are defined by

$${}_s Y_{lm} = \begin{cases} \sqrt{\frac{(l-s)!}{(l+s)!}} \bar{\partial}^s Y_{lm} & \text{for } s \geq 0 \\ (-1)^s \sqrt{\frac{(l-|s|)!}{(l+|s|)!}} \bar{\partial}^{|s|} Y_{lm} & \text{for } s \leq 0. \end{cases} \quad (6.30)$$

The spin s spherical harmonics are a complete, orthonormal set for expanding (square-integrable) spin s functions on the sphere. Some of their properties are reviewed below in Sec. 6.2.1. Before doing so, let us gain some intuition for E -modes and B -modes by considering the following example.

Example: E and B -mode polarization from axisymmetric scalar potentials. In this case, P_E (and P_B) only depend on θ and so the Stokes parameters reduce to

$$Q \pm iU = (1 - \mu^2) \frac{d^2}{d\mu^2} (P_E \pm iP_B), \quad (6.31)$$

where $\mu \equiv \cos \theta$. It follows that P_E generates only non-zero Q polarization, with

$$Q = (1 - \mu^2) \frac{d^2 P_E}{d\mu^2}, \quad (6.32)$$

and P_B generates only non-zero U polarization with

$$U = (1 - \mu^2) \frac{d^2 P_B}{d\mu^2}. \quad (6.33)$$

In the case of P_E , the polarization is thus radial or tangential about the symmetry axis, while for P_B it is at $\pm 45^\circ$ to the radial direction. Equivalently, the E -mode polarization is directed along or perpendicular to the direction in which its magnitude varies, while for B -modes it is rotated by 45° .

Quite generally, a given potential $P_E = P_B = P$ generates E - and B -mode polarization with

$$\begin{aligned} (Q + iU)^{(E)} &= \bar{\partial} \bar{\partial} P && (E\text{-mode}) \\ (Q + iU)^{(B)} &= i \bar{\partial} \bar{\partial} P && (B\text{-mode}). \end{aligned} \quad (6.34)$$

It follows that $(Q + iU)^{(B)} = i(Q + iU)^{(E)}$, so the B -mode pattern is obtained from the E -mode by rotating the polarization direction everywhere by $\pi/4$.

6.2.1 Properties of spin-weighted spherical harmonics

Here we list, mostly without proof, some useful properties of the spin-weighted spherical harmonics.

- The ${}_s Y_{lm}$ are orthonormal over the sphere (for a given s):

$$\int d\hat{\mathbf{n}} {}_s Y_{lms} Y_{l'm'}^* = \delta_{ll'} \delta_{mm'} . \quad (6.35)$$

- Under complex conjugation,

$${}_s Y_{lm}^* = (-1)^{s+m} {}_{-s} Y_{l-m} . \quad (6.36)$$

- The parity properties of the ${}_s Y_{lm}$ are

$${}_s Y_{lm}(-\hat{\mathbf{n}}) = (-1)^l {}_{-s} Y_{lm}(\hat{\mathbf{n}}) . \quad (6.37)$$

- The spin s spherical harmonics can be expressed in terms of the Wigner D matrices [c.f. Eq. (1.50)] as

$$D_{-ms}^l(\phi, \theta, 0) = (-1)^m \sqrt{\frac{4\pi}{2l+1}} {}_s Y_{lm}(\theta, \phi) . \quad (6.38)$$

- Equation (1.51) reducing the product of two D -matrices implies that

$$\begin{aligned} {}_{s_1} Y_{l_1 m_1}(\hat{\mathbf{n}}) {}_{s_2} Y_{l_2 m_2}(\hat{\mathbf{n}}) &= \sum_{LMS} (-1)^{l_1 + l_2 + L} \sqrt{\frac{(2l_1 + 1)(2l_2 + 1)(2L + 1)}{4\pi}} \\ &\times \begin{pmatrix} l_1 & l_2 & L \\ m_1 & m_2 & M \end{pmatrix} \begin{pmatrix} l_1 & l_2 & L \\ s_1 & s_2 & S \end{pmatrix} {}_S Y_{LM}^*(\hat{\mathbf{n}}) . \end{aligned} \quad (6.39)$$

6.3 Statistics of CMB polarization

The decomposition of the polarization field into E and B modes is invariant under rotations, and the E and B multipoles transform like those for the temperature anisotropies, so that

$$\begin{aligned} E_{lm} &\rightarrow \sum_{m'} D_{mm'}^l E_{lm'} \\ B_{lm} &\rightarrow \sum_{m'} D_{mm'}^l B_{lm'} . \end{aligned} \quad (6.40)$$

It follows that the two-point correlators are of the form

$$\begin{aligned}\langle E_{lm} E_{l'm'}^* \rangle &= C_l^E \delta_{ll'} \delta_{mm'} \\ \langle B_{lm} B_{l'm'}^* \rangle &= C_l^B \delta_{ll'} \delta_{mm'}.\end{aligned}\quad (6.41)$$

The cross-correlations of Θ , E and B are further restricted if we assume that the fluctuations are statistically invariant under parity transformations. This will be the case if the mechanism that generates the primordial fluctuations respects parity, and there is no parity-violating physics that subsequently processes the primordial fluctuations. Under active spatial inversion, the electric field at \mathbf{e} maps to minus the field at $-\mathbf{e}$. Moreover, $\hat{\theta}(\mathbf{e}) = \hat{\theta}(-\mathbf{e})$ but $\hat{\phi}(-\mathbf{e}) = -\hat{\phi}(-\mathbf{e})$. It follows that $Q(\mathbf{e}) \rightarrow Q(-\mathbf{e})$ while $U(\mathbf{e}) \rightarrow -U(-\mathbf{e})$. If \hat{P} is the parity operator, we therefore have

$$\begin{aligned}[\hat{P}(Q \pm iU)](\mathbf{e}) &= (Q \mp iU)(-\mathbf{e}) \\ \Rightarrow \sum_{lm} \hat{P}(E_{lm} \pm iB_{lm})_{\pm 2} Y_{lm}(\mathbf{e}) &= \sum_{lm} (E_{lm} \mp iB_{lm})_{\mp 2} Y_{lm}(-\mathbf{e}) \\ &= \sum_{lm} (-1)^l (E_{lm} \mp iB_{lm})_{\pm 2} Y_{lm}(\mathbf{e}).\end{aligned}\quad (6.42)$$

It follows that under parity $E_{lm} \rightarrow (-1)^l E_{lm}$ (*electric* parity) while $B_{lm} \rightarrow (-1)^{l+1} B_{lm}$ (*magnetic* parity).

Combining statistical isotropy and parity invariance means there can be no 2-point correlations between B and either Θ or E .

6.4 Boltzmann equation for polarization

The polarization of the CMB is first order in perturbations; it vanishes at zero order by isotropy. We can therefore ignore the gravitational redshifting effect of perturbations on the comoving energy since this only generates second-order terms in the polarized Boltzmann equation. We noted earlier that the Stokes parameters are (exactly) conserved along the photon path in phase space provided we define them relative to the screen-projection of a parallel-propagated basis. Generally, as \mathbf{e} evolves along the photon path (due to gravitational lensing), a screen-projected parallel-propagated polarization basis will rotate with respect to the polar basis vectors $\hat{\theta}(\mathbf{e})$ and $\hat{\phi}(\mathbf{e})$. However, this rotation is also first-order in perturbations so we can treat the Stokes parameters as constant on the polar basis vectors that we use to define polarization directions in the absence of scattering.

With these considerations, we can write the Boltzmann equation for the linear polarization as

$$\frac{\partial}{\partial \eta} (Q \pm iU) + \mathbf{e} \cdot \nabla (Q \pm iU) = \left. \frac{d}{d\eta} (Q \pm iU) \right|_{\text{scatt.}}, \quad (6.43)$$

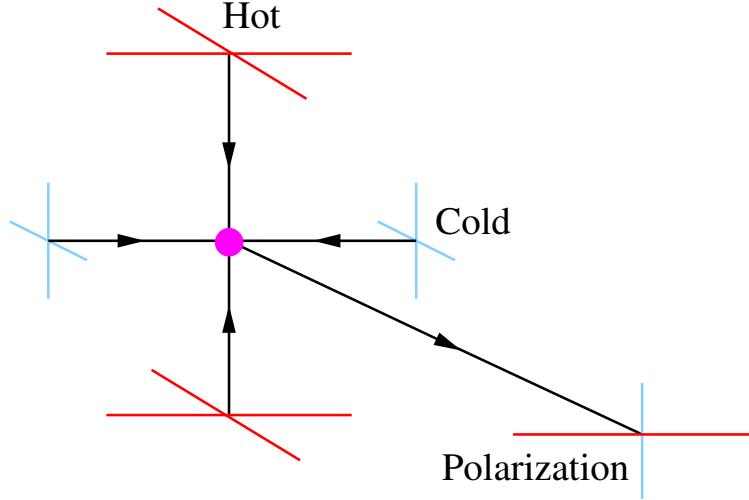


Figure 15: Generation of polarization by Thomson scattering of unpolarized radiation. A quadrupole variation of the radiation temperature in the plane perpendicular to the scattering direction generates linear polarization.

where the right-hand side describes the effect of scattering.

6.4.1 Generation of polarization by scattering (details non-examinable)

Linear polarization of the CMB is generated by Thomson scattering of the quadrupole of the temperature anisotropy around recombination. A cartoon of this mechanism is given in Fig. 15. In this section we derive the scattering rate for this process that appears in the polarized Boltzmann equation (6.43).

To keep things simpler, we shall first discuss scattering of initially unpolarized radiation from electrons at rest. Consider scattering from \mathbf{e} into \mathbf{e}' , and introduce polarization bases at \mathbf{e} and \mathbf{e}' with the local x (and x') axes in the scattering plane and the local y (and y') axes perpendicular to the plane. Denote the polarization produced by scattering into \mathbf{e}' , on the basis adapted to the scattering plane, by $d\bar{Q}(\mathbf{e}')$ and $d\bar{U}(\mathbf{e}')$. Then, if the scattered total intensity is

$$dI(\mathbf{e}') = \frac{3}{16\pi} |d\tau| (1 + \cos^2 \beta) I(\mathbf{e}) d\mathbf{e}, \quad (6.44)$$

where $d\tau$ is the increment in optical depth and β is the (scattering) angle between \mathbf{e} and \mathbf{e}' , standard results for dipole scattering give

$$d\bar{Q}(\mathbf{e}') = \frac{3}{16\pi} |d\tau| (-1 + \cos^2 \beta) I(\mathbf{e}) d\mathbf{e}, \quad d\bar{U}(\mathbf{e}') = 0. \quad (6.45)$$

Before we can integrate over \mathbf{e} , we must rotate the polarization at \mathbf{e}' to a common basis. We shall use the polar-coordinate basis there, for which $\boldsymbol{\theta}'$ and $\boldsymbol{\phi}'$ are obtained from the scattering-plane x' and y' directions by some rotation $-\gamma'$ in a right-handed sense about \mathbf{e}' . The corresponding angle at \mathbf{e} is $-\gamma$. Denoting the scattered polarization on the polar-coordinate basis by $dQ(\mathbf{e}')$ and $dU(\mathbf{e}')$, we have

$$\begin{aligned} d(Q \pm iU)(\mathbf{e}') &= e^{\pm 2i\gamma'} d(\bar{Q} \pm i\bar{U}) \\ &= -\frac{3}{16\pi} |d\tau| e^{\pm 2i\gamma'} \sin^2 \beta I(\mathbf{e}) d\mathbf{e}. \end{aligned} \quad (6.46)$$

To handle the $e^{\pm 2i\gamma'} \sin^2 \beta$ factor, we can use the spin-weighted generalisation of the addition theorem for spherical harmonics:

$$\sum_{|m| \leq l} {}_s Y_{lm}^*(\mathbf{e}) {}_{s'} Y_{lm}(\mathbf{e}') = \frac{2l+1}{4\pi} D_{ss'}^l(\gamma, \beta, -\gamma'). \quad (6.47)$$

To derive this result, note that we can obtain the triad formed from the scattering-plane x and y directions and the outward normal \mathbf{e} from the global x , y and z frame by applying a rotation $D(\phi, \theta, \gamma)$. Similarly, we obtain the triad at \mathbf{e}' by rotating with $D(\phi', \theta', \gamma')$. However, we can also form the triad at \mathbf{e}' by first rotating the global x , y , z triad with $D(0, \beta, 0)$ and subsequently rotating with $D(\phi, \theta, \gamma)$. It follows that

$$D(\phi', \theta', \gamma') = D(\phi, \theta, \gamma) D(0, \beta, 0) \Rightarrow D^{-1}(\phi, \theta, \gamma) D(\phi', \theta', \gamma') = D(0, \beta, 0). \quad (6.48)$$

We now express the rotations in terms of their matrix representations, $D_{mm'}^l$ and use unitarity to deal with the inverse rotation $D^{-1}(\phi, \theta, \gamma)$. Finally, the relation between the spin-weighted harmonics and the rotation matrices, Eq. (6.38), establishes the addition theorem in Eq. (6.47).

Armed with the addition theorem, and noting that $\sin^2 \beta = \sqrt{8/3} d_{0\pm 2}^2(\beta)$, we can write

$$\begin{aligned} \sin^2 \beta e^{\pm 2i\gamma'} &= \sqrt{\frac{8}{3}} d_{0\pm 2}^2(\beta) e^{\pm 2i\gamma'} \\ &= \sqrt{\frac{8}{3}} D_{0\pm 2}^2(0, \beta, -\gamma') \\ &= \sqrt{\frac{8}{3}} \frac{4\pi}{5} \sum_{|m| \leq 2} {}_2 Y_{2m}^*(\mathbf{e}) {}_{\pm 2} Y_{2m}(\mathbf{e}'). \end{aligned} \quad (6.49)$$

Integrating Eq. (6.46) over \mathbf{e} is now trivial, and isolates the quadrupole of the incident intensity. Expressing things in terms of temperature anisotropies and temperature-equivalent Stokes parameters, we simply have

$$d(Q \pm iU)(\mathbf{e}') = -\frac{3}{5} |d\tau| \sum_{|m| \leq 2} \frac{1}{\sqrt{6}} \Theta_{2m \pm 2} Y_{2m}(\mathbf{e}'), \quad (6.50)$$

where $\Theta(\mathbf{e}) = \sum_{lm} \Theta_{lm} Y_{lm}(\mathbf{e})$. Writing $Q \pm iU = \sum_{lm} (E_{lm} \pm iB_{lm})_{\pm 2} Y_{lm}$, we see that the scattered radiation is locally an E -mode quadrupole.

Although we derived Eq. (6.50) for electrons at rest, in linear theory we can use this for electrons with a bulk velocity. This is because polarization is necessarily first-order in perturbations, so the change in polarization in transforming from the electrons' rest-frame is second order.

It is left as an exercise (on Examples 3) to generalise the results of this section for scattering of polarized radiation. You should find that the result for the linear polarization is

$$d(Q \pm iU)(\mathbf{e}') = -|d\tau|(Q \pm iU)(\mathbf{e}') + \frac{3}{5}|d\tau| \sum_{|m| \leq 2} \left(E_{2m} - \frac{1}{\sqrt{6}}\Theta_{2m} \right) {}_{\pm 2}Y_{2m}(\mathbf{e}'), \quad (6.51)$$

where the first term arises from photons scattering out of the beam, and the second from scattering into the beam. Note how the in-scattering term is modified by the E -mode quadrupole of the incident radiation. Moreover, the collision term for the total intensity is modified by the incident polarization:

$$d\Theta(\mathbf{e}') = -|d\tau|[\Theta(\mathbf{e}') - \Theta_0] + \frac{1}{10}|d\tau| \sum_{|m| \leq 2} \left(\Theta_{2m} - \sqrt{6}E_{2m} \right) Y_{2m}(\mathbf{e}'), \quad (6.52)$$

where, recall, $\Theta_0 = \Theta_{00}/\sqrt{4\pi}$ is the monopole part of the temperature anisotropy (which has a spherical multipole Θ_{00}).

Finally, we can write the Boltzmann equation for linear polarization:

$$\frac{d(Q \pm iU)}{d\eta} = \dot{\tau}(Q \pm iU) - \frac{3}{5}\dot{\tau} \sum_{|m| \leq 2} \left(E_{2m} - \frac{1}{\sqrt{6}}\Theta_{2m} \right) {}_{\pm 2}Y_{2m}(\mathbf{e}), \quad (6.53)$$

where the derivative $d/d\eta$ is along the background lightcone. Integrating gives the observed polarization

$$(Q \pm iU)(\eta_0, \mathbf{x}_0, \mathbf{e}) = -\frac{\sqrt{6}}{10} \sum_m \int_0^{\eta_0} d\eta' g(\eta') (\Theta_{2m} - \sqrt{6}E_{2m})(\eta', \mathbf{x}_0 - \chi\mathbf{e}) {}_{\pm 2}Y_{2m}(\mathbf{e}), \quad (6.54)$$

where $\chi = \eta_0 - \eta'$. In the absence of reionization, the only contribution to the line-of-sight integral is from around recombination. There, the polarization is locally a pure E -mode quadrupole. However, due to spatial inhomogeneities over the last-scattering surface, the observed polarization will not generally retain the simple character it has at last scattering. To see how this works in detail, it is simplest to consider the important cases of density perturbations and gravitational waves separately.

6.5 Polarization from scalar perturbations

If we ignore reionization and approximate last scattering as instantaneous, Eq. (6.54) reduces to

$$(Q \pm iU)(\eta_0, \mathbf{x}_0, \mathbf{e}) \approx -\frac{\sqrt{6}}{10} \sum_m (\Theta_{2m} - \sqrt{6}E_{2m})(\eta_*, \mathbf{x}_*) \pm_2 Y_{2m}(\mathbf{e}). \quad (6.55)$$

Here, the point (η_*, \mathbf{x}_*) lies on the last-scattering surface of the observation event, i.e., $\eta_0 = \eta_* + \chi_*$ and $\mathbf{x}_0 = \mathbf{x}_* + \chi_* \mathbf{e}$ where χ_* is the distance to last-scattering.

We saw in Sec. 4.1 that, for scalar perturbations, we can write the spherical multipoles of the Fourier transform of the temperature anisotropy as

$$\Theta_{lm}(\eta, \mathbf{k}) = (-i)^l \frac{4\pi}{2l+1} \Theta_l(\eta, \mathbf{k}) Y_{lm}^*(\hat{\mathbf{k}}), \quad (6.56)$$

so that

$$\Theta_{2m}(\eta, \mathbf{k}) = -\frac{4\pi}{5} \Theta_2(\eta, \mathbf{k}) Y_{2m}^*(\hat{\mathbf{k}}). \quad (6.57)$$

Since the polarization fields E and B are scalar fields over the sphere, like Θ , they can be expanded in normal modes in the same way, so that

$$(Q \pm iU)(\eta, \mathbf{x}, \mathbf{e}) = \int \frac{d^3\mathbf{k}}{(2\pi)^{3/2}} (Q \pm iU)(\eta, \mathbf{k}, \mathbf{e}) e^{i\mathbf{k}\cdot\mathbf{x}}$$

with $(Q \pm iU)(\eta, \mathbf{k}, \mathbf{e}) = \sum_{lm} (-i)^l \frac{4\pi}{2l+1} (E_l \pm iB_l)(\eta, \mathbf{k}) Y_{lm}^*(\hat{\mathbf{k}}) \pm_2 Y_{lm}(\mathbf{e}). \quad (6.58)$

Taking the Fourier transform of Eq. (6.55) (with respect to \mathbf{x}_0) we have

$$(Q \pm iU)(\eta_0, \mathbf{k}, \mathbf{e}) = \frac{\sqrt{6}}{10} \frac{4\pi}{5} (\Theta_2 - \sqrt{6}E_2)(\eta_*, \mathbf{k}) e^{-ik\chi_* \hat{\mathbf{k}} \cdot \mathbf{e}} \sum_m Y_{2m}^*(\hat{\mathbf{k}}) \pm_2 Y_{2m}(\mathbf{e}). \quad (6.59)$$

Following our treatment of gravitational waves, it is simplest to consider first the case where \mathbf{k} is along the z -direction. The observed E and B -mode multipoles from other Fourier modes can then be obtained by a suitable rotation.

The local form of the polarization at the last-scattering event is dictated by the last term (i.e., the summation over m) in Eq. (6.59). Taking $\hat{\mathbf{k}} = \hat{\mathbf{z}}$, we only have $m = 0$ modes, and, using

$$\pm_2 Y_{20}(\mathbf{e}) = \sqrt{\frac{5}{4\pi}} D_{0\pm_2}^2(\phi, \theta, 0) = \sqrt{\frac{5}{4\pi}} \sqrt{\frac{3}{8}} \sin^2 \theta, \quad (6.60)$$

we see that at last scattering $Q \propto \sin^2 \theta$ and $U = 0$. However, what we observe across the sky after free-streaming through χ_* is this local polarization modulated by

the plane wave $\exp(-ik\chi_* \cos \theta)$. The observed polarization still has $U = 0$, but the modulation transfers signal in Q from $l = 2$ to higher multipoles (with most appreciable power appearing at $l = k\chi_*$). Importantly, the E -mode character of the polarization is preserved since we have radial or tangential polarization with azimuthally-symmetric variation. This leads to the very significant result that *linear scalar perturbations do not produce any B -mode polarization*. As we shall see, the same is not true for gravitational waves, so that the B -mode of polarization is an excellent observable with which to search for the imprint of gravitational waves.

Let us now derive the multipoles of the observed polarization in detail. For $\hat{\mathbf{k}} = \hat{\mathbf{z}}$, Eq. (6.59) reduces to

$$\begin{aligned} (Q \pm iU)(\eta_0, k\hat{\mathbf{z}}, \mathbf{e}) &= \frac{\sqrt{6}}{10} \frac{4\pi}{\sqrt{5}} (\Theta_2 - \sqrt{6}E_2)(\eta_*, k\hat{\mathbf{z}}) \sum_{L \geq 0} (-i)^L \sqrt{2L+1} j_L(k\chi_*) Y_{L0}(\mathbf{e}) {}_{\pm 2}Y_{20}(\mathbf{e}) \\ &= \frac{\sqrt{24\pi}}{10} (\Theta_2 - \sqrt{6}E_2)(\eta_*, k\hat{\mathbf{z}}) \sum_{L \geq 0} \left[(-i)^L (2L+1) j_L(k\chi_*) \right. \\ &\quad \times \left. \sum_{l \geq 2} \sqrt{2l+1} \begin{pmatrix} 2 & L & l \\ \pm 2 & 0 & \mp 2 \end{pmatrix} \begin{pmatrix} 2 & L & l \\ 0 & 0 & 0 \end{pmatrix} {}_{\pm 2}Y_{l0}(\mathbf{e}) \right], \end{aligned} \quad (6.61)$$

where we have used the plane-wave expansion and Eq. (6.39) to reduce the product of (spin-weighted) spherical harmonics. This is very similar in form to temperature anisotropies from gravitational waves; performing the sum over L using the explicit form of the $3j$ symbols returns

$$\begin{aligned} (Q \pm iU)(\eta_0, k\hat{\mathbf{z}}, \mathbf{e}) &= -\frac{3\sqrt{4\pi}}{20} (\Theta_2 - \sqrt{6}E_2)(\eta_*, k\hat{\mathbf{z}}) \\ &\quad \times \sum_l (-i)^l \sqrt{2l+1} \sqrt{\frac{(l+2)!}{(l-2)!} \frac{j_l(k\chi_*)}{(k\chi_*)^2}} {}_{\pm 2}Y_{l0}(\mathbf{e}). \end{aligned} \quad (6.62)$$

Note that the spatial-to-angular projection of scalar polarization is controlled by the same function, $j_l(k\chi_*)/(k\chi_*)^2$, as for the temperature anisotropies from gravitational waves. Extracting the E and B -mode multipoles gives

$$\begin{aligned} (E_{lm} \pm iB_{lm})(\eta_0, k\hat{\mathbf{z}}) &= -\delta_{m0} \frac{3\sqrt{4\pi}}{20} (\Theta_2 - \sqrt{6}E_2)(\eta_*, k\hat{\mathbf{z}}) \\ &\quad \times (-i)^l \sqrt{2l+1} \sqrt{\frac{(l+2)!}{(l-2)!} \frac{j_l(k\chi_*)}{(k\chi_*)^2}}, \end{aligned} \quad (6.63)$$

from which we see that $B_{lm} = 0$, as anticipated from the discussion above. Finally, we can obtain the multipoles for a general \mathbf{k} by rotating. Since $D_{m0}^l(\phi_{\mathbf{k}}, \theta_{\mathbf{k}}, 0) =$

$\sqrt{4\pi/(2l+1)}Y_{lm}^*(\hat{\mathbf{k}})$, we find

$$E_{lm}(\eta_0, \mathbf{k}) = -(-i)^l \frac{6\pi}{10} \sqrt{\frac{(l+2)!}{(l-2)!}} (\Theta_2 - \sqrt{6}E_2)(\eta_*, \mathbf{k}) \frac{j_l(k\chi_*)}{(k\chi_*)^2} Y_{lm}^*(\hat{\mathbf{k}}), \quad (6.64)$$

and $B_{lm} = 0$. This is the expected normal-mode form for scalar perturbations, with

$$\frac{E_l(\eta_0, \mathbf{k})}{2l+1} = -\frac{3}{20} \sqrt{\frac{(l+2)!}{(l-2)!}} (\Theta_2 - \sqrt{6}E_2)(\eta_*, \mathbf{k}) \frac{j_l(k\chi_*)}{(k\chi_*)^2}, \quad (6.65)$$

and $B_l(\eta_0, \mathbf{k}) = 0$.

The angular power spectrum of the E -mode polarization is (for adiabatic perturbations)

$$\begin{aligned} C_l^{EE} &= \frac{4\pi}{(2l+1)^2} \int d\ln k \left[\frac{E_l(\eta_0, \mathbf{k})}{\mathcal{R}(\mathbf{k})} \right]^2 \mathcal{P}_{\mathcal{R}}(k) \\ &\approx 4\pi \left(\frac{3}{20} \right)^2 \frac{(l+2)!}{(l-2)!} \int d\ln k \left[\frac{(\Theta_2 - \sqrt{6}E_2)(\eta_*, \mathbf{k})}{\mathcal{R}(\mathbf{k})} \right]^2 \left(\frac{j_l(k\chi_*)}{(k\chi_*)^2} \right)^2 \mathcal{P}_{\mathcal{R}}(k). \end{aligned} \quad (6.66)$$

To compute the polarization power spectrum, we need Θ_2 and E_2 on the last-scattering surface. Generally, these can be obtained by solving truncated Boltzmann hierarchies for Θ_l and E_l (see Examples 3). However, we can gain some useful insight by working to first order in the tight-coupling approximation (Sec. 4.6.3). There, we argued that to first-order in tight-coupling

$$\Theta_2 = \frac{20}{27} k\dot{\tau}^{-1} v_b. \quad (6.67)$$

Allowing for the polarization correction to the scattering rate for Θ [Eq. (6.52)], the factor 20/27 gets replaced by 8/9 (exercise on Examples 3). Moreover, applying the tight-coupling approximation to the polarization Boltzmann equation (6.53), we have

$$\sqrt{6}E_2 = -3\Theta_2/2 \quad (\text{tight coupling}), \quad (6.68)$$

so that the polarization source term

$$\Theta_2 - \sqrt{6}E_2 = \frac{20}{9} k\dot{\tau}^{-1} v_b. \quad (6.69)$$

There are two immediate consequences of this result, as follows.

- Polarization traces the baryon velocity at last scattering. Since the velocity undergoes acoustic oscillations on sub-sound-horizon scales, we expect acoustic peaks in the polarization power spectrum. However, since the velocity is $\pi/2$ out of phase with the density by virtue of the continuity equation, the peaks in the polarization spectrum will be at multipoles that correspond to acoustic troughs in the power spectrum of the temperature anisotropies.

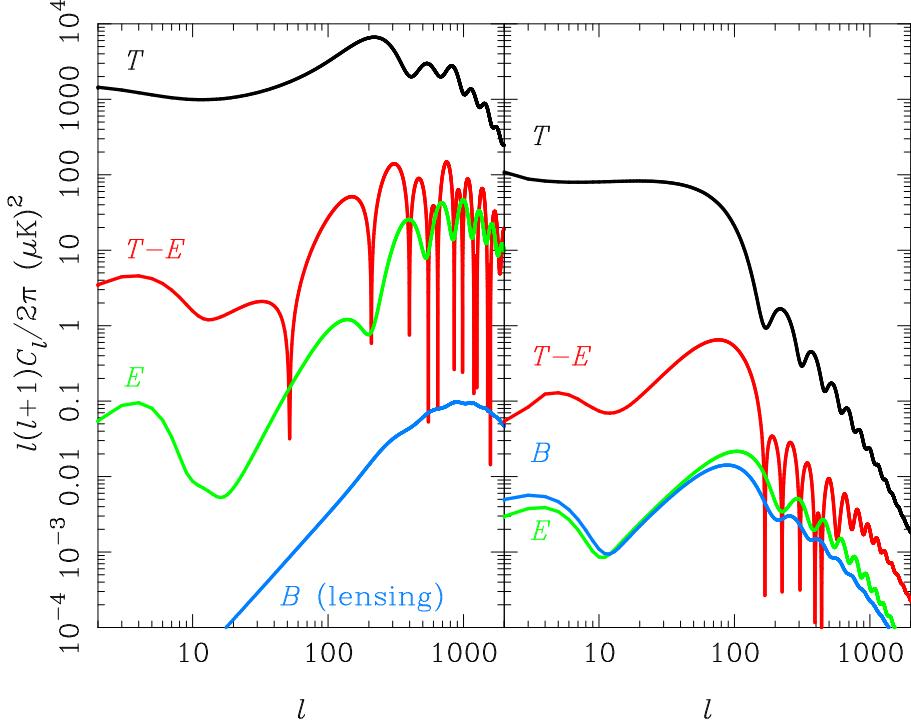


Figure 16: Temperature (black), E -mode (green), B -mode (blue) and $T-E$ cross-correlation (red) CMB power spectra from scalar perturbations (left) and tensor perturbations (gravitational waves; right). The amplitude of the tensor perturbations is shown for $r = 0.2$. At second order in perturbation theory, B -modes are produced from density perturbations. The dominant such effect is from weak gravitational lensing, and the associated spectrum of B -modes is also shown in the left-hand panel (blue).

- The temperature quadrupole at last scattering is small for large-scale modes since very little quadrupole anisotropy can build up by free-streaming between collisions when the mean free path is short compared to the perturbation wavelength. We therefore expect the large-angle polarization to be small. The polarization will peak at angular scales that correspond to the angle subtended by the mean free path at last scattering, since these modes have the largest quadrupole anisotropy at last scattering. (Smaller scales than this are diffusion damped.)

The polarization power spectra produced by adiabatic density perturbations are plotted in the left panel of Fig. 16. These are the result of a full numerical calculation. The E -mode power spectrum (plotted in green) shows the expected features discussed above. Note that it peaks around $l \sim 1000$. Large-angle polarization from the last-scattering surface is very small. The increase in polarization on large scales in Fig. 16 is due to re-scattering at reionization. Fourier modes that are comparable to the horizon size at reionization have developed a significant quadrupole temperature anisotropy by free-

streaming. Scattering of this quadrupole generates large-angle E -mode polarization, with amplitude proportional to the optical depth to reionization. Measurements of the large-angle polarization by Planck indicate that the optical depth to reionization is around $\tau \approx 0.08$, so that around 10% of CMB photons re-scattered after decoupling at recombination.

The temperature and E -mode polarization are correlated, since the baryon velocity is correlated with $\Theta_0 + \psi$. The cross-power spectrum is also plotted in Fig. 16. This has zeroes corresponding to Fourier modes that are caught at either the midpoint (which maximise E -mode polarization) or extrema of their oscillation (which maximise the temperature anisotropy) at last scattering.

The current best measurements of the E -mode power spectrum, and the cross-correlation between Θ and E are from Planck (Fig. 17). They show the remarkable agreement between the polarization measurements and the predicted theory spectra, based on the best-fitting Λ CDM model to the temperature anisotropies (red lines in the plots).

6.6 Polarization from gravitational waves

We now consider the generation of polarization from gravitational waves. We see from Eq. (5.37) that the Fourier transform of the Θ_{lm} take the form

$$\Theta_{lm}^{(\pm 2)}(\eta, \mathbf{k}) = \frac{1}{\sqrt{2}}(-i)^l \Theta_l^{(\pm 2)}(\eta, \mathbf{k}) \sqrt{\frac{4\pi}{2l+1}} D_{m\pm 2}^l(\phi_{\mathbf{k}}, \theta_{\mathbf{k}}, 0), \quad (6.70)$$

for helicity ± 2 gravitational waves, with equivalent results for $E_{lm}^{(\pm 2)}$ and $B_{lm}^{(\pm 2)}$. The Fourier transform of the observed polarization from a single helicity- p gravitational waves becomes [c.f. Eq. (6.59)]

$$(Q \pm iU)(\eta_0, \mathbf{k}, \mathbf{e}) = \frac{\sqrt{3}}{10} \sqrt{\frac{4\pi}{5}} (\Theta_2^{(p)} - \sqrt{6} E_2^{(p)}) (\eta_*, \mathbf{k}) e^{-ik\chi_* \hat{\mathbf{k}} \cdot \mathbf{e}} \sum_m D_{mp}^2(\phi_{\mathbf{k}}, \theta_{\mathbf{k}}, 0) {}_{\pm 2} Y_{2m}(\mathbf{e}). \quad (6.71)$$

Taking \mathbf{k} along $\hat{\mathbf{z}}$, we have only $m = p$ modes. Noting that

$${}_{\pm 2} Y_{2p}(\mathbf{e}) = \sqrt{\frac{5}{4\pi}} \left[e^{2i\phi} \begin{Bmatrix} \sin^4(\theta/2) \\ \cos^4(\theta/2) \end{Bmatrix} \delta_{p2} + e^{-2i\phi} \begin{Bmatrix} \cos^4(\theta/2) \\ \sin^4(\theta/2) \end{Bmatrix} \delta_{p-2} \right], \quad (6.72)$$

the form of the polarization at last scattering is (exercise!)

$$\begin{aligned} Q &\propto \frac{1}{2} ({}_{+2} Y_{2p} + {}_{-2} Y_{2p}) = \frac{1}{4} \sqrt{\frac{5}{4\pi}} (1 + \cos^2 \theta) (e^{2i\phi} \delta_{p2} + e^{-2i\phi} \delta_{p-2}) \\ iU &\propto \frac{1}{2} ({}_{+2} Y_{2p} - {}_{-2} Y_{2p}) = -\frac{1}{4} \sqrt{\frac{5}{4\pi}} \cos \theta (e^{2i\phi} \delta_{p2} - e^{-2i\phi} \delta_{p-2}). \end{aligned} \quad (6.73)$$

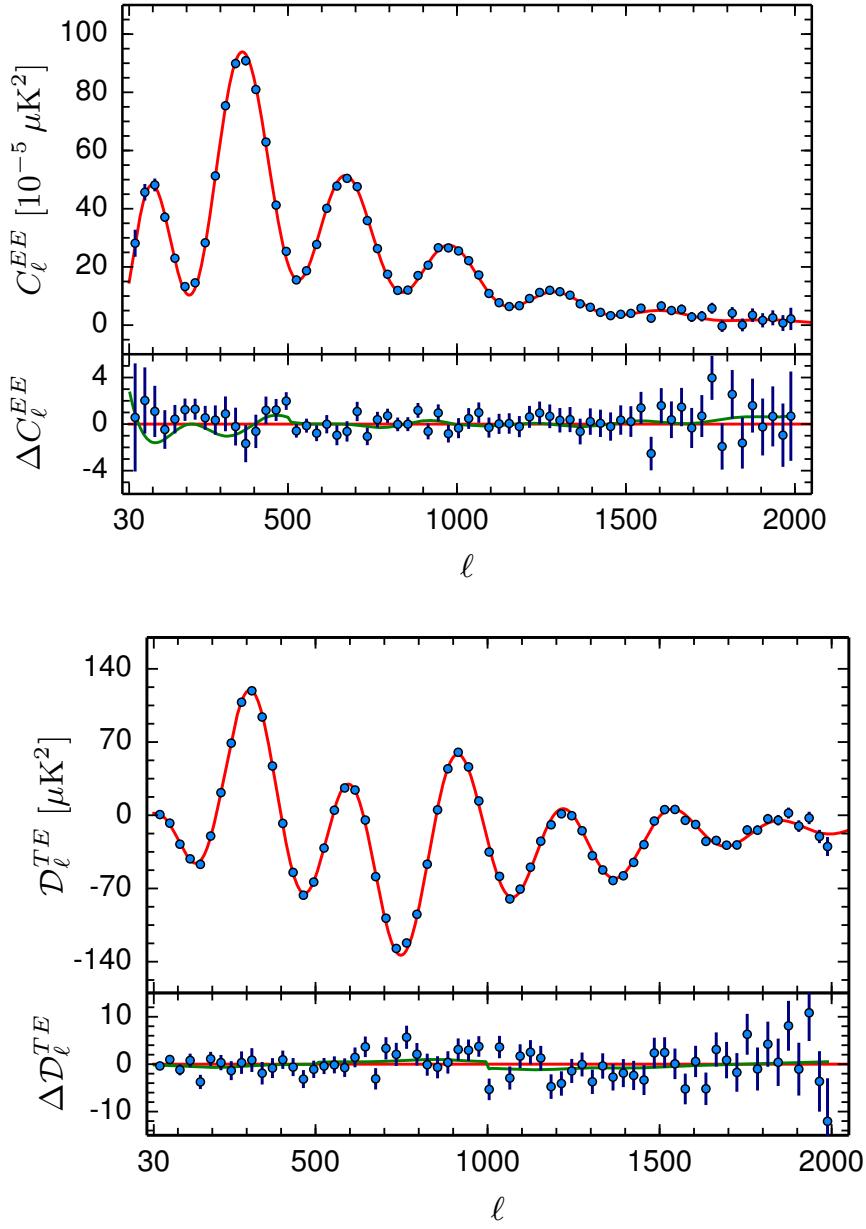


Figure 17: Polarization power spectra from the 2015 Planck release. The C_ℓ^E is shown in the top plot and the cross-correlation between Θ and E in the bottom plot. Note that $\mathcal{D}_\ell \equiv l(l+1)C_\ell/(2\pi)$. The (blue) points with error bars are the data points, while the (red) line is the prediction for ΛCDM models fit to the *temperature anisotropies alone*.

Unlike scalar perturbations, U is now non-zero. The observed polarization is this local polarization modulated by the plane wave $\exp(-ik\chi_* \hat{\mathbf{k}} \cdot \mathbf{e})$, and this generally has both E - and B -mode components.

To see this in detail, the observed polarization from a helicity- p gravitational wave is

$$\begin{aligned}
 (Q \pm iU)(\eta_0, k\hat{\mathbf{z}}, \mathbf{e}) &\propto {}_{\pm 2}Y_{2p}(\mathbf{e}) e^{-ik\chi_* \cos \theta} \\
 &= {}_{\pm 2}Y_{2p}(\mathbf{e}) \sum_L \sqrt{4\pi} \sqrt{2L+1} (-i)^L j_L(k\chi_*) Y_{L0}(\mathbf{e}) \\
 &= \sqrt{5} \sum_L \left[(2L+1) (-i)^L j_L(k\chi_*) \right. \\
 &\quad \times \sum_l \sqrt{2l+1} \begin{pmatrix} 2 & L & l \\ -p & 0 & p \end{pmatrix} \begin{pmatrix} 2 & L & l \\ \pm 2 & 0 & \mp 2 \end{pmatrix} {}_{\pm 2}Y_{lp}(\mathbf{e}) \Big] \\
 &= -\sqrt{5} \sum_l (-i)^l \sqrt{2l+1} {}_{\pm 2}Y_{lp}(\mathbf{e}) \left[\epsilon_l(k\chi_*) \pm \frac{p}{2} i \beta_l(k\chi_*) \right], \quad (6.74)
 \end{aligned}$$

where the functions

$$\epsilon_l(x) \equiv \frac{1}{4} \left[\frac{d^2 j_l(x)}{dx^2} + \frac{4}{x} \frac{dj_l(x)}{dx} + \left(\frac{2}{x^2} - 1 \right) j_l(x) \right] \quad (6.75)$$

$$\beta_l(x) \equiv \frac{1}{2} \left[\frac{dj_l(x)}{dx} + \frac{2}{x} j_l(x) \right] \quad (6.76)$$

follow from substituting the explicit form of the $3j$ symbols and using the recursion relations for the spherical Bessel functions. The new feature here is the $\beta_l(x)$, which arises from the $L = l \pm 1$ modes of the plane wave in Eq. (6.74) – these L modes are absent for scalar perturbations. It is the $\beta_l(x)$ terms that give rise to B -mode polarization. Reinstating factors from Eq. (6.71), extracting the E and B modes, and rotating to a general \mathbf{k} gives

$$\begin{aligned}
 E_{lm}^{(\pm 2)}(\eta_0, \mathbf{k}) &= -\frac{\sqrt{12\pi}}{10} (-i)^l \sqrt{2l+1} \left(\Theta_2^{(\pm 2)} - \sqrt{6} E_2^{(\pm 2)} \right) (\eta_*, \mathbf{k}) \epsilon_l(k\chi_*) D_{m\pm 2}^l(\phi_{\mathbf{k}}, \theta_{\mathbf{k}}, 0) \\
 B_{lm}^{(\pm 2)}(\eta_0, \mathbf{k}) &= \mp \frac{\sqrt{12\pi}}{10} (-i)^l \sqrt{2l+1} \left(\Theta_2^{(\pm 2)} - \sqrt{6} E_2^{(\pm 2)} \right) (\eta_*, \mathbf{k}) \beta_l(k\chi_*) D_{m\pm 2}^l(\phi_{\mathbf{k}}, \theta_{\mathbf{k}}, 0). \quad (6.77)
 \end{aligned}$$

The projection functions $\epsilon_l(x)$ and $\beta_l(x)$ are shown in Fig. 18. The projection onto E -modes peaks sharply at $k\chi_* = l$, while the projection onto B -modes is less sharp.

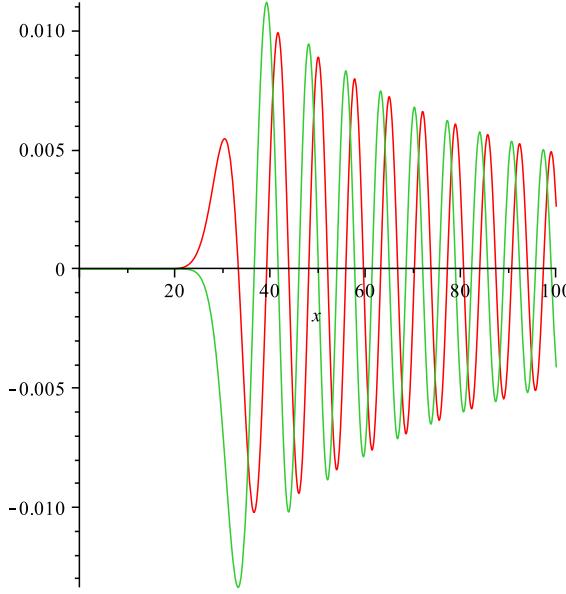


Figure 18: Polarization projection functions $\epsilon_l(x)$ (green) and $\beta_l(x)$ (red) for gravitational waves (with $l = 30$).

The E - and B -mode power spectra follow directly from Eq. (6.77):

$$\begin{aligned} C_l^E &\approx \frac{6\pi}{25} \int d \ln k \left[\frac{(\Theta_2^{(p)} - \sqrt{6}E_2^{(p)})(\eta_*, \mathbf{k})}{h^{(p)}(\mathbf{k})} \right]^2 \epsilon_l^2(k\chi_*) \mathcal{P}_h(k) \\ C_l^B &\approx \frac{6\pi}{25} \int d \ln k \left[\frac{(\Theta_2^{(p)} - \sqrt{6}E_2^{(p)})(\eta_*, \mathbf{k})}{h^{(p)}(\mathbf{k})} \right]^2 \beta_l^2(k\chi_*) \mathcal{P}_h(k). \end{aligned} \quad (6.78)$$

The source term $\Theta_2^{(p)}$ can be estimated in the tight-coupling limit from the general integral solution for the $\Theta_l^{(p)}(\eta, \mathbf{k})$ given in Eq. (5.38). Setting η_0 to the time of interest, and integrating over a mean free path l_p , in the limit that $kl_p \ll 1$ and $\mathcal{H}l_p \ll 1$ (so we can treat the $h^{(p)}$ and the projection function as constant), we find

$$\Theta_2^{(p)}(\eta, \mathbf{k}) \approx \frac{l_p}{\sqrt{6}} \dot{h}^{(p)}(\eta, \mathbf{k}). \quad (6.79)$$

This is equivalent to our physical expectation that a temperature quadrupole should build up over a scattering time due to the shear of a long wavelength gravitational wave:

$$\Theta(\mathbf{e}) \sim -\frac{1}{2} l_p \dot{h}_{ij} e^{\hat{i}} e^{\hat{j}}. \quad (6.80)$$

A more careful treatment of the tight-coupling limit, including the effects of polarization-dependent scattering, gives the source term in the tight-coupling limit as

$$(\Theta_2^{(p)} - \sqrt{6}E_2^{(p)})(\eta, \mathbf{k}) \approx \frac{5}{3\sqrt{3}}l_p \dot{h}^{(p)}(\eta, \mathbf{k}). \quad (6.81)$$

On scales larger than the horizon size at matter–radiation equality, the solution of the Einstein equation (5.19) for the $h^{(p)}(\eta, \mathbf{k})$ in matter domination is

$$h^{(p)}(\eta, \mathbf{k}) = 3 \frac{j_1(k\eta)}{k\eta} h^{(p)}(\mathbf{k}). \quad (6.82)$$

This solution at matter–radiation equality η_{eq} (with $k\eta_{\text{eq}} \ll 1$) matches smoothly onto the primordial $h^{(p)}(\mathbf{k})$. It follows that the shear is $O(k^2\eta)$ on super-horizon scales, and the large-angle polarization from recombination is small. The large-angle behaviour of the polarization is explored further in a problem on Examples 3, where it is shown that we expect roughly equal and constant power in E and B -mode polarization on large scales.

The polarization power spectra from a scale-invariant background of gravitational waves with tensor-to-scalar ratio $r = 0.2$ are shown in the right-hand panel of Fig. 16. As for temperature anisotropies, the spectra peak on scales corresponding to the horizon size at recombination. Note that for this r , which is close to the upper limit allowed by observations of the temperature anisotropies, the gravitational wave contribution to the temperature anisotropies and E -mode polarization is sub-dominant to the scalar perturbations. However, gravitational waves dominate the B -mode polarization on large angular scales. (On small scales, the non-linear B -modes induced by the action of gravitational lensing on the largely E -mode primary polarization are dominant.) This is important: it means that with sensitive enough observations, B -mode polarization can improve constraints on r significantly over what can be achieved with the temperature anisotropies. Note that the increase in polarization power on the largest scales in Fig. 16 is due to reionization.

Currently there are only upper limits on the B -mode power spectrum. There was much excitement in March 2014 when the BICEP2 team announced a detection of B -mode polarization on degree scales – the level of power they detected was consistent with primordial gravitational waves with $r \approx 0.2$. However, the BICEP2 measurement is at a single frequency (150 GHz), so that contamination from dust emission in our Galaxy could not be quantified. This has now become possible with the Planck polarization data at 353 GHz, which is a high enough frequency that dust dominates over the CMB even in parts of the sky with very low Galactic emission (such as the field observed by BICEP2). We now know that the BICEP2 signal is dominated by Galactic dust, and once this is cleaned out the direct constraint on the tensor-to-scalar ratio becomes

$$r < 0.12 \quad (95\% \text{ C.L.; } B\text{-modes only}). \quad (6.83)$$

This is very similar to the constraint from the temperature anisotropies. Combining with the temperature data from Planck gives the slightly tighter joint constraint $r < 0.08$ (see the green contours in Fig. 14), ruling out simple inflation models like $m^2\phi^2$ at high significance.

There are experiments taking data now that forecast constraints on r at the $O(10^{-2})$ level. Indeed, it is thought that there are no fundamental technical or cosmological obstacles to achieving $r \sim O(10^{-4})$, although it calls for arrays of thousands of detectors in space (or hundreds of thousands below the Earth's atmosphere). The search for B -mode polarization from gravitational waves takes us to the frontier of observational CMB research, and is a fitting place to end this part of the course.