

Theory and Observations of the Cosmic Background Radiation

J. Richard Bond

CIAR Cosmology Program

Canadian Institute for Theoretical Astrophysics

Abstract

These lecture notes describe the state of CMB research up to Spring 1996. The lectures were originally delivered at the Les Houches Session LX in August 1993, but as the written version evolved an attempt was made to keep up with the tremendous pace of development in CMB theory and observation. A beta-release preprint version was generated in Winter 1995. This final version corresponds to the published one, with minor further debugs corrected. It was updated to include the “final” COBE/FIRAS data, 4-year COBE/DMR and various smaller angle anisotropy data sets, and a brief discussion of the CMB satellites, MAP and Planck. Among the topics covered are: A comprehensive treatment of energy injection of all kinds at high redshift, and the constraints derived from the limits on spectral distortion of the CMB. The observation of anisotropies and phenomenological temperature power spectrum estimates. How 3D random fields illuminated by differential visibility map via transport to various 2D radiation patterns, with applications to primary sources such as the integrated Sachs Wolfe effect, photon bunching and the Doppler effect, and to secondary sources such as the Sunyaev-Zeldovich thermal effect and emission from dusty galaxies. The theory of primary anisotropies (those derived from linear perturbation theory) is developed and the historical path in calculational procedures to the current high precision emphasis is described. Procedures for calculating secondary anisotropies (those derived from nonlinear development) are also sketched. CMB observations are related to early universe fluctuation power spectra amplitudes and shapes for inflation-based models, and are connected to large scale structure observations and galaxy formation to constrain structure formation theories and the cosmological parameters that define them. Technical appendices give a comprehensive treatment of perturbation theory of the Einstein-Boltzmann equations, beginning from the nonlinear ADM equations of transport, conservation, constraint and geometrodynamics. Included are: a full treatment of momentum space gauge freedom as well as the coordinate freedom in the radiative transport equation, full derivations of polarized photon source functions, tensor (gravity wave) perturbations, gravitational lensing, the shear viscosity and thermal diffusions with polarization in the tight-coupling regime, *etc.*

in *Cosmology and Large Scale Structure*, pp. 469-674,
eds. R. Schaeffer, J. Silk, M. Spiro and J. Zinn-Justin, Elsevier (1996)

Contents

1	Introduction and basic properties	5
2	Spectral observations and constraints	9
3	Spectral distortion theory	13
3.1	Radiative transport in the expanding universe	13
3.2	Source functions for spectral distortions	17
3.2.1	Compton scattering and the Kompaneets source term	17
3.2.2	Bremsstrahlung	19
3.2.3	Double Compton scattering	21
3.2.4	Rayleigh scattering	22
3.2.5	Line radiation	22
3.2.6	Synchrotron	22
3.2.7	Dust grains	23
3.3	The cosmic photosphere and Bose–Einstein distortions	27
3.4	Recombination and photon decoupling	28
3.4.1	Hydrogen and Helium Recombination	28
3.4.2	Visibility and decoupling	33
3.5	Reionization of the universe	34
3.6	Post-recombination energy sources	36
4	Phenomenology of CMB anisotropy	41
4.1	Statistical measures of the radiation pattern: $C(\theta), \mathcal{C}_\ell, \dots$	42
4.2	Experimental arrangements and their filters	45
4.2.1	Pixel–pixel correlation filters	45
4.2.2	Beams and <i>dmr</i> and <i>firs</i>	48
4.2.3	2-Beams, 3-beams, oscillating beams,	49
4.3	Primary power spectra for inflation-based theories	51
4.4	2D spectra with tilt and a Gaussian coherence angle	55
4.5	Experimental band-powers: past and present	57
4.6	Measuring cosmological parameters with the CMB	65
5	Primary and secondary sources of anisotropy	76
5.1	Angular power spectra from 3D random source-fields	76
5.1.1	Simple sample sources	78
5.1.2	Angular power spectra for simple sample sources	79
5.1.3	Products of Bessel functions	81
5.1.4	Fourier derivation of the simple sample spectra at high ℓ	83
5.1.5	Narrow and broad visibilities	84
5.1.6	Power-law spectra with coherence scales in 3D and 2D	85
5.2	The primary primary anisotropy effects	86
5.2.1	Sachs–Wolfe, photon-bunching and Doppler sources	87
5.2.2	Longitudinal and synchronous pictures of the Sachs–Wolfe effect	87

5.2.3	Differential power spectrum and form factors	88
5.2.4	Damping	90
5.2.5	Early reionization form factors	91
5.2.6	The isocurvature effect on low multipoles	91
5.3	Secondary anisotropies	92
5.3.1	Sample secondary anisotropy power spectra	92
5.3.2	Anisotropy power from dusty primeval galaxies	94
5.3.3	SZ and nonlinear Thomson scattering from clusters	96
5.3.4	Single-cluster observations of the SZ effect	98
5.3.5	The maximum entropy nature of Gaussian anisotropies	101
5.3.6	Quadratic nonlinearities in Thomson scattering	101
5.3.7	The influence of weak gravitational lensing on the CMB	104
6	Perturbation theory of primary anisotropies	105
6.1	Overview of fluctuation formalism	105
6.2	Perturbed Einstein equations	107
6.2.1	Time-hypersurface and gauge freedom	107
6.2.2	Scalar mode Einstein equations	109
6.2.3	Useful gauge invariant combinations for scalar modes	111
6.2.4	Longitudinal and synchronous gauges	112
6.2.5	Tensor mode metric equations	113
6.3	Connection with primordial post-inflation power spectra	115
6.4	Relating scalar and tensor power measures to the $d\mathbf{m}r$ band-power	118
6.5	The Boltzmann transport equation	120
6.5.1	Scalar mode transfer equations	121
6.5.2	Tensor mode transfer equations	132
7	Connection with other cosmic probes of k-space	135
7.1	Density power spectra and characteristic scales	135
7.2	The observable range in $\mathbf{k}$ -space	137
7.3	Relating the cluster-amplitude σ_8 and the $d\mathbf{m}r$ band-power	140
7.4	The future	145
A	The ADM formalism and perturbation theory	146
A.1	The ADM equations	147
A.2	Scalar perturbations	151
A.3	Tensor perturbations	155
B	Transport theory in General Relativity	156
B.1	The distribution function and the BTE in GR	156
B.2	Number, energy and momentum conservation equations	159
B.3	The transport of extremely relativistic particles	161
B.4	momentum space gauge transformations	163

C	Polarized transport for Thomson scattering	166
C.1	The polarization matrix and Stokes parameters	166
C.2	Scalar perturbation source terms	171
C.2.1	Thomson source functions	171
C.2.2	The moment equations for photons	172
C.2.3	CDM and baryon transport	173
C.2.4	The transport of massless neutrinos	174
C.2.5	Hot and warm dark matter transport	175
C.3	Numerically useful regimes for scalar perturbations	178
C.3.1	Tight-coupling, shear viscosity and thermal diffusion . .	178
C.3.2	Free-streaming	182
C.4	Modifications with mean curvature	183
C.5	Lensing	185
C.6	Tensor perturbation source terms	188

1 Introduction and basic properties

We shall look back on this decade as a golden age for cosmic background radiation research, with signals unveiled by very high precision spectrum and angular anisotropy experiments revealing much about how structure arose in the Hubble patch in which we live. Although the theory was reasonably well developed before the observation of anisotropy, much new work on all aspects of CMB theory and phenomenology has occurred, to better place the new experimentation in a cosmological framework. Sample lecture notes and reviews giving earlier snapshots of the state of the art in theory are [1, 2, 3, 4, 5, 7, 6]. Peebles' book [8] covers some of the theoretical ground, the White et al. [7] Annual Review article gives a shorter overview and extensive references, while Partridge [9] covers experimental techniques. [10] gives a recent overview of cosmology and how the CMB fits in. These lectures will be about equally divided among: the spectrum and what the remarkable lack of distortion tells us (section 2, 3); the observations and phenomenology of anisotropies (section 2); primary and secondary sources of anisotropy (section 5); the coupled Einstein–Boltzmann equations which describe the development of primary anisotropies (section 6), including a set of Appendices providing much more detail about these equations and their solution; how the CMB results connect with large scale structure results (section 7). Emphasis is on inflation-based models of cosmic structure formation. What is not covered here is how topological defect theories of structure formation impact upon the CMB: for cosmic strings see e.g., [281, 282, 283] and for texture defects see e.g., [284, 285, 286, 287], and references therein.

In section 2, I review the status of spectrum observations: although important historic work from the ground, in balloons and in rockets shortward of a centimeter will now be a footnote to FIRAS, ground-based radio telescopes still control the spectral constraints at the long wavelength end. The long-sustained assault on the mm-wave CMB peak led to one strong distortion after another, each one stimulating a flurry of theoretical papers which, by now, have largely sorted out the issues of how early energy release in the Universe would have been processed into observable signals (section 3). The photon transport is rather simple for spectral distortion calculations, homogeneous radiative transfer, a warmup for the more complicated inhomogeneous transfer in random media required for the treatment of CMB anisotropies. The source functions describing the predominant emission, absorption, and scattering processes are given there (bremsstrahlung from Coulomb-scattered and Compton-scattered electrons, low energy Compton scattering, and interactions with any primeval dust present). Of course, it is the *secondary* anisotropies that would accompany these distortions that can give us insights into the structure at emission time. *Primary anisotropies* (section 6) are those that we can calculate with linear perturbation theory. The primary goal of theoretical anisotropy research is to work out detailed predictions within a given cosmic structure formation model of primary and secondary anisotropies as a function of scale. Because of the

linearity, primary anisotropies are the simplest to predict and offer the least ambiguous glimpse of the underlying fluctuations that define the structure formation theory. With detailed high precision observations, we expect to be able to use CMB anisotropies to measure various cosmological parameters to high accuracy (section 4.6). The nonlinearity inherent in secondary anisotropies makes those predictions more ambiguous.

If energy is injected early enough in the Universe, it is just reprocessed by interaction with the plasma into a Planck spectrum, albeit with a higher entropy than the starting state. We must rely on indirect arguments based on primordial nucleosynthesis to constrain exactly when the entropy of our Hubble patch came into being, and this only if it was injected later than a redshift of ten billion. The cosmic photosphere exists around a redshift of ten million or so. With a FIRAS temperature of $T_\gamma(\text{now}) \equiv T_{c*} = 2.728 \pm 0.004$ K [12, 11], the entropy per comoving volume is

$$s_{\gamma*} = \frac{4\pi^2}{45} (\bar{a}T_\gamma)^3 \left[\frac{k_B}{\hbar c} \right]^3 = 1.48 \times 10^3 \text{ cm}^{-3}. \quad (1)$$

The (mean) scale factor of the Universe is $\bar{a}$, which I take to be normalized to be unity at the present time, so that it is related to the redshift z by $\bar{a} = (1+z)^{-1}$. I also invariably take the temperature to be in energy units, which is equivalent to taking Boltzmann's constant k_B to be unity. Recall that $1 \text{ eV} = 1.16 \times 10^4 \text{ K}$. As well, $\hbar$ and c are taken to be unity. Returning from these theorist units to the real world requires insertion of as many factors of $\hbar c = 0.1973 \text{ eV } \mu\text{m}$ as are necessary to take the energy factors into lengths, and once that is done, c is inserted to take the lengths to time. Recall that for a Planck distribution of photons, we have a comoving number and energy density and a pressure

$$n_{\gamma*} = \frac{2\zeta_3}{\pi^2} T_{c*}^3, \quad \rho_{\gamma*} = \frac{3}{4} s_{\gamma*} T_\gamma \approx 2.7 n_{\gamma*} T_\gamma, \quad p_\gamma = \frac{1}{3} \rho_\gamma \approx 0.9 n_\gamma T_\gamma, \quad (2)$$

where $T_{\gamma*} \equiv \bar{a}T_\gamma = T_{c*}$; numerically,

$$n_{\gamma*} = 412 \text{ cm}^{-3}, \quad \rho_{\gamma*} = 0.261 \text{ eV cm}^{-3}, \quad \Omega_\gamma h^2 = 2.46 \times 10^{-5}, \quad (3)$$

scaling as the appropriate power of $(T_{\gamma*}/2.728 \text{ K})$. The Hubble parameter is $h \equiv H_0/(100 \text{ km s}^{-1} \text{ Mpc}^{-1})$.

Because the expansion of the Universe is adiabatic, the photon entropy per comoving volume is a conserved quantity. If we suppose the entropy was generated early enough that neutrinos and e^+e^- pairs would have been in thermodynamic equilibrium with the photons ($T > 1 \text{ MeV}$), then the annihilation of the e^+e^- pairs into photons, when the temperature was about a few hundred keV, would have increased the photon entropy from $\frac{4}{11} s_{\gamma*}$ to the $s_{\gamma*}$ we observe. The comoving neutrino temperature would have remained that associated with the lower entropy level per particle, $\bar{a}T_\nu = (4/11)^{1/3} \bar{a}T_\gamma$, i.e., 1.95 K. The total entropy, apart from the minor bits residing in the gas and

stellar radiation of our Universe, is fully determined by the single number T_γ and the number of low mass neutrino degrees of freedom:

$$s_{tot*} = s_{\gamma*} + s_{\nu*} = s_{\gamma*} + \frac{7}{8} \frac{4}{11} N_\nu s_{\gamma*} = \frac{43}{22} s_{\gamma*} = 2.90 \times 10^3 \text{ cm}^{-3}, \quad (4)$$

for $N_\nu = 3$ light neutrino generations, each contributing a left-handed particle and an antiparticle, but no right-handed components. The associated energy density parameter for relativistic particles is $\Omega_{er} h^2 = (1 + \frac{7}{8} (4/11)^{4/3} N_\nu) \Omega_\gamma h^2$, 4.1×10^{-5} for $N_\nu = 3$. The origin of s_{tot*} is, of course, a mystery, enshrouded by the cosmic photosphere. It used to be considered to be a gift of the Planck era. In inflation models, our patch of the Universe was once in accelerated expansion, during which any primordial temperature would have dropped to essentially zero. s_{tot*} could then have arisen only once deceleration began, for only then could coherent field energy have been able to dissipate into entropy.

It is usual to divide the entropy by another (partially) conserved quantity, the comoving baryon number density, $n_{B*} = 1.13 \times 10^{-5} \Omega_B h^2$, expressed in terms of the baryon density parameter (relative to the critical density) Ω_B . After the identification of physical processes that could plausibly have led to the generation of baryon number from a hot medium, it became usual to invert the large numbers

$$\frac{s_{tot*}}{n_{B*}} = 2.56 \times 10^{10} \left(\frac{\Omega_B h^2}{0.01} \right)^{-1}, \quad \frac{s_{\gamma*}}{n_{B*}} = 1.31 \times 10^{10} \left(\frac{\Omega_B h^2}{0.01} \right)^{-1},$$

and try to explain why the baryon number is so tiny relative to the entropy through the extreme weakness of baryon-violating interactions.

A nice way to picture CMB transport from the early Universe to the present is to consider when and where various phenomena occurred on our past light cone, *when* defined by redshift, *where* defined by comoving distance to that redshift. For concrete numbers, I shall take the example of a universe with a critical density in nonrelativistic matter, $\Omega_{nr} = 1$, where Ω_{nr} has contributions from cold dark matter, baryons, etc. The cosmic photosphere is then $5924 \text{ h}^{-1} \text{ Mpc}$ away from us, very close to our “horizon”, $\approx 2H_0^{-1} = 6000 \text{ h}^{-1} \text{ Mpc}$. (Of course, inflation could have made the true event horizon much bigger; some process must have.) FIRAS gives stringent upper limits to distortions of various types. For example, the photon chemical potential constraint strongly limits the energy output that occurred just shortward of the cosmic photosphere (within about $200 \text{ h}^{-1} \text{ kpc}$ comoving distance from it). Barring early energy input which escapes the COBE bounds, the photons decouple at a redshift ≈ 1000 , a distance $5796 \text{ h}^{-1} \text{ Mpc}$ away, $128 \text{ h}^{-1} \text{ Mpc}$ from the photosphere. The shell between the photosphere and this last scattering surface where the Compton depth is unity defines an electron scattering “atmosphere”, quite thick to photons. In particular, when helium recombines, the photons are very tightly coupled.

The theory of the hydrogen atom (section 3.4) is so well known that we can be quite confident that we have the physics of recombination well described.

The essential ingredients were worked out immediately after the discovery of the CMB, and the novel feature is the dominant role that the two-photon decay of the $2s$ state to the $1s$ state plays. The width of the region over which decoupling takes place is only about $10 h^{-1}$ Mpc comoving distance (section 3.4.2). That the width is nonzero plays a fundamental role in defining how small the scale of anisotropies is that we can see. The relation between angular scale and comoving distance at high redshifts is about $d \approx 100 \Omega_{nr}^{-1/2} h^{-1}$ Mpc ($\theta/1^\circ$), hence we might expect that fluctuations on scales below about $10'$ are affected: they are strongly damped below this “coherence” angle and this will define which experiments are most useful to do if we wish to probe the moment when the photons were first released to freely propagate from their point of origin to us, without much further modification, apart from some gravitational redshifts, some lensing, and possibly some scattering from hot gas.

Even with the FIRAS limits, it is still quite possible that enough energy was injected either prior to recombination, or sufficiently shortly after (above redshift ~ 150) so that the photons had their decoupling delayed (section 3.5). The decoupling position then moves forward to $\sim 5570 h^{-1}$ Mpc. The thickness of the region over which decoupling would have taken place is more than an order of magnitude larger, $\sim 200 h^{-1}$ Mpc comoving distance, corresponding to a few degrees. Damping of anisotropies below a few degrees is the result, although nonlinear effects can lead to interesting short distance signatures of such early reionization (section 5.3.6).

Distortions of the background may occur before or after recombination. If it is Compton cooling of hot gas, the spectral signature of the y -distortion to the background radiation has allowed very powerful FIRAS constraints to be given on the Compton y -parameter which strongly rules out many models. If pregalactic dust, or dust in primeval galaxies, exists, it will absorb higher frequency radiation (UV and optical) and down-shift it into the infrared; combined with the redshift, a sub-mm background is expected but, with FIRAS, is now quite strongly constrained (section 3.2.7). Accompanying these secondary backgrounds are anisotropies that carry invaluable information about the epochs that the relevant structures formed. Even if the angle-averaged distortions are well below the level that absolute spectrum experiments like FIRAS probe, it is certain that these secondary anisotropies are accessible to experiments: the question is only for what fraction of the sky do they rise above experimental noise and the primary signal. A major goal of experimental/phenomenological anisotropy research is to design experiments and statistical processing procedures that will allow the various primary, secondary and foreground contributions to anisotropy to be separated (section 4). With the wealth of signals to be unveiled, we have a CMB future that “looks marvelous, simply marvelous”.

2 Spectral observations and constraints

We now know from COBE’s Far Infrared Absolute Spectrophotometer, FIRAS, that the CMB is well fit by a blackbody with $T \approx 2.728 \pm 0.004$ K over the region from $5000 \mu\text{m}$ to $500 \mu\text{m}$ [12, 11], a number compatible with the COBRA rocket experiment of Gush et al. [15] covering the same band, and also with ground based measurements at centimeter wavelengths – although there is still room for significant spectral distortion longward of 1 cm. Figure 1 gives a view of the current state of the data on thermodynamic temperature $T(\lambda)$ as a function of wavelength for FIRAS and selected experiments described below.

Following the Penzias and Wilson [13] discovery, during the 60s and 70s there were a large number of radio observations with coherent receivers that obtained $T(\lambda)$ in the Rayleigh–Jeans (RJ) portion of the spectrum. These results were reviewed by Weiss in 1980 [14], and a best fit temperature of 2.74 K was given, with a ± 0.08 K “one sigma” error. Throughout the 80s, a Berkeley–Italian team [17], the White Mountain collaboration, made measurements at many wavelengths, from 12 cm down to 0.33 cm, using corrugated horn antennas with 15° beamwidths switched from sky to a 3.8 K “cold load” calibrator. And Johnson and Wilkinson [18] used a balloon to get a temperature estimate at 1 cm.

Wavelengths longer than 10 cm are extremely difficult to explore, both because of large Galactic corrections that must be made and because of contamination from man-made radio signals. For many years a rather heroic early experiment by Howell and Shakeshaft in 1967 [16] was all that defined the constraints at long wavelengths. Recent experiments at the relatively radio-quiet South Pole [19, 20, 21, 22] have considerably improved the error bars. There are hints of deviation from the FIRAS temperature extrapolated into the RJ regime but the corrections are large.

One of the more remarkable aspects of the CMB story is that the population of rotational states of diatomic molecules – found by optical observations of interstellar absorption lines in the spectrum of bright stars – can be used to estimate CMB temperature. The first molecules discovered in interstellar space were CH and CN, found using the spectrum of ζ Ophiuchi. In 1941, McKellar inferred a 2.3 K excitation temperature to explain the relative intensity ratios of the lines originating from the $K = 0$ and $K = 1$ levels in the $3883 \text{ \AA}$ band of CN (frequency difference $(2640 \mu\text{m})^{-1}$, just longward of the CMB peak). This observation was very well known to the astronomical community, since it was given prominent play in the classic Herzberg text [23], although it was dismissed as having “only a very restricted meaning”. Had Gamow or his students Alpher and Herman made the connection, how different the development of cosmology might have been, but it was only in 1966, after the Penzias and Wilson discovery, that the connection was made. In 1972, Thaddeus [24] reviewed the CN work and gave $T(2640 \mu\text{m}) = 2.78 \pm 0.10$ K for ζ Ophiuchi, with much larger errors for other stellar spectra. In the 80s and 90s great improvements were made using very high signal-to-noise spectra

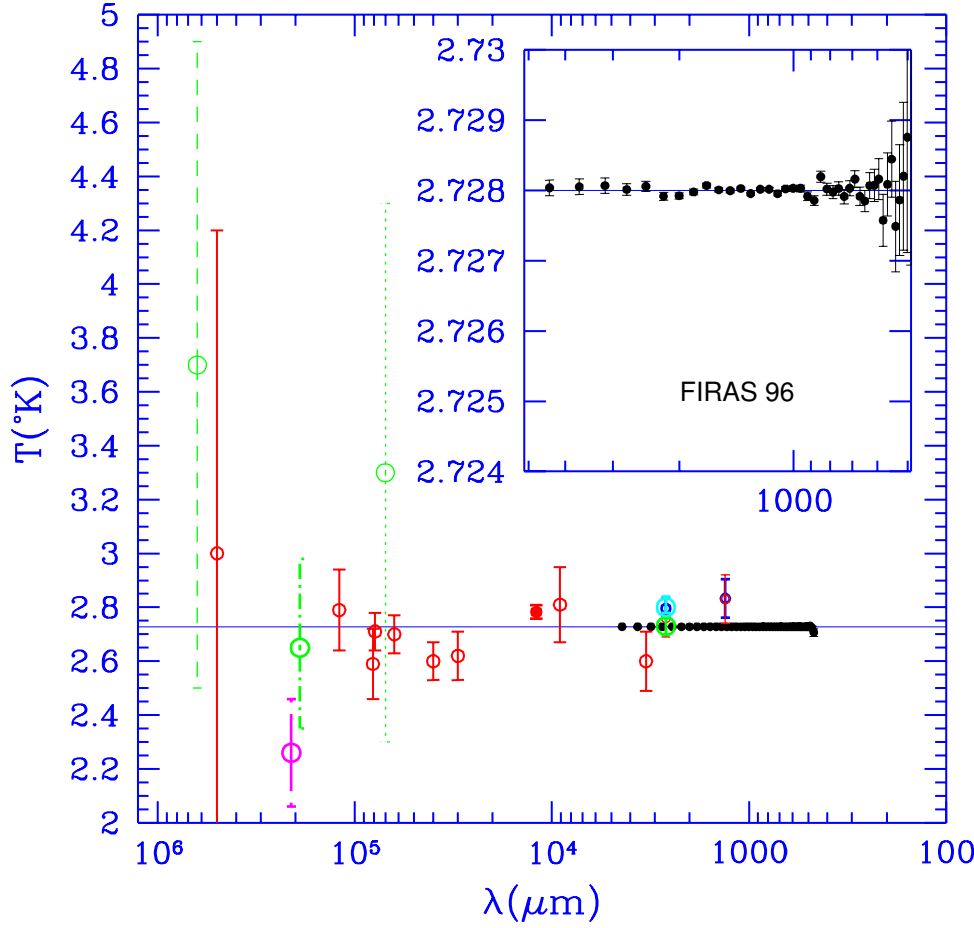


Figure 1: Selected old and new data on CMB distortions in terms of thermodynamic temperature. The dotted point at 7 cm is the original Penzias and Wilson (1965) result, the long-dashed point at 63 cm is from Howell and Shakeshaft (1966). The situation in the Rayleigh–Jeans region was improved quite a bit with the White Mountain collaboration results (solid). Results from Bersanelli (1995) at 21 cm and Staggs and Wilkinson (1995) at 19 cm are shown. The point with the small error bar at $\lambda = 1.2$ cm is that of Johnson and Wilkinson (1987). Cyanogen results are given at $2640 \mu\text{m}$ (Roth et al. 1993, Crane 1989, 1995). The tiny error bars are from FIRAS (Fixsen et al. 1996). The inset gives a blowup of the region for FIRAS.

of ζ Ophiuci; e.g., Meyer and Jura [25] got 2.73 ± 0.04 K at $\lambda = 0.264$ cm and 2.8 ± 0.3 K at $\lambda = 0.132$ cm, important at the time because it failed to confirm large reported excesses found with other techniques. Of course, excitation temperatures are really only upper bounds on the CMB temperature, since local contributions to the excitation, e.g., through collisions with electrons, might enhance the upper level's population. These frequencies overlap with those probed by FIRAS, and, within their much larger errors, agree with FIRAS.

The assault on the CMB peak and into the Wein region by the experimentalists proved very difficult, with distortion reports being the norm rather than the exception. In the 60s and early 70s there were rocket and balloon experiments which reported significant post-peak excesses, but Muehlner and Weiss (1973, reviewed in Weiss [14]), using five broad band filters, were able to show that these large excesses were not there. Around 1980, Woody and Richards [28] and Gush [29] used Fourier transform spectroscopy to get the spectrum around the peak, and both reported large (but qualitatively different) distortions, that cyanogen results and an experiment by Peterson, Richards and Timusk [30] failed to confirm. In 1988, a Nagoya-Berkeley rocket experiment with 6 broad band filters (Matsumoto et al. [31]) indicated a large excess energy content (about 20% of that in the CMB) that spurred much theoretical exploration of energy injection.

The issue was forever settled, to a standing ovation, at the famous January 1990 AAS presentation by COBE team leader John Mather of the perfect blackbody that 9 minutes of data taken shortly after the November 1989 launch revealed, with $T_{c*} = 2.735 \pm 0.06$ K, a result beautifully confirmed shortly after by the COBRA rocket experiment [15], with $T_{c*} = 2.736 \pm 0.02$ K.

Both experiments also used the elegant method of Fourier transform spectroscopy, with on-board reference blackbodies to compare with the sky signal. FIRAS also had an auxiliary external calibrator blackbody which could be moved in for further in-flight calibration. The FIRAS calibrators could range from 2 to 25 K in temperature, but when the sky was being observed, temperatures near 2.7 K were chosen to nearly null the difference between the internal and sky signals. FIRAS used polarizing Michelson interferometers, with mirrors that moved at constant velocity so that path difference was proportional to time lag, to construct the correlation between sky and reference blackbody as a function of time lag, an interferogram. FIRAS made two million of them. Fourier transform gives the power as a function of frequency. A dichroic filter split the FIRAS signal into low and high frequency parts, < 500 and $> 500 \mu\text{m}$ (with the best results from the low frequency 10^4 to $500 \mu\text{m}$ part).

With Fourier transform spectroscopy, determining the absolute temperature (which requires absolute calibration of the reference blackbodies) cannot be done with nearly the same precision as determining the level of deviation from a blackbody. The most complete analysis of the FIRAS data is [12], who used the full (all channels, nine months) low frequency data set, whereas [11] concentrated on the last 6 weeks of the FIRAS experiment, for which calibrations were frequent and the instruments were operating very stably. Models

were needed for the dipole, determined from the DMR experiment on COBE, and for the Galactic emission – modelled by $G(\ell, b)g(\nu)$, with the geometrical function of Galactic longitude and latitude, $G(\ell, b)$, taken from the DIRBE 240 μm map, and $g(\nu)$ from the FIRAS data. This is of course dominated by emission from the Galactic plane. There is evidence that the high b gas is colder than that using $g(\nu)$ determined in this way [48, 49].

The dipole amplitude is 3.372 ± 0.007 mK (95% *CL* error bars), i.e., $\Delta T/T = 1.2 \times 10^{-3} = v/c$, with v our velocity relative to the CMB local rest frame. Although this is small, the precision of FIRAS was such that the difference between the spectrum determined for a patch in the dipole direction and that from the opposite direction could be taken. This should be proportional to the derivative of a blackbody, and indeed it is to a very high degree of accuracy (an rms deviation consistent with the level of detector noise). From the 4-year DMR data, the derived value is 3.353 ± 0.024 mK [85, 86], in good agreement with the FIRAS result. The DMR-derived direction in Galactic coordinates (ℓ, b) is $(264.26 \pm 0.33, 48.22 \pm 0.13)$.

At the 95% confidence limits, the temperatures determined from the monopole spectrum and from the dipole spectrum in [12] are:

$$\text{monopole:} \quad T_{c*} = 2.728 \pm 0.004 \text{ K} \quad (95\% \text{ } CL), \quad (5)$$

$$\text{dipole:} \quad T_{c*} = 2.717 \pm 0.014 \text{ K}, \quad (2.725 \pm 0.020 \text{ K } DMR). \quad (6)$$

Thus the dipole temperature agrees to within the errors with the monopole temperature. The 0.004 K error should be compared with the error bars on the monopole $T(\lambda)$ shown in the inset of fig. 1.

The data was also used to place stringent constraints on distortions to the spectrum. We now turn to the implications of these, but here just quote the values [12]:

$$\text{Compton } y\text{-parameter:} \quad \bar{y} < 1.5 \times 10^{-5} \quad (95\% \text{ } CL), \quad (7)$$

$$\text{chemical potential:} \quad |\mu_\gamma|/T_\gamma < 0.9 \times 10^{-4} \quad (95\% \text{ } CL), \quad (8)$$

$$\text{general distortions:} \quad \frac{\delta E}{E_{cmb}}(500\text{--}5000 \text{ } \mu\text{m}) < 0.00025 \text{ } (1\sigma). \quad (9)$$

For the general distortions, the constraint on the fractional energy release over the waveband from 5000 to 500 μm follows from the FIRAS team result using the monopole spectrum that over this band the maximum 1-sigma intensity deviation from a blackbody was $< 0.012\%$ of the peak brightness. The rms intensity deviation from a blackbody over all channels in this range is even more stringent, 0.005%. (If a sub-mm background mimicked Galactic emission the constraints would not be as severe. See section 3.6.)

3 Spectral distortion theory

3.1 Radiative transport in the expanding universe

The development of spectral distortions or angular anisotropies in the microwave background is described by radiative transfer equations for the photon distribution function, which are coupled to Einstein's equations for the gravitational field and to the hydrodynamic and transport equations for the other types of matter present. The photon distribution function for the total intensity, $f_t(q, \hat{q}, x^i, \tau)$ is a dimensionless general relativistic invariant giving the average photon occupation number as a function of the photon momentum q^I , $I = 1, 2, 3$, with magnitude q and direction vector $\hat{q}$, in the neighborhood of the spatial point x^i , $i = 1, 2, 3$, at time τ . Not only is f_t a general relativistic scalar under the change of the spacetime coordinates (x^i, τ) , it also remains invariant under change of the 3-momentum coordinates q^I . Apart from f_t , there are three other photon distribution functions needed to describe the state of polarization: f_t, f_U, f_V, f_Q , correspond to the four Stokes parameters I, U, V, Q .

Because physical momentum p redshifts as the Universe expands, the comoving momentum, $q \equiv \bar{a}(t)p$ is a better choice than p . The comoving photon energy qc and comoving wavelength λ are therefore related to the physical frequency ν , physical energy $\omega = h\nu$, and wavelength λ_e by

$$qc = \frac{2\pi\hbar c}{\lambda} = \bar{a}(t)\omega = \bar{a}(t)\frac{2\pi\hbar c}{\lambda_e}. \quad (10)$$

Thus, if λ_e is the wavelength at emission at time t , λ is the observed wavelength at t_0 in the absence of frequency shifts beyond that from cosmological expansion. A Planck distribution function is of form $f_{Pl} = [\exp(q/T_{\gamma*}) - 1]^{-1}$, where $T_{\gamma*} \equiv \bar{a}T_\gamma$ is the (comoving) photon temperature. We denote the Planckian with the observed CMB temperature $T_{c*} = 2.728$ K by

$$f_c \equiv (e^x - 1)^{-1}, \quad x \equiv \frac{q}{T_{c*}} = \frac{5273 \mu\text{m}}{\lambda} \frac{2.728 \text{ K}}{T_{c*}} = 1.76 \frac{\nu}{100 \text{ GHz}}. \quad (11)$$

The dimensionless x remains constant as the universe expands. Instead of the distribution function, it is often convenient to work with a generalized (comoving) thermodynamic temperature,

$$T_{t*}(q, \hat{q}, x^i, \tau) \equiv q / \ln(f_t^{-1} + 1). \quad (12)$$

We are interested in the fluctuations in f_t and T_{t*} ,

$$\Delta f_t \equiv f_t - f_c, \quad \Delta_t \equiv \left(\frac{\Delta T}{T} \right)_t \equiv \frac{(T_{t*} - T_{c*})}{T_{c*}}. \quad (13)$$

These will generally store both distortion and anisotropy information and may often be nonlinear. Let us denote the spatial averages of f_t and Δ_t at a given

time by $\bar{f}_t \equiv f_c + \overline{\Delta f_t}$ and $\overline{\Delta f_t}$ and the spatial fluctuation in f_t by $\delta f_t = \Delta f_t - \overline{\Delta f_t}$.

The specific intensity I_ν historically used by astronomers to describe radiative transfer is related to the distribution function by

$$I_\nu(\nu, \hat{q}, x^i, t) = 2\pi \frac{\hbar}{c^2} \nu^3 2 f_t, \quad (14)$$

with the 2 coming from the two photon polarizations. The energy per unit 3-volume radiated into solid angle $d\Omega_{\hat{q}}$ in the frequency interval ν to $\nu + d\nu$ is $I_\nu d\nu d\Omega_{\hat{q}}$. For a generally inhomogeneous spacetime, both the 3-volume and the momentum (hence ν) can be transformed by a coordinate (gauge) transformation, which is why invariant distribution functions are far preferable to work with.

For spectral distortions, and for anisotropies that arise from secondary processes such as Compton cooling of hot gas in clusters and emission from point sources, the following form of the transfer equation is sufficient:

$$\left. \frac{\partial f_t}{\partial \tau} \right|_q + \hat{q} \cdot \nabla f_t = \bar{a} \mathcal{S}[f_t], \quad (15)$$

where $\mathcal{S}$ is the source function describing the difference between the rate at which photons are being added to the momentum volume $d^3\mathbf{q}/(2\pi)^3$, and the rate at which they are being removed. Instead of “cosmic time”, it is more convenient to use conformal time $d\tau = dt/\bar{a}$ and comoving space coordinates x^i in the transfer equation. In terms of $(\Delta T/T)_t$, the transfer equation takes the form

$$\left[\left. \frac{\partial}{\partial \tau} \right|_q + \hat{q} \cdot \nabla \right] \left(\frac{\Delta T}{T} \right)_t \equiv \mathcal{G}(\mathbf{r}, q, \hat{q}, \tau) \equiv \frac{\bar{a} \mathcal{S}[f_t] (1 + \Delta_t)^2}{x(f_c + \Delta f_t)(1 + f_c + \Delta f_t)}. \quad (16)$$

The solution of the transfer equation for a source at position $\mathbf{r}_s$ emitting in a burst at time τ_s , hence with $\mathcal{S} \propto \delta(\tau - \tau_s) \delta(\mathbf{r} - \mathbf{r}_s)$, is the Green function at time τ_0 and position $\mathbf{r}_0$:

$$\left[\left. \frac{\partial}{\partial \tau} \right|_q + \hat{q} \cdot \nabla \right]^{-1} = \vartheta(\tau_0 - \tau_s) \delta^{(3)}(\mathbf{r}_0 - \mathbf{r}_s - \hat{q}(\tau_0 - \tau_s)), \quad (17)$$

where ϑ is the Heaviside unit function, 0 for $\tau < \tau_s$, 1 otherwise. It describes the *free-streaming* of the radiation along the line-of-sight to the source, with q kept constant over the look-back. It can be used to map the radiation pattern from a time just after all emissions, absorptions and scatterings have become negligible (so $\mathcal{S} \approx 0$) to the present. If there is a contribution $-\bar{a}\bar{\Gamma}_a f_t$ of uniform absorbers in $\bar{a}\mathcal{S}$ as well, then the Green function is

$$\begin{aligned} & \left[\left. \frac{\partial}{\partial \tau} \right|_q + \hat{q} \cdot \nabla + \bar{a}\bar{\Gamma}_a \right]^{-1} \\ &= e^{-\zeta_a(\tau_0|\tau_s)} \vartheta(\tau_0 - \tau_s) \delta^{(3)}(\mathbf{r}_0 - \mathbf{r}_s - \hat{q}(\tau_0 - \tau_s)), \end{aligned} \quad (18)$$

where the absorption depth for the process Γ_a is

$$\zeta_a(\tau_0|\tau_s) \equiv \int_{\tau_s}^{\tau_0} \bar{a}(\tau) \bar{\Gamma}_a(\tau) d\tau. \quad (19)$$

With inhomogeneous absorbers, the Green function naturally depends on the absorption depth along the line-of-sight.

In the tight coupling limit valid in the early universe, sources and sinks in $\mathcal{S}$ approximately balance, so $\mathcal{S} \approx 0$, but the solutions then are an equilibrium, with a small perturbation describing diffusion and viscous coupling of the photon fluid to other matter present. As usual with radiative transport, most of the complications arise in the transition from tight coupling to free streaming. If the spatial fluctuation $\delta f = f_t - \bar{f}_t$ is small enough so that the spatial average of $\bar{a}\mathcal{S}[f_t]$ can be replaced by $\bar{a}\mathcal{S}[\bar{f}_t]$ to zeroth order, then $\bar{f}_t$ obeys the zeroth-order (background) transfer equation,

$$\partial \bar{f}_t / \partial \tau \Big|_q = \bar{a} \bar{\mathcal{S}}[\bar{f}_t]. \quad (20)$$

In both the tight coupling and free streaming regimes, any form-invariant function of q is a solution, in particular a Planckian with T_{t*} constant, or a Bose–Einstein distribution $(\exp[q/T_{t*} + \alpha] - 1)^{-1}$ with the chemical potential parameter $\alpha \equiv -\mu_\gamma/T_t$ constant as well. If the distortion and/or anisotropy fluctuation Δf_t is small compared with f_c , then

$$\Delta f_t \approx T_c \frac{\partial f_c}{\partial T_c} \frac{\Delta T}{T} \equiv x f_c (1 + f_c) \frac{\Delta T}{T}, \quad x f_c (1 + f_c) = \frac{x e^x}{(e^x - 1)^2}. \quad (21)$$

We typically use this transformation to go from distribution function to temperature fluctuation, although it is sometimes not a good approximation, e.g., in the Wien region with dust emission sources, since f_c drops so rapidly.

The full treatment of the transport theory with gravitational redshift and lensing effects and polarization effects is developed in Appendix B, and discussed in section 6.5. The transport operator, the left-hand side of eq. (15), is augmented by a term that depends upon the connection coefficients of the spacetime metric, $-q^{-1} \Gamma_{\alpha\beta}^i q^\alpha q^\beta \partial f / \partial q^i$. The Green function describing free-streaming from a source is a delta function along the photon’s geodesic path. The bending of $\hat{q}$ is essentially a lensing effect, a nonlinear correction involving a product of the perturbed metric and Δf_t . The gravitational frequency shift as the photon climbs into and out of local pockets of curvature is very important in linear theory, the Sachs–Wolfe effect [32]. It is legitimate to take the Sachs–Wolfe term to the right-hand side and treat it as a source. If we write the metric as $ds^2 = \bar{a}^2(\eta_{\alpha\beta} + h_{\alpha\beta}) dx^\alpha dx^\beta$, with $\eta_{\alpha\beta} = \text{diag}(-1, 1, 1, 1)$,¹ where $h_{\alpha\beta}$ is the metric fluctuation, then the effective $\mathcal{G}$ in linear perturbation theory (from eq. (289)) is

$$\mathcal{G}_{tSW} = -\frac{1}{2} \dot{h}_{ij} \hat{q}^i \hat{q}^j + \frac{1}{2} \hat{q}^i \partial_i h_{00} + \hat{q}^i \hat{q}^j \partial_i h_{0j}. \quad (22)$$

¹The MTW [196] sign conventions and the summation convention are used. Mean curvature is ignored here but is discussed in Appendix B.

It is usual in perturbation theory to simplify this by adopting a coordinate system in which $h_{i0} = 0$ – called a time-orthogonal gauge choice. Also, by means of a change of the momentum variable in f_t , one can modify this form. For example, $\Delta_t - \frac{1}{2}h_{00}$ has $\mathcal{G}_{tSW} = -\frac{1}{2}\dot{h}_{ij}\hat{q}^i\hat{q}^j - \frac{1}{2}\dot{h}_{00}$.

Linear perturbations in the expanding universe can be separated into scalar, vector and tensor modes, which are mutually independent. For a flat universe, we can Fourier transform the transport equation. If $\mathbf{k}$ is the comoving wavevector and we choose the 3-axis to be $\hat{k}$, then *tensor modes* involve the two (transverse traceless) gravitational wave polarization modes, $h_{(T+)} = (h_{11} - h_{22})/2$ and $h_{(T\times)} = h_{12}$, *vector modes* describing vorticity² involve h_{13}, h_{23} (and h_{01}, h_{02}), and *scalar modes* involve

$$\nu \equiv -\frac{1}{2}h_{00}, \quad \varphi \equiv \frac{h - h_{33}}{4}, \quad \text{where } h \equiv \delta^{ij}h_{ij}, \quad (23)$$

$$\psi \equiv -\frac{h - 3h_{33}}{4k^2}, \quad \Psi_n \equiv -i\frac{h_{03}}{\bar{a}k}, \quad \Psi_\sigma \equiv \Psi_n + \bar{a}\dot{\psi}. \quad (24)$$

The tensor modes are invariant under coordinate changes, whereas the scalar mode potentials defined by eq. (24) do change, i.e., are gauge dependent. For scalar perturbations, the Einstein equations and the various transport equations only involve ν , φ and Ψ_σ , and the perturbations to the various matter densities and velocity potentials, as well as to the distribution functions; further, ν , φ and Ψ_σ only depend upon the choice of time surfaces upon which they are measured, not on changes of spatial coordinates on the hypersurfaces. The time can be chosen to make some linear combination of the three vanish. The two standard choices that have been used in the computation of radiative transport in linear perturbation theory are the *synchronous gauge*, for which $\nu = 0$ (and $\Psi_n = 0$), and the *longitudinal gauge*, for which $\psi = 0$ (and $\Psi_n = 0$).¹

In terms of these metric variables,

$$\begin{aligned} \mathcal{G}_{tSW}^{(S)} &= -i\hat{q} \cdot \mathbf{k}\nu - \dot{\varphi} - (\hat{q} \cdot \mathbf{k})^2 \bar{a}^{-1}\Psi_\sigma, \\ \mathcal{G}_{tSW}^{(T\{+, \times\})} &= -(1 - (\hat{q} \cdot \hat{k})^2) \frac{1}{2}\dot{h}_{(T\{+, \times\})}. \end{aligned} \quad (25)$$

The source function for Compton scattering when energy transfers are important is described in the next section, and the source function $\mathcal{G}_{tC}$ in the low energy Thomson scattering limit including polarization and angular anisotropy effects in the scattering is derived in detail in Appendix C. The dominant terms

²The vector parts of the vector parts, $h_{0i}^{(V)}$ and w_i of $h_{ij}^{(V)} = k_i w_j + k_j w_i$, are curls of vectors.

¹Many different notations are used for the perturbation variables $\{\nu, \varphi, \psi, \Psi_n, \Psi_\sigma\}$; e.g., Bardeen (1980): $\{A, H_L + \frac{1}{3}H_T, -k^{-2}H_T, k^{-1}\bar{a}B, k^{-1}\bar{a}(B - k^{-1}\dot{H}_T)\}$ [171, 173]; Bardeen (1988): $\{\alpha, \varphi, -\gamma, -\bar{a}\beta, -\chi\}$ [178]; Mukhanov et al. $\{\phi, -\psi, -E, \bar{a}B, \bar{a}(B - \dot{E})\}$ [176]. In [2, 4, 88, 134, 195, 215, 216], we used a $(+, -, -, -)$ metric signature, hence h, h_{33} are of opposite signs to those given here. I use ψ because it is basically the displacement potential familiar from use in the Zeldovich approximation, Ψ_n and Ψ_σ because they are velocity potentials for the 4-velocity and shear of observers following the flow of time.

for scalar (S) and tensor (T) modes do not depend upon these complications:

$$\begin{aligned}\mathcal{G}_{tC}^{(S)} &= -n_e \sigma_T \bar{a} (\Delta_t^{(S)} - \Delta_{t0}^{(S)} - \hat{q} \cdot \mathbf{v}_B^{(S)}) + (\text{anisotropy, polarization}), \\ \mathcal{G}_{tC}^{(T\{+, \times\})} &= -n_e \sigma_T \bar{a} \Delta_t^{T\{+, \times\}} + (\text{anisotropy, polarization}),\end{aligned}\quad (26)$$

where n_e is the electron density, σ_T is the Thomson cross section and $\Delta_{t0}^{(S)}$ is the angle-averaged temperature fluctuation.

3.2 Source functions for spectral distortions

Provided the temperature of the universe is well below $m_e c^2$ where e^+e^- pairs recombine, only a small number of processes have to be included to adequately describe the photon transport. In the following expressions, ρ_γ is the photon energy density, n_B is the baryon number density, ϵ_B is the energy per baryon in gas, $Y_e = n_e/n_B$ is the electron fraction per baryon, T_e and T_γ are the electron and photon temperatures (in energy units), $x_e = \omega/T_e$, $\sigma_T = (8\pi/3)\alpha^2/m_e^2$ is the Thomson cross section, m_e and m_N are the electron and nucleon masses, and α is the fine structure constant. The energy rates are those appropriate to near-equilibrium transfer from photons to plasma. All source functions and the photon energy ω are given in the reference frame in which the electrons are at rest (the comoving-baryon gauge).

3.2.1 Compton scattering and the Kompaneets source term

For the nonrelativistic electrons appropriate to the period after pair recombination, Compton scattering is primarily Thomson scattering, a conservative scattering process in which the outgoing photon energy ω' equals that of the incoming one ω , so momentum but not energy is transferred. The associated source function can describe the development of anisotropy, but will give rise to no spectral distortion. For this source function to vanish, it is necessary that the radiation field be isotropic in the comoving frame of the electrons.

The general structure of the source function for $\gamma e \rightarrow \gamma e$ scattering is

$$\mathcal{S}[f](\mathbf{q}) = \sum_{\mathbf{q}'} R(\mathbf{q}' \rightarrow \mathbf{q}) f(\mathbf{q}') (1 + f(\mathbf{q})) - R(\mathbf{q} \rightarrow \mathbf{q}') f(\mathbf{q}) (1 + f(\mathbf{q}')).$$

The first term describes stimulated emission, of photons in momentum state $\mathbf{q}$, the second describes stimulated absorption. Here R is the scattering kernel, which is related to the Klein–Nishina cross section averaged over the thermal electron distribution, $n_e \langle d\sigma_{KN}/[(2\pi)^3 d^3q] \rangle$. If the electrons are in thermal equilibrium at temperature T_e , R obeys the detailed balance relation

$$R(\mathbf{q} \rightarrow \mathbf{q}') = R(\mathbf{q}' \rightarrow \mathbf{q}) e^{(\omega - \omega')/T_e}.$$

The source function for scattering vanishes if $f^{-1} + 1$ is proportional to e^{ω/T_e} , that is if the distribution function is a Bose–Einstein one,

$$f_{BE} = [\exp(\omega/T_e + \alpha) - 1]^{-1}, \quad \alpha \equiv -\mu_\gamma/T_e. \quad (27)$$

The photon chemical potential μ_γ enters because photon number is a conserved quantity in Compton scattering.

For homogeneous transfer, $R(\mathbf{q} \rightarrow \mathbf{q}') = R_0(\omega \rightarrow \omega')$ is a function only of the energies in and out. In the Thomson (very heavy electron) limit, $R_0 \propto \delta(\omega - \omega')$, hence $\mathcal{S}[f] \rightarrow 0$: inhomogeneity is needed to have nonzero sources for Thomson scattering. In the next order in m_e^{-1} , small energy transfers $\Delta\omega = \omega - \omega'$ do occur. Let us introduce a redistribution probability $\phi(\omega \rightarrow \omega')$, defined by

$$\phi(\omega \rightarrow \omega') \equiv \frac{1}{n_e \sigma_T} R_0(\omega \rightarrow \omega') \frac{(\omega')^3}{\pi^2}, \quad \text{with} \quad \int \frac{d\omega'}{\omega'} \phi(\omega \rightarrow \omega') = 1.$$

It is sharply peaked, concentrated near $\Delta\omega \approx 0$, with deviations of order m_e^{-1} , as described by moments taken with respect to ϕ :

$$\left\langle \frac{\Delta\omega}{\omega} \right\rangle_\phi = 4 \frac{T_e}{m_e} - \frac{\omega}{m_e}, \quad \left\langle \left(\frac{\Delta\omega}{\omega} \right)^2 \right\rangle_\phi = 2 \frac{T_e}{m_e}, \quad (28)$$

describing both a net upward drift in the scatters if the photon energy is smaller than $4T_e/m_e$ (i.e., the electrons on average Compton-cool) and a random walk of the photon energy about the net drift.

To derive the Kompaneets form [33] for $\mathcal{S}$, one relies on the peaked nature of ϕ to “punch out” the distribution function at ω , using the Taylor expansion in $\Delta\omega$ of $f(\omega')$ and of the detailed balance relation:

$$\mathcal{S}_K[f] = \frac{T_e}{m_e} n_e \sigma_T c \frac{1}{T_e \omega^2} \frac{\partial}{\partial \omega} \left[\omega^4 \left(T_e \frac{\partial f}{\partial \omega} + f(1+f) \right) \right]. \quad (29)$$

The following properties can be readily verified: (1) no photons are created or destroyed ($\frac{1}{\pi^2} \int \omega^2 d\omega \mathcal{S}_K[f] = 0$); (2) $\mathcal{S}_K[f]$ vanishes only if f is of the Bose–Einstein form (since $[f(1+f)]^{-1} df = -d(\omega/T_e)$); (3) The rate per unit volume at which photons are heated, $d\rho_\gamma/dt$, is the negative of the Compton cooling rate per unit volume of the electrons, $n_B(d\epsilon_B/dt)$:

$$\left[-n_B \frac{d\epsilon_B}{dt} \right]_K = \left[\frac{d\rho_\gamma}{dt} \right]_K = \int \frac{\omega^2 d\omega}{\pi^2} \omega \mathcal{S}_K[f_t] \approx \frac{4n_e \sigma_T}{m_e} \rho_\gamma (T_e - T_\gamma), \quad (30)$$

where the last term assumes $f_t \approx f_{Pl}$, the Planck function. Thus the electron temperature is driven by Compton cooling towards the photon temperature.

If Compton scattering dominates energy redistribution, but it is not so strong as to shape a Bose–Einstein “kinetic equilibrium” distribution, a y -distortion spectrum is the solution to the Kompaneets equation. For small distortions of the distribution function, $\Delta f_t \ll f_c$, we have

$$\mathcal{G}_K = -2\bar{a} n_e \sigma_T \frac{(T_e - T_c)}{m_e} \psi_K(x), \quad \psi_K(x) \equiv 2 - \frac{x(e^x + 1)}{2(e^x - 1)}. \quad (31)$$

The solution of the radiative transfer equation can therefore be written in terms of the Compton y -parameter (along a line-of-sight from the current (conformal) time τ_0 back to time τ):

$$\left(\frac{\Delta T}{T}\right)_K = -2y\psi_K(x), \quad y \equiv \int_{\tau}^{\tau_0} \bar{a} d\tau n_e \sigma_T \frac{(T_e - T_c)}{m_e}. \quad (32)$$

In terms of the Thomson scattering (optical) depth,

$$\zeta_C(\tau_0|\tau) \equiv \int_{\tau}^{\tau_0} n_B Y_e \sigma_T \bar{a} c d\tau, \quad (33)$$

this is $\zeta_C(\tau_0|\tau)\langle T_e - T_c \rangle/m_e$. For a fully ionized medium with free electron abundance Y_e per baryon, $\zeta_C \approx 0.0465 Y_e \Omega_{B, gas} h \Omega_{nr}^{-1/2} (1+z)^{3/2}$ for $z \gg 1$. There is another useful solution to the Kompaneets equation given by Zeldovich and Sunyaev [34] which is valid for larger y than the perturbation expansion allows, but with the restriction that the electron temperature is well in excess of T_γ , which eq. (32) does not require:

$$f(x) \approx \frac{1}{\sqrt{4\pi y}} \int_0^\infty \frac{d\xi}{\xi} \left[\frac{1}{e^\xi - 1} - f_c(x) \right] \exp \left[-\frac{(3y - \ln x + \ln \xi)^2}{4y} \right].$$

In either case, the asymptotic Rayleigh–Jeans temperature is related to the unperturbed photon temperature by $T_{RJ} = e^{-2y} T_c$ and the total energy is $\rho_\gamma = e^{4y} \rho_{cmb}$. The FIRAS constraint eq. (7) implies that energy injected into the medium (below a redshift $z_y \sim 10^5$ defined below) which Compton-cooled can be at most

$$\frac{\delta E_{Compton cool}}{E_{cmb}} = 4y < 6.0 \times 10^{-5} \quad (95\% CL). \quad (34)$$

The spectral signature of this y -distortion is uniquely characteristic: $-2y$ on the Rayleigh Jeans side, xy on the far Wein side, passing through zero at $x = 3.83$, as shown in fig. 2 for $y = 0.001$. A Bose–Einstein curve with $\alpha = 0.0057$ is also shown. Both correspond to 0.4% energy injections relative to the CMB. (For a few years in the late eighties there was a flurry of activity as theoreticians tried to come to grips with a $y = 0.016$ distortion reported by Matsumoto et al. [31].)

3.2.2 Bremsstrahlung

The source function for free–free emission and absorption from ionized hydrogen and helium is

$$\begin{aligned} \mathcal{S}_{ff}[f_t] &= -\Gamma'_B(\Delta f_t - (f_{eq} - f_c)), \quad f_{eq} = (e^{x_e} - 1)^{-1}, \\ \Gamma'_B &= \Gamma_B(1 - e^{-x_e}), \quad x_e \equiv \omega/T_e, \\ \Gamma_B &= \frac{(2\pi)^{3/2}}{\sqrt{3}} \alpha n_B^2 \sigma_T \left(\frac{m_e}{T_e}\right)^{1/2} Y_e(Y_p + Y_{\text{HeII}} + 4Y_{\text{HeIII}}) \frac{g(x_e)}{T_e^3 x^3}, \end{aligned} \quad (35)$$

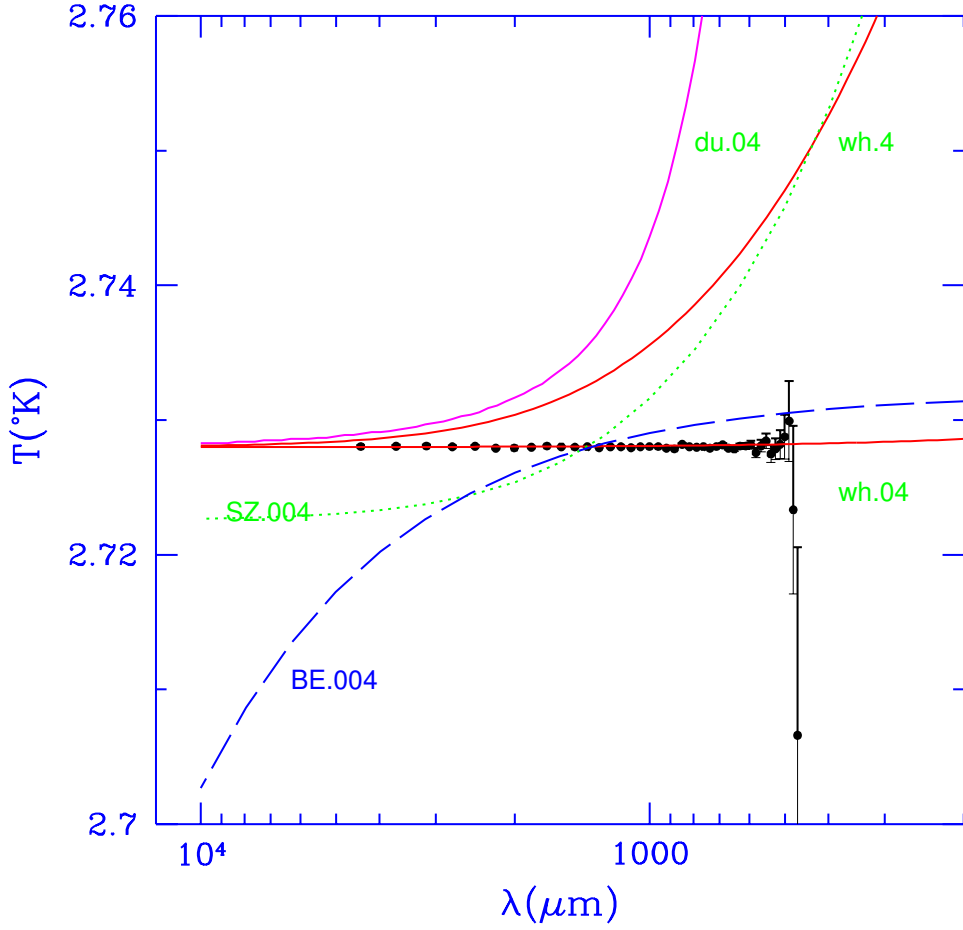


Figure 2: Sample types of spectral distortions are compared with the FIRAS data (Fixsen et al. 1996). SZ.004 is a y -distortion with $y = 0.001$, BE.004 is a Bose–Einstein distortion with $\alpha = 0.0057$, du.04 is a model with ordinary dust grains with abundance 10^{-6} reprocessing injected energy which was taken to be 4% of that in the CMB between redshifts 50 and 25. Two models mimicking the effect of an optically thin abundance of needle-like grains (whiskers) acting over the same redshift, with 40% and 4% of the CMB energy injected, are also shown.

$$\text{Gaunt factor: } g(x_e) \approx 1, \quad x_e > 1; \quad \approx \frac{\sqrt{3}}{\pi} \ln \left[\frac{2.25}{x_e} \right], \quad x_e < 1.$$

Since the rate at which photons are emitted into the energy interval ω to $\omega + d\omega$ by free-free processes, $dn_\gamma/dt \sim (\omega^2 d\omega/\pi^2) \Gamma'_B(f_{eq} - f_c)$, $\rightarrow d\omega/\omega$ at low ω , bremsstrahlung is very efficient at filling in an equilibrium Planck distribution (with zero chemical potential) at low energies. Although it is also not inefficient at high energies, Compton scattering dominates.

It is convenient to characterize the strength of bremsstrahlung by a parameter y_{ff} analogous to the Compton y -parameter:

$$y_{ff} \equiv \int dt [\Gamma_B x^3] (1 - T_c/T_e). \quad (36)$$

The approximate constancy of $\Gamma_B(x)x^3$ has been exploited in this formula. The current 2σ limit on this parameter is $y_{ff} < 1.9 \times 10^{-5}$ [19]. The total energy input relative to the CMB for $T_e \gg T_c$ and over the long wavelength range up to say $\lambda = 2$ cm is

$$\begin{aligned} \frac{\delta E_{brems}}{E_{cmb}}(\text{tot}) &= 15 y_{ff} \left\langle \frac{T_{e*}}{T_{c*}} \right\rangle, \\ \frac{\delta E_{brems}}{E_{cmb}}(\lambda > 2 \text{ cm}) &\approx 15 y_{ff} x \lesssim 7 \times 10^{-5}, \quad \text{with } x = 0.26. \end{aligned} \quad (37)$$

This can be used to constrain reionized models, with the caveat that $y_{ff} \propto n_e^2 T_e^{-1/2}$ is dominated by dense regions and so is very sensitive to clumping in the medium.

The source for the temperature fluctuations can be written

$$\mathcal{G}_{brems} = \bar{a} \frac{dy_{ff}}{dt} \psi_{ff}(x), \quad \psi_{ff}(x) \equiv \frac{g(x_e)}{x^3} \frac{(e^{-x_e} - e^{-x})}{1 - T_c/T_e} \frac{(e^x - 1)}{x}. \quad (38)$$

The signature of bremsstrahlung in the thermodynamic temperature is

$$\left(\frac{\Delta T}{T} \right)_{brems} = y_{ff} \langle \psi_{ff} \rangle, \quad \psi_{ff} \approx x^{-2} \ln(2.35/x_e) \quad \text{for } x \ll 1. \quad (39)$$

Thus for low frequencies, the thermodynamic temperature follows a ν^{-2} law. (For large x , but x_e small ($T_e \gg T_c$), $\psi_{ff} \approx x^{-4} e^x \ln(2.35/x_e)$, so the slope eventually turns positive.)

3.2.3 Double Compton scattering

In Compton scattering, the electron can shake off a soft photon, $\gamma e \rightarrow \gamma e + \gamma$, basically a bremsstrahlung process with a form very similar to that for free-free emission. In particular, there is a logarithmic divergence in the number of low energy photons emitted. The source functions for this Double Compton

scattering were derived in [2] using the cross sections given by Gould [35]. These revealed a different dependence on photon energy than bremsstrahlung at high energies:

$$\begin{aligned}
\mathcal{S}_{DC}[f] &= -\Gamma'_{DC}(f - f_{eq}), \quad \Gamma'_{DC} = \Gamma_{DC}(1 - e^{-x_e}), \\
\Gamma_{DC} &= \frac{16\pi^3}{45} \alpha n_e \sigma_T \left(\frac{T_e}{m_e} \right)^2 \frac{g_{DC}(x_e)}{x_e^3}, \\
g_{DC}(x) &= \frac{15}{4\pi^4} \int_{2x}^{\infty} f(y)(1 + f(y - x))y^4 dy \left[\frac{x/y F(x/y)}{2} \right], \\
\frac{wF(w)}{2} &= \frac{1}{2}(1 - w) \left[1 + (1 - w)^2 + \frac{w^2(1 + w^2)}{(1 - w)^2} + w^4 + w^2(1 - w)^2 \right].
\end{aligned} \tag{40}$$

For small w , $[wF(w)/2] \rightarrow 1$. Burigana et al. [36] fit $g_{DC}(x_e)$ by $\exp(-x_e/2)$, valid for $x_e < 1$, an improvement (in the cosmologically-relevant regime) over a more complicated approximation I gave in [2]. The net effect is that the Double Compton process is usually subdominant to free-free emission for cosmologically interesting parameters unless $\Omega_B h^2$ is quite low.

3.2.4 Rayleigh scattering

Photon scattering from neutral hydrogen and helium has an identical source function to that for Thomson scattering, but with a frequency dependence given by the fourth power law. For hydrogen, the ratio of the rates $\sim (n_{\text{HI}}/Y_e)(\omega/\omega_\alpha)^4$, where $\omega_\alpha = 10.2$ eV is the Lyman α transition energy. For typical photon energies at recombination ($z \approx 1000$) this is small, $2 \times 10^{-5} n_{\text{HI}}/n_e$, and it declines precipitously as the radiation temperature drops, so is never significant. Although helium is neutral above $z = 1000$, Thomson scattering dominates even in the very tightly coupled regime.

3.2.5 Line radiation

Lines formed during the recombination of helium when the temperature is a few eV and of hydrogen at $\sim 1/4$ eV are either too weak to be easily observable, or are buried in the background associated with interstellar dust emission [37]. Such processes play a very important role in the recombination process itself of course, and this is discussed later.

3.2.6 Synchrotron

Since synchrotron emission requires both magnetic fields and energetic electrons to have been generated, it seems unlikely that a synchrotron background from high redshift will generate a measurable distortion. However, at radio frequencies the anisotropy from extragalactic radio sources and Galactic emission is significant and will contaminate anisotropies from other signals. Fortunately

the spectral signature is sufficiently different from primary anisotropies that with enough frequency coverage this component could be isolated. The synchrotron intensity is parameterized by a power law index p_s : $I_\nu \sim \nu^{-p_s}$. For extragalactic radio sources, $p_s \sim 0.5$ is the conventional value, but from deep VLA counts there is evidence for a flatter population of sources [38]; how flat and how abundant in the frequency range of interest for anisotropy observations is currently not well known. For Galactic sources, one has $p_s \sim 0.3$ – 0.7 at low frequencies; e.g., using maps at 408 MHz [39], 1.4 GHz [40], and 2 GHz [22] gives $p_s \sim 0.6$ for moderate Galactic latitudes. One expects the index to steepen at higher frequency, and there are indications that around 15 GHz, $p_s \sim 1$ may be more appropriate [103]. In any case, (small) spatial variations in p_s are both expected and observed. The thermodynamic temperature is

$$\left(\frac{\Delta T}{T}\right)_{synch} \propto \psi_{synch}, \quad \psi_{synch} \equiv \frac{1}{x^{3+p_s}} \frac{(e^x - 1)^2}{xe^x}, \quad (41)$$

going as $\nu^{-(2+p_s)}$ for low frequencies, an even steeper law than the bremsstrahlung ν^{-2} .

3.2.7 Dust grains

Radiation from heated primeval dust at high redshift would naturally reside at submillimeter wavelengths, with the energy density peaking at several hundred microns (e.g., BCH2 [42] and references therein); e.g., with 30 K dust typical of starburst galaxies, the dust temperature would only be a factor of two above the CMB at redshift 5. Of course Galactic sources abound to obscure cosmological signals: dust at 20 K and possibly cold dust (~ 5 K) at high Galactic latitude [43]. The dust source function for emission/absorption is

$$\mathcal{S}_{dust} = \Gamma_a (f_{eq} - f_c) - \Gamma_a \Delta f, \quad (42)$$

$$f_{eq} = (e^{x_d} - 1)^{-1}, \quad x_d \equiv \frac{\omega}{T_d}, \quad \Delta f \equiv f - f_c, \quad f_c \equiv (e^x - 1)^{-1}, \quad (43)$$

$$\Gamma_a = \frac{\rho_d}{\rho_{id}} \left(\frac{\lambda_e}{2\pi c} \right)^{-1} A_d(\lambda_e). \quad (44)$$

There is also a dust scattering source term. Here f_{eq} is the equilibrium distribution function for dust in thermal equilibrium at a (single) temperature T_d (obtained by balancing the energy absorbed from the local radiation field to the energy emitted – usually). ρ_d is the mass density of grains, ρ_{id} is their internal density ($\approx 3 \text{ g cm}^{-3}$), and the parameterizing function A_d depends upon the photon energy $\omega = 2\pi/\lambda_e$ and grain properties. In BCH2, we adopted a three parameter form (α_d, A_{d100}, r_d), only two of which were needed at the infrared emission wavelengths of relevance for CMB observations, an amplitude A_{d100} at 100 μm and a slope:

$$A_d(\lambda_e, z) = A_{d100} (100 \mu\text{m}/\lambda_e)^{\alpha_d - 1}, \quad (45)$$

where $A_{d100} \equiv A_d(100 \mu\text{m})$. For almost all types of grains and plausible conditions, the absorption part $-\Gamma_a \Delta f$ of eq. (42) is of relevance only in the UV and visible. However, the large contribution of the unperturbed cosmic background itself must be included as an absorption component $-\Gamma_a f_c$ that partly counteracts the emission component $\Gamma_a f_{eq}$ at long wavelengths. The source function for the thermodynamic temperature is therefore

$$\mathcal{G}_{dust} = K \bar{a}^{1-\alpha_d} \frac{\rho_d}{\rho_{id}} A_{d100} \psi_{dust}, \quad (46)$$

$$\psi_{dust}(x) \equiv x^{\alpha_d} \frac{(e^{-x_d} - e^{-x})}{(1 - e^{-x_d})} \frac{(e^x - 1)}{x}, \quad (47)$$

where K is a constant, hence the shape of the thermodynamic temperature spectral form is

$$\left(\frac{\Delta T}{T}\right)_{dust} \propto \psi_{dust}, \quad \approx x^{\alpha_d} \frac{(T_d - T_c)}{T_c} \quad \text{for } x \ll 1. \quad (48)$$

For $x_d \ll 1, x > 1$, $\psi_{dust} \approx x^{\alpha_d-2} e^x$.

Of course the dust population will be a mix of grains of differing composition, size, shape and temperature. Unravelling the components making up “conventional” Galactic dust remains a hotly debated subject, nicely reviewed in [50]. An example of a recently proposed mix to explain all of the data from the UV to the sub-mm [44] is: most of the mass in “usual” $\sim 0.01\text{--}0.1 \mu\text{m}$ silicate grains, with an added carbon-dominated coating and separate amorphous carbon and graphite grains; $\sim 6\%$ of the mass in very small ($\sim 10 \text{ \AA}$) carbon-dominated grains; and $\sim 6\%$ in $\sim 10\text{--}100 \text{ \AA}$ polycyclic aromatic hydrocarbon molecules (PAHs), of which “bucky-balls” are an example. Dust which is porous and fractal [52], consisting of large random aggregates of small grains, and grains which are triaxial, possibly with extreme elongations (needles or whiskers), are also proposed constituents of the Galactic mix. For spherical amorphous carbon, graphite and silicate grains, $A_{d100} \sim 0.3$. The slope α_d depends upon the mix of grains and their shapes. On broad theoretical grounds, one expects $\alpha_d \approx 2$ for large λ . If the FIRAS sub-mm to mm emission is fit to a single temperature dust model, $\alpha_d = 1.65$ (and $T_d = 23 \text{ K}$) are obtained [43]; similar slopes are inferred from other CMB experiments, while earlier data over the $100\text{--}1300 \mu\text{m}$ range gave $\alpha_d \sim 1\text{--}2$, with the steeper slopes inferred for star forming regions and molecular clouds, and the shallower ones inferred for the Galactic center, dust forming stars and compact HII regions. Forcing α_d to be exactly 2, but allowing two dust temperatures, gives a better fit to FIRAS [43, 47], with the 20 K dust augmented by a 5 K component. A cold component persists at high Galactic latitude, which could be Galactic [48] or due to redshifted extragalactic sources [49]. The dust temperatures associated with most of the Galactic IR luminosity, from diffuse HI clouds, and also from molecular clouds, are around 20 K, with warmer 30 K dust in lower density HII regions, which do not contribute much luminosity. In starburst galaxies, 30 K dust dominates.

The dust absorption law at short wavelengths is also of concern because it determines how efficiently stellar and other radiation is absorbed to be re-emitted in the infrared. If the absorption cross section was geometrical, πr_d^2 , where r_d is the grain radius, then $A_d = 0.75\lambda_e/(2\pi r_d)$ and Γ_a is approximately constant – a rough guide, but for realistic materials A_d is broadly frequency independent at intermediate wavelengths with resonance features superposed. Galactic dust is observed to have $A_d \approx 0.8$ at $\lambda = 0.1 \mu\text{m}$, rising from the visual to the UV (until $0.1 \mu\text{m}$), probably due to very small grains, with a strong resonant feature at $\lambda \approx 0.22 \mu\text{m}$ associated with graphite. There is a strong broad resonance at $\lambda \approx 10 \mu\text{m}$, attributed to silicates in an amorphous or disordered state, another silicate feature at $19 \mu\text{m}$, and a resonance feature around $3 \mu\text{m}$, attributed to carbonaceous grains or coating on the silicates [44]. Dust grains in molecular clouds exhibit more resonances. The size distribution of grains can be derived from Galactic extinction data only with specific assumptions about the nature of the dust; e.g., [51] apply the $0.1\text{--}5 \mu\text{m}$ extinction data to spherical grain models and obtain $dn_d/dr_d \sim r_d^{-3.1} \exp[-r_d/0.14 \mu\text{m}]$ for silicates and $\sim r_d^{-3.5} \exp[-r_d/0.28 \mu\text{m}]$ for graphite and/or amorphous carbon – not far off the oft-used MRN law $dn_d/dr_d \sim r_d^{-3.5}$ [46].

The usual way to make predictions about dust emission in the extragalactic, protogalactic and pregalactic realms is to assume the dust is similar to Galactic dust – as it is currently envisaged. For emission redshifts below about 10, resonances would not appear in the $5000\text{--}500 \mu\text{m}$ FIRAS band but pregalactic dust emission at $z \sim 50\text{--}100$ would bring broadened resonance features into the FIRAS band to aid in emission epoch determination – if distortions had been found. Complicating the constraints that one can impose from the FIRAS limits on high redshift emission is the freedom one has with dust models. In particular, fractal grains would have large effective sizes which would lower the effective α_d in the far infrared and thereby increase $A_d(\lambda)$; it could easily be by more than an order of magnitude over the conventional dust value. A much more radical absorption rate would result if grains were long conducting needles, basically little antennae for which A_d could be thousands of times bigger than the conventional value at long wavelengths. The cosmological importance of these whiskers was suggested by Layzer and Hively [53], and the subject has been developed by Rana [56], Hoyle [54], Wright [57], and is the mainstay of the attempts by Hoyle, Burbidge and Narlikar [55] to create a viable neo-steady-state model.

In the limit in which the photon wavelength is large compared with the scale r_d of the grain (volume $\equiv (4\pi/3)r_d^3$), A_d can be written in terms of the trace of the (electric) polarizability tensor, α_{ij}^e , of the grain (which can be treated as a coherent unit in this limit): $A_d = \frac{4\pi}{3} \sum \Im[\alpha_{jj}^e]$, where $\Im$ denotes imaginary part. For example, for homogeneous ellipsoidal grains with a complex isotropic dielectric tensor $\epsilon(\omega)\delta_{ij}$,

$$A_d = \Im[\mathcal{A}_d], \quad \mathcal{A}_d = \frac{1}{3} \sum_{j=1}^3 \frac{\epsilon(\omega) - 1}{1 + L_j(\epsilon(\omega) - 1)} \quad (\text{for } \omega r_d \ll 1). \quad (49)$$

The sum is over the axes of the grains, and the L_j are “depolarizing factors”, functions of the axis ratios for ellipsoidal grains. The L_j sum to unity. Setting all $L_j = 1/3$ gives the classical Mie expression for spherical grains; it is used together with laboratory data on $\epsilon(\omega)$ to estimate A_d .

For conductors at IR wavelengths, the dielectric function is of form $\epsilon(\omega) \approx \epsilon_d + i2\sigma_c\lambda$, where ϵ_d is the static (real) dielectric constant and σ_c is the conductivity. For iron grains, $(2\sigma_c)^{-1} \approx 0.015 \mu\text{m}$, and for carbon (graphite) grains, it is $\approx 0.6 \mu\text{m}$. For needle-like grains L_j is $1/2$ in the transverse directions and nearly vanishes along the needle. For example, assuming a prolate ellipsoid with semi-minor axis b_d much smaller than the semi-major axis a_d , the depolarizing factor along the needle is $L_{\parallel} \approx (b_d/a_d)^2 \ln(a_d/b_d)$, hence eq. (49) gives

$$A_d \approx \frac{1}{3} \left[\frac{2\sigma_c\lambda}{1 + (L_{\parallel}2\sigma_c\lambda)^2} + 8 \frac{2\sigma_c\lambda}{(1 + \epsilon_d)^2 + (2\sigma_c\lambda)^2} \right], \quad (50)$$

$\propto \lambda$ until λ exceeds $L_{\parallel}^{-1}(2\sigma_c)^{-1}$, which could be in the centimeter range if a_d/b_d could be above a thousand. Thus, $\alpha_d \approx 0$, perhaps rising to 2 only beyond the FIRAS range. Formation scenarios that could lead to such elongated grains have been proposed but there is no evidence that they are produced in nature. Wright et al. [58] argue that the FIRAS data implies such a good blackbody that a large optical depth to needles is needed in a whisker-impregnated steady state model, and this would mask high redshift objects (we have also seen the SZ effect in a cluster at $z = 0.55$, section 5.3.4). Although it seems improbable that the entire CMB could be just dust-emitted radiation, a small fraction of grains in the whisker form could hide more modest energy injections. Figure 2 illustrates what happens when one flattens the dust index to $\alpha_d = 0$ and uses a whisker-motivated value for A_{d100} (2222 was chosen) on a model with energy injected in a burst between redshifts 50 and 25 and a dust abundance $\Omega_d = 10^{-7}$, arranged to give a depth just below unity. Whereas a model with injected energy 4% of the CMB is strongly ruled out for normal dust with $\alpha_d = 1.5$ and $A_{d100} = 0.3$, the redshifted whisker temperature remains so near the CMB temperature that the distortion is small (but the 40% injected energy model is ruled out). Of course, a mix of grain types with only a small percentage of whiskers will give larger distortions [57]; and even the whisker-only model will be enhanced by nonequilibrium effects: these antennae are such efficient radiators that a balance between absorbed and emitted energy leading to a steady dust temperature will not happen, but rather there will be strong temperature fluctuations as absorbed energy is immediately radiated away, a phenomenon expected in very small grains as well [59]. With improved exploration of the sub-mm and mm sky, a necessary part of the next generation of CMB anisotropy experiments, we can expect that the exotic dust loophole will be more strongly constrained.

3.3 The cosmic photosphere and Bose–Einstein distortions

To determine what happens to injected energy at early epochs, we must solve $(\partial f / \partial t)_q = \mathcal{S}_{bremss} + \mathcal{S}_{DC} + \mathcal{S}_K$. The other processes mentioned above are not important. To be accurate, numerical solutions are required; Burigana, Danese and DeZotti [36] give the most detailed to date. Three redshifts characterize the solutions: Energy injection prior to

$$z_{Pl} \approx 10^{6.9} \left(\frac{\Omega_B h^2}{0.01} \right)^{-0.39} \quad (51)$$

is redistributed into a Planckian form, hence z_{Pl} defines the redshift of the cosmic photosphere. Between z_{Pl} and

$$z_{BE} \approx 10^{5.6} \left(\frac{\Omega_B h^2}{0.01} \right)^{-1/2} \quad (52)$$

injected energy is redistributed into a Bose–Einstein shape characterized by a chemical potential. Below

$$z_y \approx 10^5 \left(\frac{\Omega_B h^2}{0.01} \right)^{-1/2} \quad (53)$$

the y -distortion formula holds. There is an intermediate range between z_{BE} and z_y when neither the Bose–Einstein nor y -distortion forms are accurate.

To understand the magnitudes of these redshifts, an analytic treatment based on Zeldovich and Sunyaev [34] is quite adequate. Assume the distribution function has the form $f = [\exp(x + \alpha(x, t)) - 1]^{-1}$ and linearize the transport equation in α . (α is more transparent to work with than the thermodynamic temperature fluctuation $\Delta_t = -\alpha/(x + \alpha)$.) In the tight coupling regime, $\mathcal{S}_K + \mathcal{S}_{bremss} + \mathcal{S}_{DC}$ approximately vanishes; this condition is satisfied for small $x_e = \omega/T_e$ if $\alpha = \alpha_0(t) \exp(-x_0/x_e)$. Thus for low frequencies, $x < x_0$, bremsstrahlung and the Double Compton process dump photons in fast enough to yield a Planck form, but for $x > x_0$ the Bose–Einstein form prevails. Here $x_0 = (4x^3(\Gamma_{bremss} + \Gamma_{DC})/\Gamma_K)^{1/2}$, where the “Kompaneets” rate is $\Gamma_K \equiv 4n_e \sigma_T T_e / m_e$. The approximate constancy of $x^3(\Gamma_{bremss} + \Gamma_{DC})$ has been exploited to obtain this result. If we assume $\dot{Y}_\gamma$ photons per baryon are being injected with average energy $\bar{E}_\gamma$ at time t , adding to the $Y_{\gamma 0}$ photons per baryon already there, then we find the scaling parameter α_0 evolves according to

$$\begin{aligned} \frac{d\alpha_0}{dt} &= -\frac{\alpha_0}{\tau_D} + \left(\frac{\bar{E}_\gamma}{3.6T_\gamma} - 1 \right) 1.87 \frac{\dot{Y}_\gamma}{Y_{\gamma 0}}, \\ \tau_D &= 1.29x_0[(\Gamma_{bremss} + \Gamma_{DC})x^3]^{-1} = 1.29(\Gamma_K/4)^{-1}x_0^{-1}. \end{aligned} \quad (54)$$

Thus there is a damping term with timescale τ_D driving α_0 towards zero, i.e., a Planck distribution, against which the injection term tries to drive the

distortion. When the damping time is shorter than the expansion rate of the universe, any injected energy input would be rethermalized into a Planckian in equilibrium with the electrons within one Hubble time. This basically defines z_{Pl} . When the Kompaneets rate is a few times the expansion rate, x_0 will be low but α_0 will not be zero, and the BE form is appropriate. This defines z_{BE} . However, it is not until the Kompaneets rate is a few times below the expansion rate that the perturbative y -distortion solution prevails. This defines $z_y \approx z_{BE}/4$. Naturally z_{BE} and z_y scale in the way defined by Γ_K/H .

To constrain the allowed energy input in the Bose–Einstein regime, we take the BE distribution and linearize it in $\Delta T/T$ and α , where now both are frequency-independent. The photon number density and photon energy density are related to the unperturbed values by

$$n_\gamma = n_\gamma^{(0)} \left(1 + 3 \frac{\Delta T}{T} - \frac{\zeta_2}{\zeta_3} \alpha \right), \quad \rho_\gamma = \rho_\gamma^{(0)} \left(1 + 4 \frac{\Delta T}{T} - \frac{\zeta_3}{\zeta_4} \alpha \right),$$

where $\zeta_s = \sum j^{-s}$ denotes the Riemann zeta function of index s . With fixed photon number throughout energy injection, we must have $n_\gamma = n_\gamma^{(0)}$ remaining invariant, hence a relationship between the temperature perturbation and the chemical potential, $\Delta T/T = (\zeta_2/\zeta_3)\alpha/3$, leading to a relative energy perturbation

$$\frac{\delta \rho_\gamma}{\rho_\gamma^{(0)}} = \left(\frac{4\zeta_2}{3\zeta_3} - \frac{\zeta_3}{\zeta_4} \right) \alpha = 0.71\alpha. \quad (55)$$

Using the FIRAS constraint eq. (8), the allowed energy injection relative to the primeval radiation in the z_{Pl} to z_{BE} epoch is at most [58]

$$\frac{\delta E_{BE}}{E_{cmb}} \lesssim 6.4 \times 10^{-5} \quad (95\% \text{ CL}). \quad (56)$$

3.4 Recombination and photon decoupling

3.4.1 Hydrogen and Helium Recombination

The subject of the recombination of the primeval plasma was well developed immediately after the discovery of the background radiation [60, 61]. In this subsection, we display the ODEs we solved in the Bond and Efstathiou papers for hydrogen recombination [134, 88]. Although helium is neutral through hydrogen recombination, helium recombination is now also explicitly included for our anisotropy calculations for increased accuracy [302], and the relevant equations are also given. In implementing these equations, it is important to use very accurate and self consistent physical parameters.

The availability of photons per baryon in the background radiation illustrates that there are not enough photons above Lyman α energy to guarantee equilibrium of the $1s$ state with states above it, though there are plenty below the Balmer continuum. Thus absorption and production timescales for the $2s \rightarrow 3p$ transitions, for example, are measured in seconds at redshifts

above 1000. We can therefore take the population of excited states with $n > 1$ to be in thermal equilibrium with the $2s$ state. In the following, we denote the abundances per baryon of various hydrogen states $\{n, \ell\}$ by $Y_{n, \ell}$, the total abundance of hydrogen atoms and ions per baryon by Y_{HT} , and the free electron and proton abundances by Y_e and Y_p . The (positive) binding energy of the state n, ℓ is denoted by B_n , and $g_{n\ell} = 4(2\ell + 1)$ is its statistical weight, with the 4 coming from the proton and electron spins. As before, T_e and T_γ are the electron and photon temperatures in energy units ($k_B = 1$). When account is taken of the equilibrium associated with the fast timescales, the network of equations describing the normal recombination transition is:

1. Equilibrium of the state $\{n, \ell\}$ with the $2s$:

$$Y_{n\ell} = (g_{n\ell}/4)Y_{2s} \exp[-(B_2 - B_n)/T_\gamma]. \quad (57)$$

2. Baryon conservation:

$$Y_p + Y_{1s} + Y_{2s}\mathcal{Z}(T) = Y_{HT}, \quad \mathcal{Z}(T) = \sum_{n>1, \ell} (g_{n\ell}/4)e^{-(B_2 - B_n)/T_\gamma}. \quad (58)$$

The partition function for states above $n = 1$ is $\mathcal{Z}(T)$.

3. Loss of free electrons through recombination: $\dot{Y}_e = -\alpha_c n_B Y_e Y_p + Y_{2s} \beta_c$. Here α_c is the recombination rate, excluding direct recombinations to the ground state since the released photon above the Lyman edge leads immediately to another ionization. The factor

$$\beta_c = e^{-B_2/T_\gamma} \left(\frac{m_e c^2 T_e}{2\pi (\hbar c)^2} \right)^{3/2} \alpha_c \quad (59)$$

describes the detailed balance relating the photoionization rate to the recombination coefficient α_c . For α_c , we use the analytic approximation

$$\alpha_c = 1.948 \times 10^{-13} (10^4 \text{ K}/T_e)^{1/2} \varphi(y) \text{ cm}^3 \text{ s}^{-1}, \quad (60)$$

$$\varphi(y) \approx \frac{1}{2} (1.735 + \ln y + y^{-1}/6) - (1 - y^{-1} - 2y^{-2}), \quad y \equiv \frac{13.6 \text{ eV}}{T_e}$$

(Bates and Dalgarno [62]).¹

4. $1s$ production:

$$\dot{Y}_{1s} = \frac{Y_{2s}}{\tau_{2\gamma}} - \frac{Y_{1s}}{\tau_{2\gamma}} e^{-(B_1 - B_2)/T_\gamma} + \frac{R}{n_B}. \quad (61)$$

¹This recombination rate is superior to the Boardman [63] form, $\alpha_c = 2.84 \times 10^{-13} T_{e4}^{-1/2}$, $T_{e4} \equiv T_e/10^4 \text{ K}$, used by Peebles and to the oft-used Seaton approximation, $\alpha_c = 2.6 \times 10^{-13} T_{e4}^{-0.85}$. The latter is accurate at 10^4 K , but differs from the Osterbrock [64] values by 3% at 5000 K, 9% at 2500 K and by 19% at 1250 K, whereas the formula we adopt differs by only a percent in all three cases (and by even less from the tabulated values of Bates and Dalgarno). The original Peebles formula differs by 12%, 26% and 37%, respectively.

The first term describes the $2s \rightarrow 1s + \gamma\gamma$ transition, with lifetime $\tau_{2\gamma}=0.12$ s. The second describes the rate at which Lyman alpha photons from the $2p \rightarrow 2s + \gamma$ transition are shifted out of the line due to the expansion of the Universe before they can be reabsorbed. Thermal CMB photons are irrelevant for this, since, at the temperatures of recombination, essentially no photons with energies as high as Ly alpha exist. (Balmer lines do yield a thermal distribution function.) Thus a detailed solution of the photon distribution function across the line, including redshift effects, is needed. This is straightforward. For given Y_{2s} and Y_{1s} , Peebles shows

$$R = H(a) \frac{(B_1 - B_2)^3}{(\pi^2 (\hbar c)^3)} \left[\frac{Y_{2s}}{Y_{1s}} - e^{-(B_1 - B_2)/T_\gamma} \right]. \quad (62)$$

Here the Hubble parameter is $H(a)$.

The net effect of the rapid equilibration of the $2s$ state with the $2p$ and higher states yields the equation

$$\dot{Y}_{2s} = \alpha_c n_B Y_e Y_p - Y_{2s} \beta_c - \frac{Y_{2s}}{\tau_{2\gamma}} + \frac{Y_{1s}}{\tau_{2\gamma}} e^{-(B_1 - B_2)/T_\gamma} - \frac{R}{n_B}. \quad (63)$$

The rates are large enough that $\dot{Y}_{2s}$ can be taken to vanish, yielding an expression for Y_{2s} in terms of Y_{1s} and Y_e . Baryon conservation gives a further relation of Y_{1s} in terms of $Y_p = Y_e$, so the entire system of equations reduces to one for the evolution of the free electron abundance, Y_e .

Denoting the ionization fraction by $x = Y_p/Y_{HT}$, putting $Y_{1s}/Y_{HT} \approx 1 - x$, and transforming the time derivative to one over the photon temperature, $T_\gamma = T_{\gamma*}/\bar{a}$, we have

$$\frac{dx}{dT_\gamma} = \frac{1}{T_\gamma H(a)} \frac{n_B Y_{HT} \alpha_c x^2 - \beta_c (1 - x) e^{-(B_1 - B_2)/T_\gamma}}{1 + \beta_c \left(\tau_{2\gamma}^{-1} + \frac{(B_1 - B_2)^3 H(a)}{(\pi^2 (\hbar c)^3) n_B Y_{HT} (1 - x)} \right)^{-1}}. \quad (64)$$

This is a stiff equation. At high redshift, Saha equilibrium holds, with $\dot{Y}_{1s}=0$, thus $Y_{2s} = Y_{1s} e^{-(B_1 - B_2)/T_\gamma}$, and $\dot{Y}_e=0$, hence

$$Y_{1s} = c_{HI} Y_e Y_p, \quad c_{HI} = n_B e^{B_1/T_\gamma} \left(\frac{m_e c^2 T_e}{2\pi (\hbar c)^2} \right)^{-3/2} \frac{g_{HI}}{g_e g_p}, \quad (65)$$

The statistical weights are $g_e = 2, g_p = 2, g_{HI} = 4$. but by $T_\gamma \approx 4000$ K one should shift over to the ODE solution.

This equation is coupled to the Compton cooling equation (30) for the evolution of the electron temperature T_e as it breaks equality with the photon temperature T_γ to follow the $(1 + z)^2$ redshift evolution of a nonrelativistic ideal fluid:

$$\frac{1}{a^2} \frac{d}{dt} T_e a^2 = - \frac{8\sigma_T \rho_\gamma}{3m_e c} \frac{Y_e}{Y_T} (T_e - T_\gamma) - \left[\frac{2}{3} \frac{\dot{Y}_e}{Y_T} (B_1 + \frac{3}{2} T_e) \right]. \quad (66)$$

Here $Y_T \equiv Y_{HT} + Y_e + Y_{HeT}$ is the number of gas particles per baryon. (The term in square brackets from the binding and thermal energy gained when an electron recombines is ignorable here.) The large value of the photon energy density ρ_γ ensures that this Compton heating keeps T_e and T_γ nearly equal until a redshift below about 400 (as shown in fig. 3(c)). These equations must be integrated numerically with stiff ODE solvers. Solutions for some CDM models are shown in fig. 3(a). If one is just interested in the development of anisotropies, the critical region is not around the redshift ~ 1500 when the universe passes from 95% to 10% ionized, but rather a redshift interval from about 1200 to 900 when the radiation passes from being tightly coupled to freely streaming, when the optical depth to Thomson scattering, ζ_C defined by eq. (33), passes through unity. The final values of the residual ionization are also of interest since those few free electrons present catalyze the formation of molecular hydrogen, which can be an important coolant in the first objects that collapse in the Universe.

Krolik [65] discusses extra Fokker–Planck diffusive terms arising from scattering in the lines, but shows that these result in numerically small corrections to recombination over that obtained using the system of equations given here.

Although photons are quite tightly coupled to the baryons when helium recombines, for high precision calculations of CMB anisotropies at small angular scales the effect should be taken into account [302]. With more free electrons present, the photons do not diffuse as easily. It seems to be adequate to solve for the Saha equilibrium rather than doing the full time evolution as is required for hydrogen recombination. One should solve for the ionization fractions of the states of helium and hydrogen together, in practice done by iterating the following equation and demanding convergence in Y_{HI} and Y_{HeII} :

$$\begin{aligned} Y_e &= Y_{HT} + 2Y_{HeT} - (Y_{HI} + 2Y_{HeI} + Y_{HeII}), \\ Y_{HI} &= Y_{HT}/(1 + (c_{HI}Y_e)^{-1}), \\ Y_{HeII} &= Y_{HeT}/(1 + (c_{HeII}Y_e)^{-1} + c_{HeI}Y_e), \\ Y_{HeI} &= Y_{HeT}/(1 + (1 + (c_{HeII}Y_e)^{-1})(c_{HeI}Y_e)^{-1}). \end{aligned} \quad (67)$$

The coefficients entering are

$$\begin{aligned} c_{HeI} &= n_B \left(\frac{m_e T_e}{2\pi(\hbar c)^2} \right)^{-3/2} e^{B_{HeI}/T_\gamma} \frac{g_{HeI}}{g_e g_{HeII}}, \\ c_{HeII} &= n_B \left(\frac{m_e T_e}{2\pi(\hbar c)^2} \right)^{3/2} e^{B_{HeII}/T_\gamma} \frac{g_{HeII}}{g_e g_{HeIII}}, \end{aligned} \quad (68)$$

with statistical weights reflecting the spinless alpha particle in the fully ionized state, $g_{HeIII} = 1$, the electron spin in the once-ionized helium hydrogenic ground state, $g_{HeII} = 2$, and the two electrons in the singlet 1S_0 ground state of neutral helium, $g_{HeI} = 1$. The partition functions can be assumed to be temperature independent. The binding energies are $B_{HeI} = 24.6$ eV, $B_{HeII} = 54.4$ eV. When c_{HeI}^{-1} and c_{HeII}^{-1} are *very* small, helium is fully recombined and the hydrogen-only Saha equation is adequate to solve.

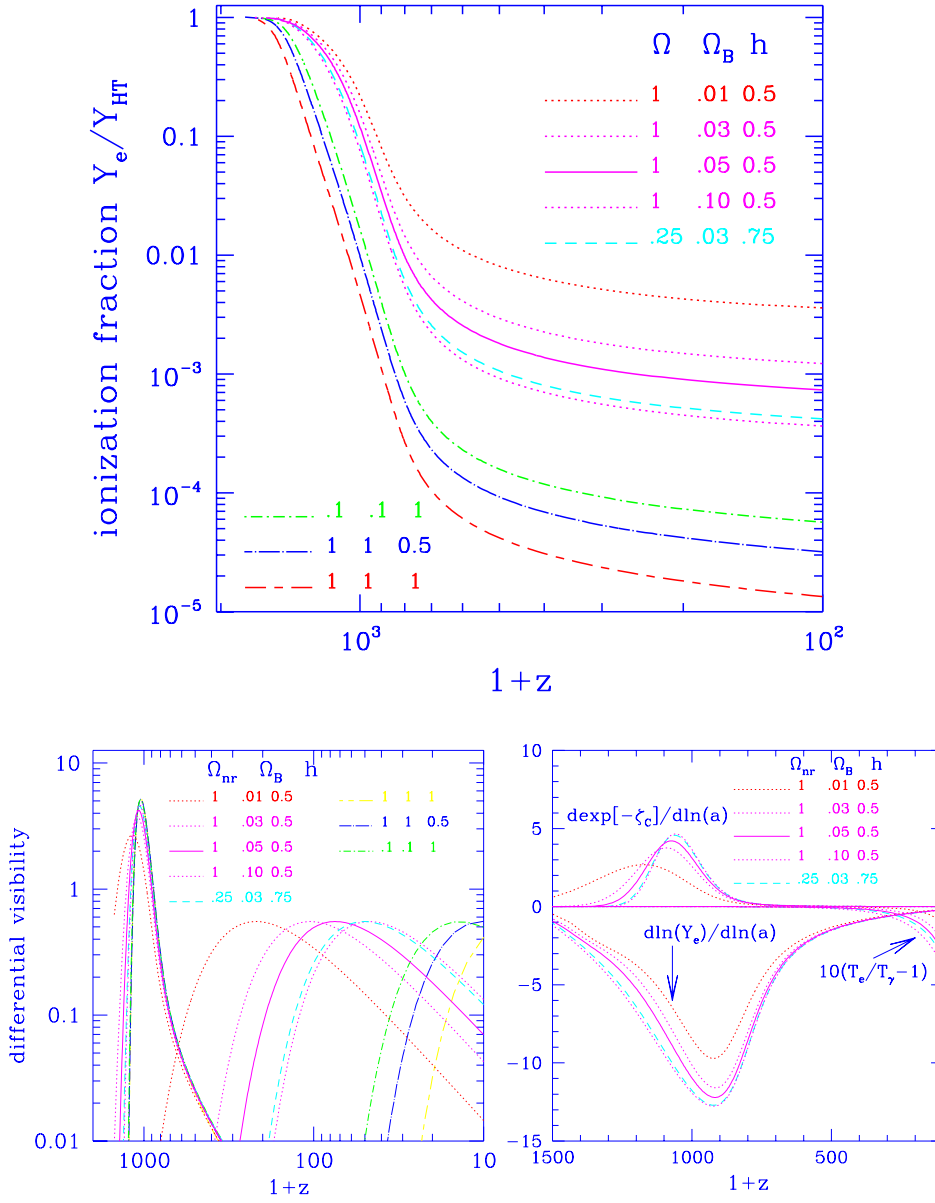


Figure 3: (a) Evolution of the ionization fraction. Effect of varying Ω_B, Ω_{nr}, h . (b) Differential visibility functions $de^{-\zeta_c}/d\ln \bar{a}$ for standard recombination (concentrated around $z \approx 1000$, rather like a Gaussian in τ) and for “no recombination”. (c) Closeup of (b), the $Y_e \sim a^{-p}$ power, and on the extreme left the relative difference between the electron and photon temperatures amplified by 10.

3.4.2 Visibility and decoupling

The visibility of the Universe to Thomson scattering is defined by $e^{-\zeta_C}$ and the differential visibility by $\mathcal{V}_C \equiv de^{-\zeta_C}/d\tau = e^{-\zeta_C}/\tau_C$, where $\tau_C^{-1} \equiv \bar{a}\bar{n}_e\sigma_T$. Figure 3(b) shows $\mathcal{V}_C/(H\bar{a})$ for the universes of (a); a closeup of a subset of the models is shown in (c). For normal recombination the differential visibility is sharply peaked, only weakly dependent on cosmological parameters. Although the distribution is somewhat skew, a Gaussian fit is not a bad approximation. We define the conformal time of decoupling τ_{dec} to be where $\mathcal{V}_C$ has a peak and the width of decoupling, $R_{\mathcal{V}_C,dec}$, to be the fwhm of $\mathcal{V}_C$ times a factor 0.425, which turns the fwhm into a dispersion for a Gaussian. The corresponding expansion factors are $\bar{a}_{dec}$ and $\sigma_{a,Cdec}$, related by

$$\begin{aligned}\tau_{dec} &= 190 \Omega_{nr}^{-1/2} \text{h}^{-1} \text{Mpc} (10^3 a_{dec})^{\frac{1}{2}} \left[\left(1 + \left(\frac{a_{eq}}{a_{dec}} \right) \right)^{\frac{1}{2}} - \left(\frac{a_{eq}}{a_{dec}} \right)^{\frac{1}{2}} \right], \\ R_{\mathcal{V}_C,dec} &= 9.5 (10 \sigma_{a,Cdec}) \Omega_{nr}^{-1/2} \text{h}^{-1} \text{Mpc} \frac{(10^3 a_{dec})^{1/2}}{\left(1 + \frac{a_{eq}}{a_{dec}} \right)^{1/2}} \approx \frac{1}{2} \sigma_{a,dec} \tau_{dec}, \\ a_{eq} &\equiv \frac{\Omega_{er}}{\Omega_{nr}} \approx [24200 \Omega_{nr} \text{h}^2]^{-1}, \quad \text{CDM: } 0.06 \lesssim \sigma_{a,Cdec} \lesssim 0.1. \quad (69)\end{aligned}$$

The $a_{eq}/a_{dec} \sim [6\Omega_{nr}(2\text{h})^2]^{-1}$ corrections usually cannot be ignored. For normal recombination CDM-dominated universes, the $\sigma_{a,dec}$ range, as measured from the fwhm of the fig. 3(b) curves, imply the Gaussian width $R_{\mathcal{V}_C,dec}$ is only about 0.03–0.05 of the horizon size at decoupling. The last scattering region is therefore quite thin, with typical (comoving) Gaussian width $\sim (5-10)\Omega_{nr}^{-1/2} \text{h}^{-1} \text{Mpc}$. The Gaussian approximation,

$$\mathcal{V}_C \equiv \frac{de^{-\zeta_C}}{d\tau} \approx \frac{\exp[-(\tau - \tau_{dec})^2/(2R_{\mathcal{V}_C,dec}^2)]}{(2\pi R_{\mathcal{V}_C,dec}^2)^{1/2}} \quad (70)$$

is not bad for these cases, and is nice for analytic purposes ([2], section 5.2).

Figure 3(c) shows how the instantaneous power law scaling $p_{e,dec} \equiv -d \ln Y_e / d \ln a$ varies with redshift. Around decoupling $p_{e,dec} \sim 10$ is typical. The Compton scattering time is related to the Hubble time at decoupling by

$$\begin{aligned}\mathcal{V}_C \text{ max:} \quad & n_e \sigma_T / H = (p_{e,dec} + 2) \quad \text{defines } \bar{a}_{dec}, \tau_{dec}, \\ \frac{\mathcal{V}_C}{(H\bar{a})} \text{ max:} \quad & n_e \sigma_T / H = (p_{e,dec} + 2 - q_{dec}), \\ & q_{dec} = \frac{1}{2} \frac{(1 + 2(a_{eq}/a_{dec}))}{(1 + (a_{eq}/a_{dec}))}. \quad (71)\end{aligned}$$

We could also define the decoupling redshift when $\mathcal{V}_C/(H\bar{a})$ has a maximum: this occurs slightly later than that determined by $\mathcal{V}_C$. Here q_{dec} is the value of the deceleration parameter at decoupling. The Compton time is therefore only

$\sim 5\%$ of the “horizon” size at decoupling. In section 5.1.2, we shall see that a local measure of the width of the visibility at time τ is useful to characterize the damping of anisotropies associated with the fuzziness of the last scattering surface:

$$R_{\mathcal{V}_C}(\tau) \equiv \left| \frac{\partial^2 \ln \mathcal{V}_C}{\partial \bar{\chi}^2} \right|^{-1/2} = \frac{(\bar{H}\bar{a})^{-1}}{\left[(p_e + 2)(-q + \frac{\bar{n}_e \sigma_T}{H} + \frac{d \ln(p_e + 2)}{d \ln a}) \right]^{1/2}}.$$

This also gives $\sigma_{a,C} \approx \bar{H}\bar{a}R_{\mathcal{V}_C}$; in particular, if we substitute $n_e \sigma_T / H = p_{e,dec} + 2$ in this, note from the figure that $d \ln(p_e + 2)/d \ln a$ is typically 2 or 3, we get $0.06 \lesssim \sigma_{a,Cdec} \lesssim 0.1$ for $7 \lesssim p_{e,dec} \lesssim 12$, in good accord with the fwhm estimates. This expression also shows that $R_{\mathcal{V}_C,dec} \approx \tau_{C,dec}$. (The time-dependent $R_{\mathcal{V}_C}(\tau)$ expression must eventually break down, once $n_e \sigma_T / H$ drops below the deceleration parameter q .)

Figure 3(b) shows the dramatic effect of early reionization on the visibility. For full ionization ($p_{e,dec} = 0$), the redshift at which $\mathcal{V}_C/(H\bar{a})$ peaks is exactly where the optical depth to us is unity,

$$z_{\zeta_C=1} \approx 10^{2.1} \left(\frac{\Omega_B h}{0.02} \right)^{-2/3} \Omega_{nr}^{1/3}. \quad (72)$$

The redshift z_{dec} at which $\mathcal{V}_C$ peaks is 20% smaller. Figure 3 shows the Gaussian approximation is not very good (the half power points in τ are at $0.75\tau_{dec}$ and $1.5\tau_{dec}$, with “Gaussian” width of $0.32\tau_{dec}$). For the typical $\Omega_{nr}=1$, $\Omega_B=0.05$ dark-matter dominated Universes, $z_{dec} \approx 130$ and $R_{\mathcal{V}_C,dec} \approx 170 h^{-1}$ Mpc, but for the $\Omega_B = \Omega_{nr} = 0.1$ universe in fig. 3, whether open or vacuum-dominated (to make $\Omega = 1$), the decoupling redshift is pushed dramatically forward, to $z_{dec} \approx 28$ and $R_{\mathcal{V}_C,dec} \approx 360 h^{-1}$ Mpc.

3.5 Reionization of the universe

Erasure of CMB temperature anisotropies is dramatic if re-ionization occurs earlier than the minimum redshift required to make the optical depth to us unity, eq. (72). Although this seems unlikely in CDM dominated models [134], it is reasonable to expect a 10% effect on $\Delta T/T \propto e^{-\zeta_C}$ even if re-ionization occurs as late as $z_{\zeta_C=1}/3$, say, since $\zeta_C = [(1+z)/(1+z_{\zeta_C=1})]^{3/2}$.

The Gunn–Peterson test shows that the cumulative optical depth to Lyman alpha radiation back to the most distant quasars at $z \sim 3$ is less than 0.05 implying the universe is extremely highly ionized with neutral hydrogen fraction $Y_H \lesssim 10^{-6}$. Quasars, which contribute a significant amount of this ionizing flux, are expected to have formed too late to have had much influence on CMB anisotropies. An early population of massive stars or more exotic sources such as decaying Big Bang relic particles with a radiative channel could reionize early enough. In [66], we estimated the fraction of the closure density in massive stars of various types required for re-ionization to occur via the overlapping of the

HII regions they generate. We found that to reionize by $z_{\zeta C=1}$ requires a cosmic abundance of ionizing stars

$$\Omega_* = K 10^{-6} \left(\frac{\Omega_B h}{0.02} \right)^{0.8} (1 + \bar{\delta}_{gas})^{1.5}, \quad (73)$$

where K is a factor depending upon the type of stars: for stars with mass $\sim 30 M_\odot$, $K \approx 30$ if they have Population III abundances (i.e., with essentially no heavy elements) and is somewhat higher if there is Population II metallicity, while for the limiting case of Very Massive Objects (mass $\gtrsim 100 M_\odot$), $K \approx 1$ for Population III abundances and $K \approx 5$ for Population II abundances. Ω_* depends upon the overdensity of the gas relative to the background, $1 + \delta_{gas}$, i.e., the clumpiness factor. Ω_* is lowest if the gas is unclumped, but the gas in the neighborhood of the stars will be overdense and the HII region would first have to break out of this gas before entering into the $\delta_{gas} \approx 0$ background medium. It is therefore unclear what to take for the average $\bar{\delta}_{gas}$ entering eq. (73), and thus how much larger a fraction than 10^{-6} in massive stars is required.

To assess whether it is plausible that such relatively large fractions of the universe can have gone into massive stars by $z_{\zeta C=1}$, we use the Press–Schechter formula [67] for the fraction of the baryons that would be in collapsed objects by redshift z , $\Omega_{Bcoll}(z) = \Omega_B \text{erfc}(\nu_{coll}/\sqrt{2})$, where $\nu_{coll}(z) \equiv f_{coll}(1+z)/\sigma_{\rho B}$. Here the factor $f_{coll} \approx 1.686$ is the average linear density fluctuation within a sphere needed for that sphere to have collapsed to infinite density when nonlinearities are included. $\sigma_{\rho B}(z)$ denotes the rms level of the gas density fluctuations at redshift z . (For rare events, i.e., high ν_{coll} , we have $\Omega_{Bcoll} \approx \Omega_B (2/\pi)^{1/2} \nu_{coll}^{-1} e^{-\nu_{coll}^2/2}$; the better physically-motivated “peak-patch picture” [68] based on collapse about peaks in the linear density field yields similar results.)

There is a natural filter $\sim 1 h^{-1} \text{ kpc}$ for the gas associated with the Jeans mass at recombination. In [134], we showed that σ_ρ on this scale is typically $\approx 20\sigma_8/(1+z)$ for initially scale invariant $\Omega_{nr} = 1$ CDM-dominated models with $h = 0.5$ and about the same for initially scale invariant nonzero Λ models with $h \approx 0.75$ and $\Omega_{nr} \approx 0.3$. Here σ_8 denotes the rms linear density fluctuations on cluster scales at the current time. For CDM and the nonzero Λ models, we have $\nu_{coll}(z_{\zeta C=1})$ ranging from about 7 to 10, hence $\Omega_{Bcoll}(z_{\zeta C=1})$ is very tiny indeed. However, by $z_{\zeta C=1}/3$ it would have grown to a number which can exceed Ω_* . Thus, although we concluded in [134] that the drastic case of extreme damping of small angle CMB anisotropies was unlikely unless there were an extremely high efficiency of massive star formation from collapsed gas, it is quite conceivable that there will be some small effect from the earliest generation of stars on the anisotropies provided there is a reasonable amount of “short-distance power” in the density fluctuation spectrum.

What complicates this enormously is that the entities which form may well be rather fragile with a small binding energy, easily disrupted by the massive

stars they generate. But it is also possible that the amount of nonlinear gas could be amplified by the explosion of such stars sweeping up shells of gas far from the parent object. It is difficult to argue definitively either way and this issue of efficiency and amplification or suppression will likely remain a subject of uncertainty in interpretation of CMB anisotropies for a long time to come. For recent discussions of the issues involved in reionization see [69, 70].

Although the influence of early reionization on inflation-based CDM models and models with nonzero Λ is ambiguous, the situation seems clearer in other models. In isocurvature baryon models with (nearly) white noise initial conditions popular in the late seventies [71], the first objects collapse at $z \sim 300$, making reionization easy, and, indeed, expected. Similarly, in models in which there are isocurvature seeds, such as in texture models, one also expects early ionization to be quite plausible, although by no means certain.

If there is no recombination, there is a constraint from the y -distortion on how early energy can be injected:

$$z_{max,reh} \approx 10^{3.8} \left(\frac{\Omega_B h}{0.02} \right)^{-2/3} \Omega_{nr}^{1/3}. \quad (74)$$

This is a result from Zeldovich and Sunyaev [34], revisited by Bartlett and Stebbins [72], which I modified to take into account the FIRAS limit [12]. This limit can be avoided if one can sustain a temperature of the cooling electrons to be nearly the CMB temperature. In any case, it is no limitation for the low Ω_B favored by standard Big Bang nucleosynthesis [73].

3.6 Post-recombination energy sources

After recombination, we expect energy release to accompany the formation of nonlinear cosmic structure as stars, black holes etc. form. Although the limits on this release in the CMB region are now very stringent, they are not as strong in the optical and near infrared. I now survey a number of sources that would be expected to contribute to a background, choosing normalization parameters to be relatively conservative. Even so they are not far off the FIRAS bound (eq. 9), $< 2.5 \times 10^{-4}$ from 500–5000 μm – a useful limit to bear in mind when considering the following energy source formulas. On the other hand, there is a tentative identification of a sub-mm background in the FIRAS data [49] in the range $\sim 200 - 1000 \mu\text{m}$, with energy $\delta E/E_{cmb} \sim 10^{-3}$ longward of $\sim 400 \mu\text{m}$, which partly mimics the Galactic contribution (and could be partly due to cold high latitude Galactic dust [48]). There are also residuals after source subtractions in the DIRBE data which could be interpreted as a cosmological infrared background at shorter ($\sim 1 - 200 \mu\text{m}$) wavelengths at the $\delta E/E_{cmb} \sim 10^{-2}$ level [79, 80]. These are shown in fig. 3.6.

We first consider an exotic source before the more prosaic ones we know must exist at some level. Decaying (cold) particles with a radiative channel

$X \rightarrow X' + \gamma$ having a branching ratio $B_{X\gamma}$ contribute a relative energy

$$\frac{E_{decay}}{E_{cmb}} \sim 0.02 B_{X\gamma} \Omega_{X,i} h^2 \frac{10^6}{(1 + z_{dec})} \quad (75)$$

to the Universe, where $\Omega_{X,i}$ is the initial density parameter of the cold particles which are destined to decay, which may easily be in excess of unity; e.g., for keV neutrinos it is 40. z_{dec} is the decay redshift, when the lifetime equals the Hubble time. In cases like this, $z_{dec} > z_{Pl}$ unless the branching ratio is tiny, i.e., with a lifetime shorter than a month. And if a considerable fraction of the CMB energy were created this way, the success of standard Big Bang nucleosynthesis would come into jeopardy. If the particle has a longer lifetime and if there is dust to reprocess the radiation into the sub-mm band probed by FIRAS, the constraints on the branching ratio are quite severe; if there is no dust so the decay radiation is just redshifted, then it would lie at shorter wavelengths where the bounds are much less stringent.

The nuclear energy output of stars with efficiency ϵ_{nuc} radiating at redshift z_* with an abundance Ω_* relative to the CMB is

$$\frac{E_*}{E_{cmb}} \approx 0.03 \left(\frac{\Omega_* h^2}{0.001} \right) \frac{5}{(1 + z_*)} \frac{\epsilon_{nuc}}{0.004}. \quad (76)$$

Massive stars have an efficiency which is not much less than the maximum value of 0.004 for Very Massive Objects [66], those with mass $> 100 M_\odot$. The radiant energy release from stars which eject a mass $Z_{ej} M$ in metals when they undergo supernova explosions is limited by the metal fraction Z they contribute to a gas of density Ω_{gas} ,

$$\frac{E_{preSN*}}{E_{cmb}} \approx 0.0008 \frac{Z}{10^{-3}} \frac{\Omega_{gas} h^2}{0.01} \frac{Z_{ej}}{0.2} \left(\frac{M}{20 M_\odot} \right)^{0.5}. \quad (77)$$

Radiation generated by mass accreting onto black holes with an efficiency ϵ_{acc} , typically taken to be about 0.1 for quasar models, delivers energy

$$\frac{E_{BHacc}}{E_{cmb}} \sim 0.0008 \frac{\Omega_{BHacc} h^2}{10^{-6}} \frac{5}{(1 + z_{acc})} \frac{\epsilon_{acc}}{0.1}. \quad (78)$$

We might reasonably expect that $\Omega_* \epsilon_{nuc}$, $Z \Omega_{gas}$ and $\Omega_{BHacc} \epsilon_{acc}$ would be larger than the normalizations indicate and so they would be in conflict with the FIRAS limit, eq. (9), if that radiation were to find its way to the sub-mm. In particular, the prospect of ($\sim 10^2$ – $10^5 M_\odot$) VMO remnant black holes forming a considerable component of the dark matter is ruled out if the unavoidable thermonuclear energy release prior to collapse passed through pregalactic dust or through dusty galaxies.

Although there is a contribution from the gravitational energy released during the collapse of various structures in the universe in all wavebands, it is typically smaller than that from other sources. Letting $\Omega_{B,coll} f_{cool}$ be the

density of baryons which have cooled in a potential well characterized by the three-dimensional virial velocity dispersion v_T which formed at redshift z_{coll} and taking the average over all collapsed structures, we get an energy release

$$\frac{E_{formation}}{E_{cmb}} \sim 0.0002 \left\langle \frac{\Omega_{B,coll} h^2 f_{cool}}{10^{-3}(1+z_{coll})} \left(\frac{v_T}{1000 \text{ km s}^{-1}} \right)^2 \right\rangle. \quad (79)$$

Taking typical parameters for gas that has cooled in forming galaxies gives a value of order ten lower than the prefactor.

The FIRAS limit on the y -distortion does place a powerful constraint on how effective explosions could have been in generating cosmic structure. As Ikeuchi and Ostriker emphasized (e.g., [74]), a predominantly hydrodynamic explanation for cosmic structure development is a perfectly reasonable extrapolation of known behavior in the interstellar medium to the pregalactic medium. In [2], I gave a conservative lower estimation of the amount of Compton cooling that would have accompanied the explosive formation of bubbles of radius R_{exp} with filling factor f_{exp} by equating the thermal energy to the minimum energy per baryon required to scour out a bubble of size R_{exp} at redshift z_{exp} :

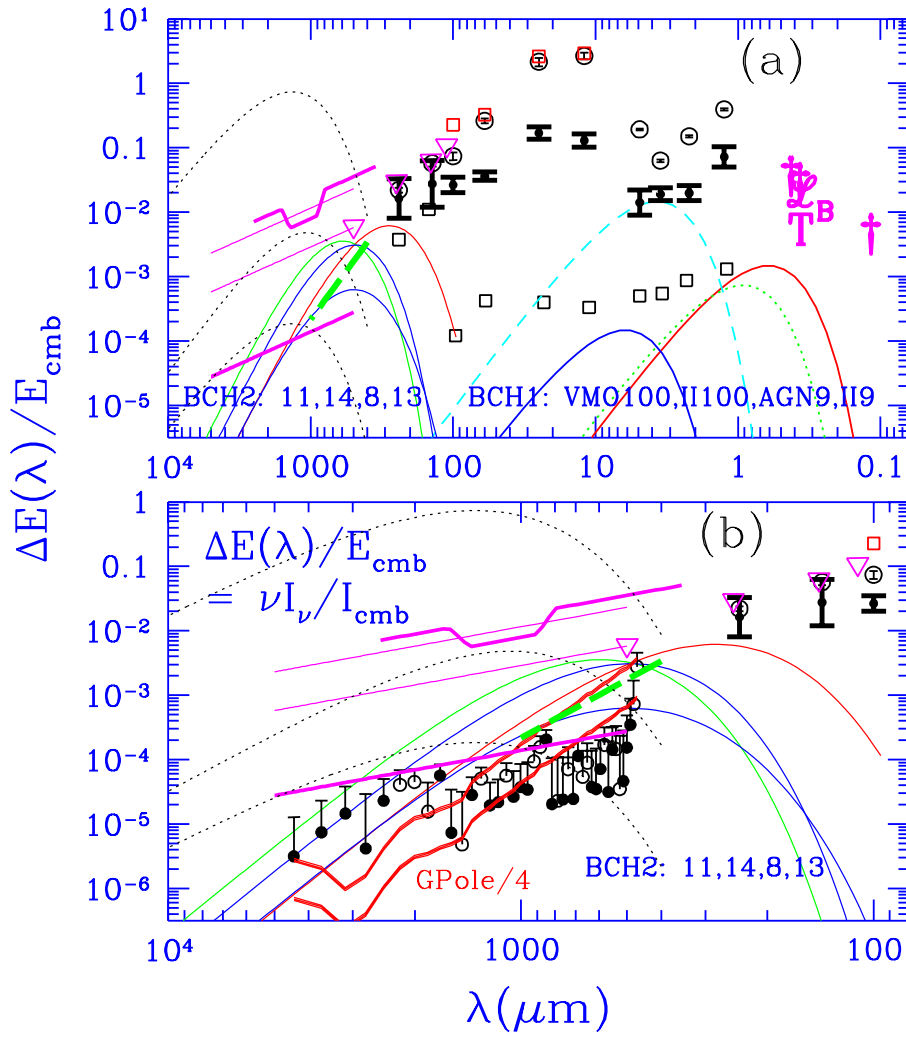
$$\frac{E_{Compton cool}}{E_{cmb}} \sim u 10^{-3} f_{exp} \left(\frac{R_{exp}}{20 \text{ h}^{-1} \text{ Mpc}} \right)^2 \Omega_B h \Omega_{nr}^{1/2}, \quad (80)$$

with $u \sim 1/2$. Chris Thompson [78] gave a more refined derivation and got the same functional form with prefactors u ranging from $1/3$ to 1 , assuming that the electrons would be much cooler than the ions. The z_{exp} dependence is weak for redshifts $\gtrsim 10$ when Compton cooling dominates, and $\sim (1+z_{exp})^{5/2}$ below. Thus the FIRAS limit of 6×10^{-5} very strongly constrains the scale R_{exp} and/or the filling factor f_{exp} . If supernova explosions were responsible for energy injection, one expects that the presupernova light radiated would be in excess of the explosive energy by a factor in excess of 100 , which would lead to even stronger restrictions on the model; and if the supernova debris is metal-enriched, the allowed amount of metals also poses a strong constraint.²

Another victim of the powerful FIRAS y -distortion limit was the superconducting cosmic string model of structure formation, in which the strings would radiate magnetohydrodynamic and (damped) extremely long frequency waves that would heat the medium, giving a picture of structure formation similar in spirit to the explosion model, but with a more exotic energy source. Thompson has estimated $\delta E/E_{cmb} \sim 10^{-2.2} (G_N \mu / 10^{-6})$ for $\Omega_B (2h)^2 = 0.1$, where G_N is Newton's constant and μ is the string tension. This is much too large for the range of μ needed to make the model viable.

Figure 3.6(a) compares current DIRBE constraints with sample theoretical models having spectra peaking in the near-infrared [81], examples of the energy

²The limits from anisotropy are not as strong: the packed shell model above gives anisotropies at the few times 10^{-5} level, but the expectation was that in the early fireball development phase, the hot gas would create large anisotropies [2, 75], although if $T_e \ll T_{ion}$ Thompson suggests these can be avoided.



(caption next page)

Table 1: Sample dust emission models

Model:	M8	M11	M14	M13	pk
Ω_{dust}	10^{-6}	10^{-5}	10^{-5}	10^{-6}	$\sim 10^{-6}$
$\delta E_{\text{inj}}/E_{\text{cmb}}$	0.04	0.004	0.004	0.004	$0.01S$
Mode	burst	burst	steep $\dot{E}$	flat $\dot{E}$	burst
z_{gf}	5	9	5	9	6–4
$\delta E_{\text{NIR}}/\delta E_{\text{inj}}$	0.82	0.01	0.14	0.81	–

Figure 4: (previous page) (a) The intensity levels νI_ν in units of the total CMB intensity $I_{cmb} = 10^{-3} \text{ erg cm}^{-2} \text{ s}^{-1} \text{ sr}^{-1}$ for a variety of near-IR and far-IR models of energy generation associated with galaxy formation are compared with current limits and potential measurement levels. The straight heavy line in the sub-mm shows the current FIRAS constraint on spectral distortions of the CMB, the light upper lines show the 1990 announcement limit, and the improvement one year later (using Baade window observations). The upper heavy line shows the COBRA limit. Typical optical and UV limits are denoted by daggers, IRAS measurements are solid squares. Open circles are DIRBE's "dark sky" values, hence upper limits to an infrared background, heavy error bars give the estimated DIRBE range of residuals at high Galactic after removing "foreground" sources, and the open squares denote the sensitivities DIRBE could in principle have gotten to with perfect source removal in 1 field-of-view after 1 year of integration. Inverted triangles are limits from the FIRAS HF channels. (b) A closeup in the sub-mm. The solid circle data points are positive FIRAS residuals the open data points are absolute values of negative FIRAS residuals. They are bounded by the solid line. If a spectrum mimics Galactic emission then it is not as strongly constrained. The solid line above *GPole*/4 is the FIRAS determination of the Galactic Pole emission, lower is 1/4 of this. The heavy large-dashed curves in both panels denote the tentative sub-mm background suggested for the FIRAS data (Puget et al. 1996). The upper dotted curve is the CMB blackbody, next is the perturbed CMB with fixed $\delta T = 0.004$, and last is the CMB times 0.00025. The solid curves below these are BCH2 primeval galaxy models of dust emission.

releases discussed above: *metal generating stars* which generate $\Omega_{metals} = 10^{-4}$ at $z = 100$ and at $z = 9$ (two solid lines), Eddington-limited accreting black holes (*AGN pre-cursors*) at $z = 9$ with $\Omega_{bh} = 10^{-5}$, VMOs with abundance $\Omega_{VMO} = 0.05$ at $z=100$ (dashed), that make $\Omega_{bh} = 0.025-0.05$. The curves scale up linearly with Ω_{metals} , Ω_{bh} , or Ω_{VMO} , and with $(2h)^2$. (All models shown have $\Omega = 1$ and $h=0.5$.) With increasing $1+z$ of formation, there is a linear increase to longer wavelength and a linear drop in the amplitude of the curves, as in the transition between the 2 solid curves.

If intervening dust is present, these curves will have the same underlying energy but be shifted, at least partially, into the sub-mm. Figure 3.6(a) shows the FIRAS bound given as a fraction of the CMB peak, while fig. 3.6(b) gives a closeup showing what freedom there really is, since the most powerful FIRAS bound was derived by modelling the Galactic emission so if cosmological sources can mimic that Galactic emission they are not as strongly constrained (e.g., [49]).

For the sub-mm theoretical curves shown, corresponding to the BCH2 [42] models listed in table 1, “normal” Galactic dust is assumed, with far infrared opacity index $\alpha_d = 1.5$, similar to the value derived for single temperature dust from the COBE observation. ($\alpha_d = 2$ is better motivated theoretically in this range, and, with a cold component added, by the COBE observations, section 3.2.7.) The dust abundance is Ω_d ($\sim 10^{-5}$ corresponds roughly to a Pop I abundance of dust in bright galaxies). The radiative energy input relative to that in the CMB is $\delta E_{inj}/E_{cmb}$, of which a fraction $\delta E_{NIR}/\delta E_{inj}$ is not absorbed by the dust, and just redshifts to appear as a near infrared background. This fraction depends upon the dust distribution. The peak of emission also depends upon how clumped the dust is; BCH2 used homogeneous models which have maximally cool dust, hence somewhat bigger emission in the FIRAS bands than the hotter compact dust of starbursts. A peak-patch model [68] for starbursting galaxies [82], including normal and dwarf contributions, radiating with a dust temperature $T_d = 30$ K that form according to a $\sigma_8 = 0.7$ CDM model, but which were not allowed to burst below redshift 4, is also given in the table, since maps based upon it are shown in fig. 15: the parameter S is variable, but should apparently be less than 0.1 to satisfy the COBE bounds of [12]; $S = 1$ would have all the normal galaxies that formed pass through a phase at birth during which their luminosity output was at the Arp 220 level – Arp 220 being the canonical strong starburst example. Wright et al. [58] and De Zotti et al. [83] give other versions of the constraints on dust emission from high (and low) redshift galaxies that can be derived from FIRAS.

4 Phenomenology of CMB anisotropy

Generally many sources will contribute to the CMB anisotropy pattern. Now that fluctuations in the temperature have been discovered, the challenge is to design experiments that can separate the many components that will be

present, in particular, the cosmological signals from those that are merely Galactic or conventionally extragalactic (e.g., radio galaxies). Ultimately, it will probably require a sophisticated combination of spectral and angular information, and cross-correlation with other datasets, such as X-ray and HI maps. With enough frequency bands covered, the prospects for separation on the basis of spectrum alone is not bad. Figure 5 draws together the spectral signatures of the different sources of anisotropy that are likely to appear, using eqs. (32), (39), (41), (48). Although the different signals are gratifyingly different, many parameters must be fit.

The angular patterns could also be used, for example to get rid of point sources. Of course, this can be dangerous since what we are trying to discover is the angular pattern in the background. We now turn to measures of this angular pattern, with special emphasis on the power spectrum as a way of codifying the contribution of different angular scales to the anisotropies for different cosmic signals. However, the patterns may be non-Gaussian, especially for secondary anisotropies, and so how the power is concentrated in hot and cold spots defines a crucial aspect of the distribution.

4.1 Statistical measures of the radiation pattern: $C(\theta), \mathcal{C}_\ell, \dots$

To relate observations of anisotropy to theory, statistical measures quite familiar from their application to the galaxy distribution have been widely used. Denote the radiation pattern as measured here and now by the two-dimensional random field $\Delta T(\hat{q})$, where $-\hat{q} = (\theta, \phi)$ is the unit direction vector on the sky (and $\hat{q}$ is the direction the photons are travelling in). The correlation function is

$$C(\theta) \equiv \left\langle \frac{\Delta T}{T_c}(\hat{q}) \frac{\Delta T}{T_c}(\hat{q}') \right\rangle, \quad \cos(\theta) = \hat{q} \cdot \hat{q}', \quad (81)$$

where T_c is the background temperature (the monopole). In theoretical treatments, a probability functional describes the distribution of the sky patterns. Generally all N -point correlation functions are required to specify the statistical distribution. The random field is statistically isotropic if all N -point functions are rotationally-invariant. In particular, this implies $C(\theta)$ is only a function of the angle between the vectors $\hat{q}$ and $\hat{q}'$. The theoretical correlation function is an ensemble-average of possible skies, while experimentally $C(\theta)$ must be an angle-averaged estimate for the patch of the sky over which the observations have taken place. Even if there were perfect resolution and all-sky coverage, the observed $C(\theta)$ and the theoretical $C(\theta)$ would differ. For realistic experiments, the errors arising from both observational sources and fluctuations because the observed patch of the sky is just one realization from the ensemble are crucial to properly include. The latter effect is called “cosmic variance” [90].

Other analogues of 3D measures that have been applied to CMB maps include: constructing the one-point distribution for $\Delta T/T$ as a function of resolution scale, the analogue of “counts-in-cells”; particular aspects are the rms fluctuation on a given resolution scale, and the skewness and kurtosis of

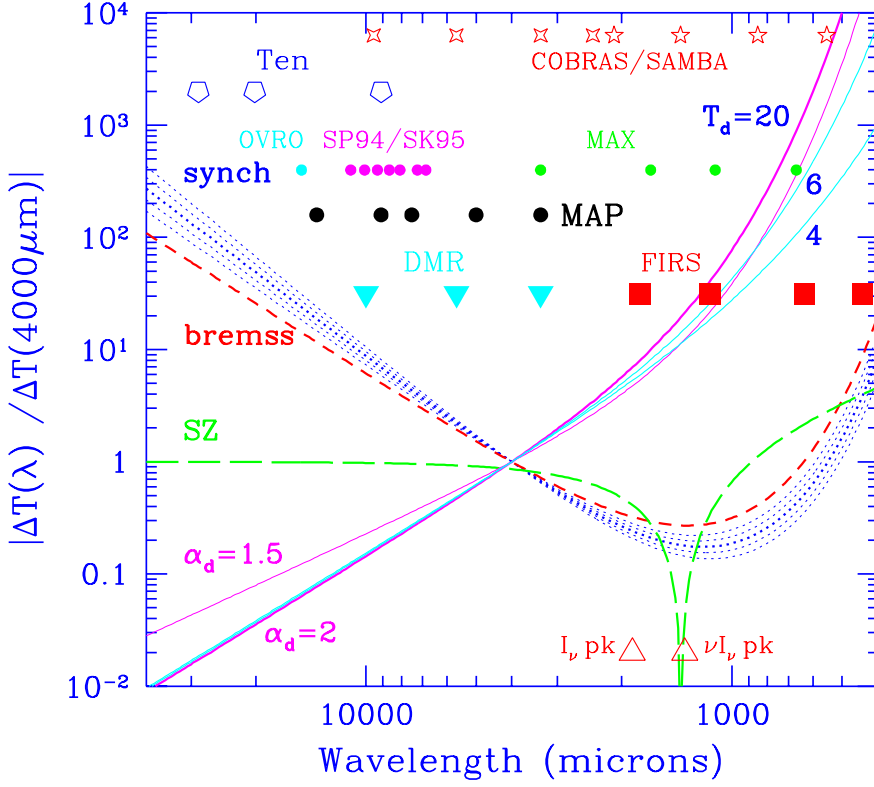


Figure 5: The flat spectrum in thermodynamic temperature predicted for primary anisotropies is contrasted with the spectral signatures for other sources of anisotropy (normalized at 4 mm): SZ anisotropies (long-dashed, with a sign change at 1300 μm); bremsstrahlung (short-dashed); synchrotron (dotted), with index varying from $p_s = 0.5$ to 0.9; and dust (with index $\alpha_d = 2$ as indicated by a two-temperature fit to COBE), both the usual Galactic dust at 20 K (heavy solid) and dust at 6 K and 4 K (light solid lines, which could represent a cold Galactic component or, e.g., 30 K dust radiating at redshift ~ 5); a shallower, less physically-motivated, $\alpha_d = 1.5$ dust opacity law for the 20K grains is also shown, appropriate for the single-temperature COBE fit. The frequency bands which various experiments probe are indicated. There is a minimum of the Galactic foregrounds at about 90 GHz, the highest frequency COBE channel.

the distribution; the statistics of hot and cold spots (high positive and negative excursions in the maps); the genus, etc. Many of these are rather obscured by the intrinsic observational noise, and only full scale Monte Carlo treatments are possible to assess how well a theory is faring.

As for the galaxy distribution on large scales, the most useful statistic is the power spectrum, $\mathcal{C}_\ell$, for a 2D distribution a function of multipole number ℓ . For CMB anisotropies, it is natural to expand the radiation pattern in spherical harmonics $Y_{\ell m}(\theta, \phi)$:

$$\begin{aligned} \frac{\Delta T}{T_c}(\hat{q}) &= \sum_{\ell m} a_{\ell m} Y_{\ell m}(\hat{q}) = \sum_{\ell} \left\{ \xi_{\ell 0} Y_{\ell 0}(\theta, 0) \right. \\ &\quad \left. + \sqrt{2} \sum_{m=1}^{\ell} Y_{\ell m}(\theta, 0) [\xi_{\ell m} \cos(m\phi) - \eta_{\ell m} \sin(m\phi)] \right\}, \end{aligned} \quad (82)$$

with the latter splitting the complex $a_{\ell m}$ into $\ell + 1$ symmetric real components, $\xi_{\ell m}$, and ℓ antisymmetric real components, $\eta_{\ell m}$, the symmetry defined by the behavior under change of the sign of the longitude:

$$\begin{aligned} a_{\ell 0} &= \xi_{\ell 0}; \quad \eta_{\ell 0} = 0; \quad a_{\ell m} = \frac{1}{\sqrt{2}}(\xi_{\ell m} + i\eta_{\ell m}), \\ a_{\ell -m} &= a_{\ell m}^* = \frac{1}{\sqrt{2}}(\xi_{\ell, m} - i\eta_{\ell, m}), \quad m \geq 1. \end{aligned} \quad (83)$$

If the temperature pattern is statistically isotropic, then $\langle a_{\ell m}^* a_{\ell' m'} \rangle = 0$ unless $\ell = \ell'$, $m = m'$. The nonzero components are the ensemble-averaged angular power spectrum,

$$\mathcal{C}_\ell \equiv \ell(\ell + 1)C_\ell/(2\pi), \quad C_\ell = \langle a_{\ell m}^* a_{\ell m} \rangle = \langle \xi_{\ell m}^2 \rangle = \langle \eta_{\ell m}^2 \rangle. \quad (84)$$

At high ℓ , this corresponds to the power in a logarithmic waveband $d \ln(\ell)$. The specific $\ell(\ell + 1)$ factor is chosen because $\mathcal{C}_\ell$ is predicted to be flat at small ℓ for theories with scale invariant adiabatic density perturbations [247] (section 5.1). In terms of a discrete “logarithmic integral”, $\mathcal{I}(f)$, of a function f_ℓ , defined by

$$\mathcal{I}(f) \equiv \sum_{\ell} f_{\ell} \frac{(\ell + \frac{1}{2})}{\ell(\ell + 1)}, \quad (85)$$

the correlation function is given by

$$C(\theta) = \sum_{\ell} \frac{2\ell + 1}{4\pi} C_{\ell} P_{\ell}(\cos(\theta)) = \mathcal{I}(C_{\ell} P_{\ell}(\cos(\theta))). \quad (86)$$

The rms fluctuations in the multipole ℓ are found by squaring the ℓ -poles of $\Delta T/T_c$ and averaging over angles:

$$\sigma_{T, \ell}^2 = \frac{1}{4\pi} \sum_{m=-\ell}^{\ell} |a_{\ell m}|^2 = \frac{1}{4\pi} \left(\xi_{\ell 0}^2 + \sum_{m=1}^{\ell} (\xi_{\ell m}^2 + \eta_{\ell m}^2) \right). \quad (87)$$

For example, $Q_{rms} = T_c \sigma_{T,2}$ is the quadrupole amplitude. The full four-year data [85] gives $Q_{rms} = 10.7 \pm 3.6 \pm 7.1 \mu\text{K}$, the first the 1-sigma statistical error, the second Galactic modelling errors; i.e., $\sigma_{T,2} = 0.4 \times 10^{-5} [1 \pm 0.3 \pm 0.7]$.

Over small patches of the sky, the curvature of the sky is not important and we can Fourier transform the radiation pattern:

$$\frac{\Delta T}{T_c}(\varpi|\hat{q}_P) = \int \frac{d^2\mathbf{Q}}{(2\pi)^2} \frac{\widetilde{\Delta T}}{T}(\mathbf{Q}) e^{i\mathbf{Q}\cdot\varpi}. \quad (88)$$

This is certainly useful for the fast Fourier transform can then be applied to small scale map construction. Here I describe the way we did this in [88]. Choose a pole $\hat{q}_P$ within the patch and, in the neighborhood of the pole, let $\varpi = (\varpi_x, \varpi_y) = \varpi(\cos\phi, \sin\phi)$, where $\varpi = 2\sin(\theta/2)$ is confined to the range $0 \leq \varpi < 2$. Its magnitude is $\varpi = |\hat{q} - \hat{q}_P|$, and to terms of order ϖ^2 , we can decompose the unit vector $\hat{q} \approx \hat{q}_P + \varpi$ into parallel and transverse pieces. This representation is an equal area projection of the sphere onto a disk in the sense that a solid angle element $d\Omega = \sin\theta d\theta d\phi$ is just $\varpi d\varpi d\phi = d^2\varpi$. However, only for $\varpi < 1$ does the map look good: as one goes into the opposite hemisphere, the distortions are severe. Note that the opposite pole to the one we are expanding about is the $\varpi = 2$ circle.¹ To evaluate the angular power spectrum, we make use of the property that in the limit of large ℓ and small ϖ ,

$$P_\ell(\cos\theta) \approx J_0((\ell + 1/2)\varpi). \quad (89)$$

We therefore have

$$C_\ell \approx \left\langle \left| \frac{\widetilde{\Delta T}}{T}(\mathbf{Q}) \right|^2 \right\rangle, \quad \text{if } Q = \ell + 1/2. \quad (90)$$

This suggests that the analogue of the power per logarithmic wavenumber interval is actually $(\ell + 1/2)^2 C_\ell / (2\pi)$. The form $\mathcal{C}_\ell \equiv \ell(\ell + 1)C_\ell / (2\pi)$ adopted differs by only 4% for $\ell = 2$ and by less than a percent for $\ell > 4$. Since the dimensions of Q are inverse radians, the ℓ -pole can be considered to probe angles around $3438/(\ell + 1/2)$ arcminutes (and angular wavelengths 2π bigger).

4.2 Experimental arrangements and their filters

4.2.1 Pixel-pixel correlation filters

We now discuss anisotropy experiments in more detail. Typically we are given the data in the form of measurements $(\Delta T/T)_p \pm \sigma_{Dp}$ of the anisotropy in the p th pixel, where σ_{Dp} is the variance about the mean for the measurements. In

¹We can choose to zero $\Delta T/T$ at this circle; the expansion is then a Fourier-Bessel series with cylindrical eigenfunctions $\propto e^{im\phi} J_m(Q_{nm}\varpi)$, where the Q_{nm} are the positive roots of $J_m(2Q_{nm}) = 0$, thus with a discrete spectrum, though not useful like the $Y_{\ell,m}$ expansion unless we are interested in small enough angles so that $\varpi = 2$ can be considered to be infinity and Q_{nm} becomes continuous.

general, there may be pixel–pixel correlations in the noise, defining a correlation matrix $C_{Dpp'}$ with off-diagonal components as well as the diagonal σ_{Dp}^2 . Also there is usually more than one frequency channel, with the generalized pixels having frequency as well as spatial designations. The signal $(\Delta T/T)_p$ can be expressed in terms of linear filters $\mathcal{F}_{p,\ell m}$ acting on the $a_{\ell m}$:

$$(\Delta T/T)_p = \sum_{\ell m} \mathcal{F}_{p,\ell m} a_{\ell m}. \quad (91)$$

The $\mathcal{F}_{p,\ell m}$ encode the experimental beam and the switching strategy that defines the temperature difference, the former filtering high ℓ , the latter low ℓ . They can also encode the frequency dependence if the signal has a fixed spectral signature, as primary CMB and secondary SZ fluctuations do. Reality implies $\mathcal{F}_{p,\ell,-m} = \mathcal{F}_{p,\ell m}^*$. The pixel–pixel correlation function of the temperature differences can be expressed in terms of a quadratic $N_{pix} \times N_{pix}$ filter matrix $W_{pp',\ell}$ acting on $\mathcal{C}_{T\ell}$:

$$C_{Tpp'} \equiv \left\langle \left(\frac{\Delta T}{T} \right)_p \left(\frac{\Delta T}{T} \right)_{p'} \right\rangle = \mathcal{I}[W_{pp',\ell} \mathcal{C}_{T\ell}], \quad (92)$$

$$W_{pp',\ell} \equiv \frac{4\pi}{2\ell+1} \sum_m \mathcal{F}_{p,\ell m} \mathcal{F}_{p',\ell m}^*; \quad \overline{W}_\ell \equiv \frac{1}{N_{pix}} \sum_{p=1}^{N_{pix}} W_{pp,\ell}. \quad (93)$$

The trace $\overline{W}_\ell$ defines the average filters [3, 42, 243] shown in fig. 6, which determine the rms anisotropies $\sigma_T[\overline{W}]$:

$$\sigma_T^2[\overline{W}] \equiv \left(\frac{\Delta T}{T} \right)_{rms}^2 \equiv \frac{1}{N_{pix}} \sum_{p=1}^{N_{pix}} C_{Tpp} \equiv \mathcal{I}[\overline{W}_\ell \mathcal{C}_{T\ell}]. \quad (94)$$

We define the *band-power* associated with the filter $\overline{W}_\ell$ to be the average power across the filter [5, 89]:

$$\langle \mathcal{C}_\ell \rangle_W \equiv \mathcal{I}[\overline{W}_\ell \mathcal{C}_{T\ell}] / \mathcal{I}[\overline{W}_\ell]. \quad (95)$$

Usually the band-power is the quantity that can be most accurately determined from the experimental data, and it is used extensively in what follows to assess what various experiments have measured, and what various theories predict.

In the high ℓ limit it is often more convenient to use the Fourier transform representation:

$$(\Delta T/T)_p = \sum_{\mathbf{Q}} \widetilde{\mathcal{F}}_p(\mathbf{Q}) \frac{\widetilde{\Delta T}}{T}(\mathbf{Q}), \quad \widetilde{\mathcal{F}}_p(\mathbf{Q}) = e^{i\mathbf{Q} \cdot \varpi_p} \mathcal{B}(\mathbf{Q}) U_p(\mathbf{Q}), \quad (96)$$

where ϖ_p is the position defining the pixel, $\mathcal{B}(\mathbf{Q})$ defines the beam profile, and $U_p(\mathbf{Q})$ contains details of the switching strategy. The associated filter for $C_{Tpp'}$

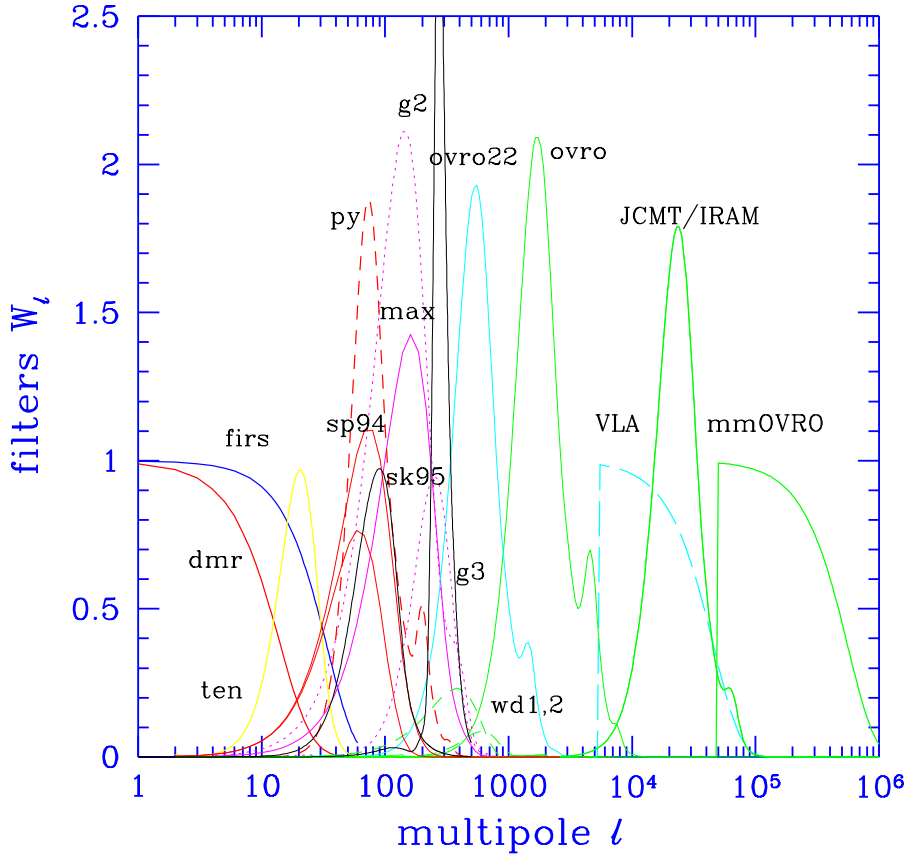


Figure 6: Filter functions for some current experiments: coBE's *dmr* and *firs* are treated as single-beam maps; *ten* is Tenerife; *sp94* is the UCSB 1994 South Pole experiment, for which two filters for different HEMT receiver systems are shown; *sk95* is the 1993-95 BigPlate (Saskatchewan) experiment, which is sensitive to a large range in ℓ , and for which two filters at $\ell \sim 100$ and $\ell \sim 300$ are shown; *py* is Python; *g2*, *g3* are 2 and 3 beam configurations for MSAM; *max* is for the MAX3,4,5 experiments; *wd1,2* are the $m = 1, 2$ analysis modes for the WhiteDish experiment. *ovro* is the filter for the 1987 OVRO experiment using a 40-m radio dish, *ovro22* uses a 5-m single dish, *JCMT/IRAM* illustrates what bolometer arrays on submillimeter telescopes are sensitive to, and *mmOVRO* and *VLA* denote approximations to the ℓ -space probed by a mm interferometer array and by the Very Large Array in a compact configuration.

is

$$\begin{aligned}\widetilde{W}_{pp'}(Q) &= \int_0^{2\pi} \frac{d\phi_{\mathbf{Q}}}{2\pi} \widetilde{\mathcal{F}}_p(\mathbf{Q}) \widetilde{\mathcal{F}}_{p'}(\mathbf{Q})^* \\ &= \int_0^{2\pi} \frac{d\phi_{\mathbf{Q}}}{2\pi} e^{i\mathbf{Q} \cdot (\varpi_p - \varpi_{p'})} \mathcal{B}^2(\mathbf{Q}) U_p(\mathbf{Q}) U_{p'}^*(\mathbf{Q}).\end{aligned}\quad (97)$$

The decomposition of the filter into a Fourier phase factor associated with the pixel position, a beam function and a switching strategy function U_p (which can depend upon the pixel position too for some experiments) is generally useful for experiments on scales below a few degrees – provided distortions in the ϖ -representation over the region of the sky mapped are not large; if they are, it is better to work with the full spherical harmonic representation. The analogue for the spherical harmonic representation $\mathcal{F}_{p,\ell m}$ of pulling out the phase associated with ϖ_p is to pull out an overall $Y_{\ell m'}(\hat{q}_p)$, but with penalty that the switching factor is a function of m' as well as the ℓm and possibly the pixel position:

$$\mathcal{F}_{p,\ell m} = \sum_{m'} Y_{\ell m'}(\hat{q}_p) \mathcal{B}(\ell, m) U_{p,\ell m m'}.\quad (98)$$

Discretization into time bins and aspects of pixelization are encoded in the functions $U_{p,\ell m m'}$ or $U_p(\mathbf{Q})$.

4.2.2 Beams and *dmr* and *firs*

Experimental beams are characterized by a full width at half maximum θ_{fwhm} . Beams must be determined experimentally, typically by determining the pattern of a point source on the sky. Usually there is a nice monotonic fall-off from the central point to low levels of power. However, beams do have side lobes which experimenters suppress as much as possible. Also the beams are not always rotationally symmetric. Still, for many experiments a Gaussian as a function of angle is not a bad approximation. The beam would then also be Gaussian in multipole (Fourier transform) space,

$$\begin{aligned}\mathcal{B}(\ell|\ell_s) &\equiv \exp\left[-\frac{(\ell + \frac{1}{2})^2}{2(\ell_s + \frac{1}{2})^2}\right], \\ \ell_s + \frac{1}{2} &= \frac{1}{2\sin(\theta_s/2)}, \quad \theta_s = \frac{\theta_{fwhm}}{\sqrt{8\ln 2}} \approx 0.425\theta_{fwhm}.\end{aligned}\quad (99)$$

The square of the COBE beam is shown in fig. 6: it falls off more rapidly than the rough 7° fwhm Gaussian used [90] before it was precisely determined [91].

One can imagine a “one-beam” experiment, with the temperature fluctuation relative to an absolute temperature being determined. In this case, the average filter is just $\overline{W}_\ell \approx \mathcal{B}^2(\ell|\ell_s)$. But this is never the case in practice although the processed COBE and *firs* maps can be analyzed as if they were one-beam experiments. COBE actually measures the difference between ΔT

values at two beam-smeared points 60° apart, but as the satellite spins and rotates, the entire sky is covered, albeit with different integration times for different sections of the sky. The set of dmr measurements give the difference in ΔT from a given beam-smeared point to enough connected points 60° away to allow a successful inversion and construction of a map: i.e., beam-smeared $\Delta T(\hat{q})$ values at 6144 pixels for each of the 2×3 frequency channels (using a convenient oversampled digitization in squares of size 2.6° of each beam-smeared point). The price one pays is that residual correlation in the experimental variance occurs between map pixels separated by 60° [92]. The gain is that the COBE maps can be thought of as giving $\Delta T(\theta, \phi)$ directly, smoothed with a “single-beam” high- ℓ filter associated with the beam-size. Of course, the monopole and dipole components are also filtered out: the $\ell = 0$ component, the average temperature on the sky, obviously is inaccessible; and because of the large dipole anisotropy induced by the motion of the earth relative to the cosmic background radiation, the “intrinsic” $\ell = 1$ component is also inaccessible.

The coverage of the *firs* experiment is more complicated than for *dmr* because it was a balloon experiment taking useful data for only about 5 hours. Nonetheless, a map with highly inhomogeneous weighting of each of the 1.3° pixels can be constructed for each of its four frequency channels. Although one may be more sophisticated in taking this into account in the construction of $\overline{W}_\ell$, it is reasonable to characterize the experiment by a one-beam filter function (99), with $\ell_s \approx 34$ corresponding to the 3.9° fwhm beam.

4.2.3 2-Beams, 3-beams, oscillating beams, ...

For a given theory, experiments could be designed to get the optimal signal. For example, MAX, MSAM and other half-degree experiments probe multipole ranges which optimize the signal from power spectra like that for primary anisotropies if the recombination of the primeval plasma occurred normally. A filter with a beam like that of *sp94* is better for probing primary anisotropies if early reionization occurred. It is of course best to get information from experiments probing the entire ℓ range, and thus the emphasis on large scale mapping experiments for the future. I now describe the U_p and $\overline{W}_\ell$ for a variety of current experimental configurations to give a flavor for what goes into fig. 6. Versions are also shown and discussed in [3, 5, 42, 89, 106, 140, 243].

A single-differencing (or 2-beam) experiment subtracts the temperature of the points on either side separated by $\theta_{throw}/2$ from the central point (the pixel label). Let us denote the separation direction by $\hat{\omega}_{throw}$. For a pixel at ϖ_p we have:

$$\Delta_p = \frac{\Delta T}{T}(\varpi_p + \frac{1}{2}\varpi_{throw}; \varpi_s) - \frac{\Delta T}{T}(\varpi_0 - \frac{1}{2}\varpi_{throw}; \varpi_s)$$

hence $U_p(\mathbf{Q}) = 2i \sin(\mathbf{Q} \cdot \varpi_{throw}/2).$ (100)

The filter is simply expressed in terms of the J_0 Bessel function:

$$\overline{W}_\ell = [2(1 - J_0(x_t))] \mathcal{B}^2(\ell|\ell_s), \quad \text{where } x_t = \frac{\ell + \frac{1}{2}}{\ell_{throw} + \frac{1}{2}}. \quad (101)$$

The $J_0(x_t)$ term is really the high ℓ approximation to $P_\ell(\cos \theta_{throw})$. $\overline{W}_\ell$ rises like $\ell^2 \mathcal{B}^2$.

In a double-difference (3-beam) experiment, the smoothed fluctuation at the pixel site has subtracted from it the average of the fluctuations at a distance θ_{throw} away. Thus

$$\begin{aligned} \Delta_p &= \frac{\Delta T}{T}(\varpi_p; \varpi_s) \\ &\quad - \frac{1}{2} \left\{ \frac{\Delta T}{T}(\varpi_p + \varpi_{throw}; \varpi_s) + \frac{\Delta T}{T}(\varpi_0 - \varpi_{throw}; \varpi_s) \right\}, \\ \text{hence } U_p(\mathbf{Q}) &= 2 \sin^2(\mathbf{Q} \cdot \varpi_{throw}/2), \end{aligned} \quad (102)$$

with the average filter

$$\overline{W}_\ell = [2(1 - J_0(x_t)) - \frac{1}{2}(1 - J_0(2x_t))] \mathcal{B}^2(\ell|\ell_s). \quad (103)$$

$\overline{W}_\ell$ rises like $\ell^4 \mathcal{B}^2$.

In the MSAM ($g2$, $g3$) experiment, the raw data was projected into both a 2-beam and 3-beam mode, an example of a growing trend to adopt switching strategies in software rather than hardware. Often the experimental filters are more complicated than 2 or 3 beam ones and the associated matrix elements must be calculated precisely, taking into account the details of the pattern on the sky. For example, the *sp91* and *sp94* experiments are similar to single-differencing experiments, except that the beam oscillates about the pixel position in a direction $\hat{\varpi}_{osc}$ with an oscillation amplitude ϖ_{osc} , frequency ω , and time behavior $\varpi_p + \varpi_{osc} \sin(\omega t)$, with the temperature positively weighted on one side and negatively weighted on the other. The *sp91* beam was 1.5° with oscillation amplitude of $2.95/2$ degrees, roughly corresponding to a 2-beam throw of about 2° , not that much larger than the beam: thus beam and throw interference result in a relatively small width and maximum of $\overline{W}_\ell$. A similar story holds for *sp94*. For such an experiment,

$$U_p(\mathbf{Q}) = 2iH_0(\mathbf{Q} \cdot \varpi_{osc}), \quad \text{where } H_0(x) \equiv \frac{2}{\pi} \int_0^{\pi/2} d\theta \sin(x \cos(\theta)) \quad (104)$$

is called the Struve function of index zero [243].

The *sp89* experiment and the MAX balloon-borne bolometer experiment both had fwhm beams of $30'$ ($\ell_s \approx 269$) and also measured temperature differences via oscillating beams, with oscillation amplitude $\approx 1.4/2$ degrees. The filters differ because MAX used a sine weighting of the temperature to make the temperature difference rather than the plus/minus step function technique of *sp89* and *sp91*. For sine weighting, $U_p(\mathbf{Q}) = 2i \frac{\pi}{2} J_1(\mathbf{Q} \cdot \varpi_{osc})$.

The 1.8' beam Owens Valley (*ov7*) experiment [94] used a 40-m radio dish at 1.5 cm to observe 7 fields on the sky with a double-differencing (or 3-beam) experiment, which (basically) subtracted the average of the temperatures 7.15' to each side of a central point from the temperature at the central point. Thus $\ell_{throw} \approx 480$ corresponds to the “throw” angle $\theta_{throw} = 7'$ and $\ell_s \approx 2246$ is the beam’s filter scale. It is theoretically advantageous to have $\theta_{throw} \gg \theta_{beam}$ to get the maximum rms signal, but it is difficult to manage experimentally.

Python was a 4-beam experiment with beam 45' and throw 2.75°, hence also a very large ratio of throw to beam; for it,

$$U_p(\mathbf{Q}) = 2i \sin^3(\tfrac{1}{2}\mathbf{Q} \cdot \varpi_{throw}),$$

$$\overline{W}_\ell = \tfrac{1}{4}[5 - \tfrac{15}{2}J_0(x_t) + 3J_0(2x_t) - \tfrac{1}{2}J_0(3x_t)]\mathcal{B}^2(\ell|\ell_s).$$

The raw data $\Delta_p(\phi)$ for the WhiteDish *wd* experiment (beam 12') are differences in the temperature of points on a circle of radius $\varpi_{throw}/2$ (= 14') centered on the pixel and the pixel. Among ways to analyze the data, the most straightforward is to form angular moments, $\int e^{-im\phi} \Delta_p(\phi) d\phi/(2\pi)$:

$$U_p^m(\mathbf{Q}) = i^m J_m(\tfrac{1}{2}Q\varpi_{throw}) (\hat{Q}_x - i\hat{Q}_y)^m, \quad \overline{W}_\ell^m = J_m^2(\tfrac{1}{2}Q\varpi_{throw})\mathcal{B}^2. \quad (105)$$

Data was given for the $m = 1, 2$ modes, derived from 5 pixels in a line.

As we move into the next generation of experiments, the goal is to make maps of extended regions. An example of the increasing sophistication is provided by the *sk95* experiment [150], which projected from 3-beam to 19-beam configurations in software, leading to an interpretation of an even more generalized pixel-space than one just including frequencies and spatial centering. The filter functions can also be designed after the fact with this approach, as in fact was done for broad-band power spectrum analysis in [150]. Two such filters at either end of the ℓ -range that *sk95* was sensitive to are shown in fig. 6.

4.3 Primary power spectra for inflation-based theories

Sample theoretical $\mathcal{C}_\ell$ ’s are shown in figs. 7, 8 for a number of inflation-inspired theories with modest variations in cosmological parameters [144, 261, 305]. The “standard” scale invariant adiabatic CDM model ($\Omega = 1$, $n_s = 1$, $h = 0.5$, $\Omega_B = 0.05$) with normal recombination shown in fig. 7 and repeated in each of the panels of fig. 8 illustrates the typical form: the Sachs–Wolfe effect dominating at low ℓ , followed by rises and falls in the first and subsequent “Doppler peaks”, with an overall decline due to destructive interference across the photon decoupling surface and damping by shear viscosity in the photon plus baryon fluid. A similar CDM model, but with early reionization (at $z > 200$), shows the Doppler peaks are damped, a result of destructive interference from forward and backward flows across the decoupling region, illustrating that the “short-wavelength” part of the density power spectrum can have a dramatic effect upon $\mathcal{C}_\ell$, since it determines how copious UV production from

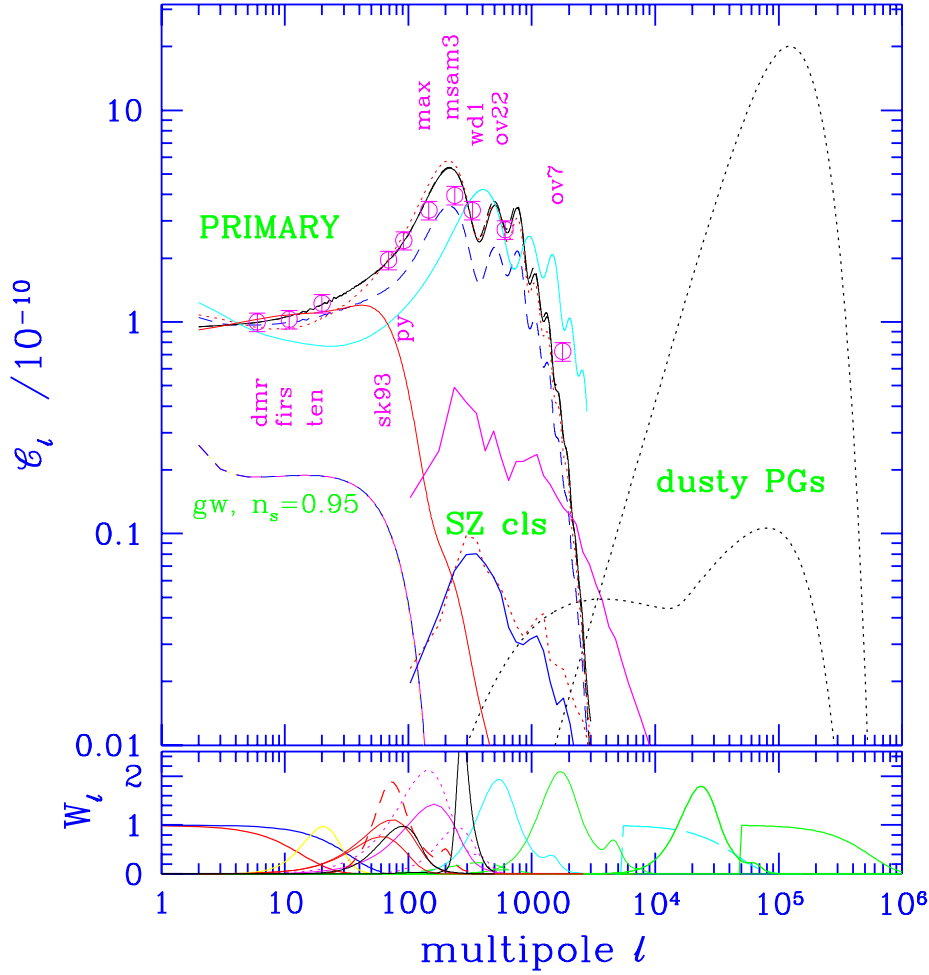


Figure 7: Temperature power spectra normalized to $\langle \mathcal{C}_\ell \rangle_{dmr} = 10^{-10}$: for a standard $n_s=1$ CDM model with standard recombination, early reionization, a (dashed) tilted primordial spectrum with $n_s = 0.95$, with the gravity wave contribution shown, a (dotted) $H_0 = 75$ model with $\Lambda \neq 0$, and an open $H_0 = 60$ CDM model (with the peaks shifted to larger ℓ). Band-powers with 10% (*dmr*-level) error bars (for selected experimental configurations) are shown for the tilted and untilted CDM models. A hot/cold hybrid model power spectrum with $\Omega_\nu = 0.2$ is plotted as well but is indistinguishable here from the standard CDM case. The power spectra of SZ maps constructed using the peak-patch method (Bond and Myers 1996) are shown for a $\sigma_8 = 1$ standard CDM model, a hot/cold hybrid model ($\Omega_\nu = 0.3$) with $\sigma_8 = 0.7$ (a tilted CDM model with $n_s = 0.8$ and $\sigma_8 = 0.7$ is also shown). Spectra for a BCH2 dust model (13) is also shown, the larger (arbitrarily normalized) part a shot-noise effect for galaxies with dust distributed over 10 kpc, the smaller a continuous clustering contribution, including a nonlinear correction. The $\sim \ell^2$ shot-noise rise also characterizes the power spectrum for extragalactic radio sources. On the other hand, Galactic foregrounds have power spectra falling $\sim \ell^{-1}$ with ℓ . Average filter functions for a variety of experiments are shown in the lower panel.

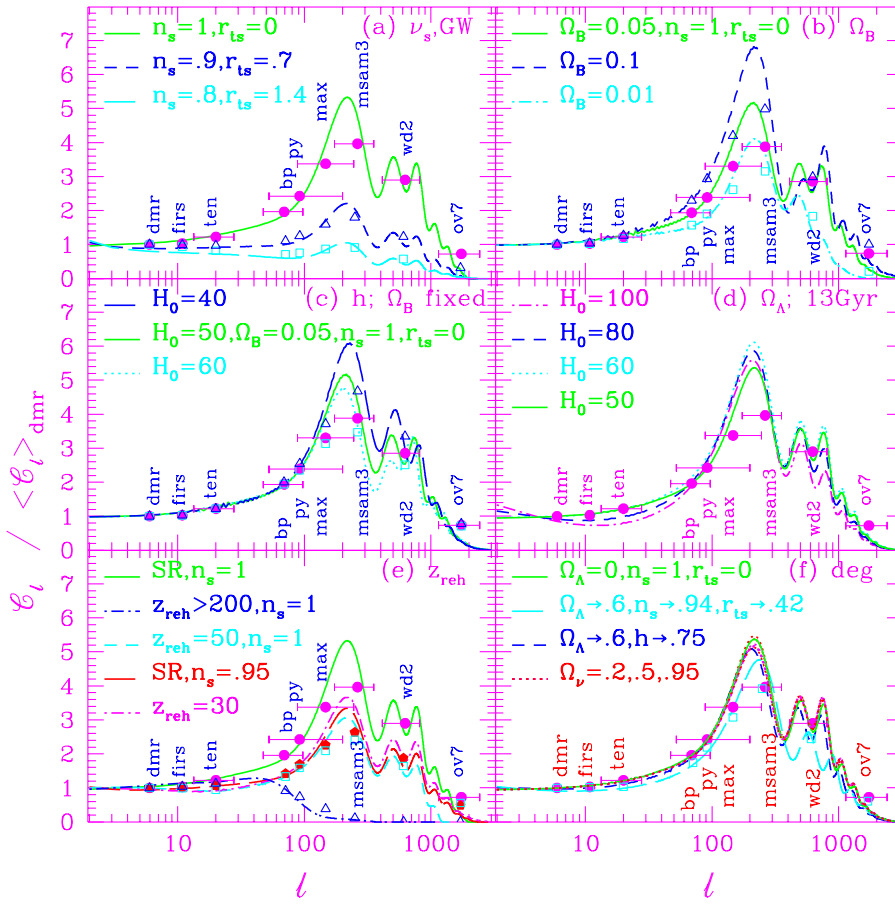


Figure 8: Spectra for a variety of inflation-inspired models, normalized to the COBE band-power. Theoretical band-powers for various experimental configurations are placed at $\langle \ell \rangle_W$, horizontal error bars extend to the $e^{-1/2} \overline{W}_{max}$ points. Unless otherwise indicated, $\Omega_B h^2 = 0.0125$, $h = 0.5$, $n_s = 1$; when the gravity wave contribution is nonzero, $\nu_t = \nu_s$ and $r_{ts} \approx -7\nu_t$ are assumed ($r_{ts} \equiv \mathcal{C}_2^{(T)}/\mathcal{C}_2^{(S)}$). The untilted $n_s = 1$, $r_{ts} = 0$ model is repeated in each panel (solid line). (a) CDM models with variable tilt n_s . (b) $n_s = 1$ models with $\Omega_B h^2$ changed, h fixed. (c) $n_s = 1$ models, with $\Omega_B h^2$ changed, Ω_B fixed. (d) $n_s = 1$ models with fixed age, 13 Gyr, but variable H_0 and $\Omega_\Lambda = 1 - \Omega_{cdm} - \Omega_B$ (.92,.79,.43,0 for 100,80,60,50). (e) CDM models with very early reionization at $z_{reh} \gtrsim 150$ (equivalent to “no recombination”), and later reionization at $z_{reh} = 30, 50$ are contrasted with standard recombination (SR). The $z_{reh} = 50$ spectrum is close to the $n_s = 0.95$ spectrum with SR (thin, dot-dashed): the moderate suppression if $20 \lesssim z_{reh} \lesssim 150$ can be partially mimicked by decreasing n_s or increasing h . (f) Sample cosmologies with nearly degenerate spectra and band-powers. Dashed curve: increasing Ω_Λ is compensated by increasing h . Dot-dashed curve: tilting to $n_s = 0.94$ ($\tilde{r}_{ts} = 0.42$) is compensated by increasing Ω_Λ to 0.6. The dotted hot/cold model curves (Bond and Lithwick 1995) (with Ω_ν indicated) are nearly identical to the standard CDM one, but even these few percent differences can be distinguished in principle by satellite all-sky experiments with currently available detector technology.

early stars was. Lower redshifts of reionization still maintain a Doppler peak, but suppressed relative to the standard CDM case (as illustrated in fig. 8(e)).

The primary spectra are calculated by solving for each mode $\mathcal{M}$ the linearized Boltzmann transport equation for photons (including polarization) and light neutrinos, coupled to the equations of motion for baryons and cold dark matter, and to the perturbed gravitational metric equations (section 6).

If the post-inflation fluctuations are Gaussian-distributed, then so are the multipole coefficients $a_{\ell m}^{(\mathcal{M})}$, with amplitudes fully determined by just the angular power spectra $\mathcal{C}_\ell^{(\mathcal{M})}$. Figures 7, 8 include adiabatic scalar and tensor contributions. The relative magnitude of each is characterized by either the ratio of the quadrupole powers, $r_{ts} = \mathcal{C}_2^{(T)}/\mathcal{C}_2^{(S)}$, or the ratio of the dmr band-powers $\tilde{r}_{ts} = \langle \mathcal{C}_\ell^{(T)} \rangle_{dmr} / \langle \mathcal{C}_\ell^{(S)} \rangle_{dmr}$. For the scale invariant cases, r_{ts} is taken to vanish.

A simple variant of CDM-like models is to tilt the initial spectrum. We deal with the physics of tilt in more detail in section 6, and just sketch the main results here. The scalar tilt is defined by $\nu_s = n_s - 1$, in terms of the usual primordial index for density fluctuations, n_s , which is one for scale invariant adiabatic fluctuations. There is a corresponding tilt which characterizes the initial spectrum of gravitational waves which induce primary tensor anisotropies, ν_t . Inflation models give $\nu_t < 0$ and usually give $\nu_s < 0$. For small tensor tilts, $r_{ts} \approx -6.9\nu_t$ and $r_{ts} \approx 1.3\tilde{r}_{ts}$ are expected (with corrections given by eq. 184). For a reasonably large class of inflation models $\nu_t \approx \nu_s$, but in some popular inflation models ν_t may be nearly zero even though ν_s is not. Figures 7 and 8(a) show $\mathcal{C}_\ell^{(S)} + \mathcal{C}_\ell^{(T)}$ derived for tilted cases when $\nu_t \approx \nu_s$ is assumed to hold. Figure 7 also shows the contribution that $\mathcal{C}_\ell^{(T)}$ makes to the total; $\mathcal{C}_\ell^{(T)}$ for both the standard and early reionization cases are actually both shown; they cannot be distinguished on this graph.

Spectra for hot/cold hybrid models with a light massive neutrino look quite similar to those for CDM only [260, 262, 261], as fig. 7 and 8(f) show, with small differences appearing at higher ℓ . This is also even true for pure hot dark matter models [134] because the scale associated with neutrino damping is near to the scale associated with the width of decoupling.

The dotted $\mathcal{C}_\ell$ in fig. 7 also has a flat initial spectrum, but has a large nonzero cosmological constant in order to have a high H_0 , in better accord with most observational determinations ($\sim 65 - 85$) e.g., [10]. The specific model has the same age (13 Gyr) as the standard CDM model, and $\Omega_\Lambda = 0.734$, $\Omega_{cdm} = 0.243$, $\Omega_B = 0.022$, $H_0 = 75$, $n_s = 1$. (The best current estimate for globular cluster ages, along with one and two sigma error estimates is $14.6^{+1.7;3.7}_{-1.6;3.0}$ Gyr [111].) Other nonzero Λ examples with this age are given in fig. 8(d). As one goes from $\ell = 2$ to $\ell = 3$ and above there is first a drop in $\mathcal{C}_\ell$ [110], a consequence of the time dependence of the gravitational potential fluctuations Φ_N (see fig. 23 for a closeup of this).

The model whose peak is shifted to high ℓ is an open CDM cosmology [305] with the same 13 Gyr age, but now $H_0 = 60$, and $\Omega_{tot} = 0.33$ (and $\Omega_{cdm} = 0.30$,

$\Omega_B = 0.035$). By $H_0 = 70$, Ω_{tot} is down to 0.055 at this age. The shift to higher ℓ for open models is a simple consequence of the cosmological angle-distance relation (section 5.1.4, eq. (130)); for closed models, the shift is to smaller ℓ .

To get a visual impression of what the spectral structure means, fig. 9 shows what the sky looks like on a few resolution scales for the standard $n_s = 1$ CDM model: on the COBE beamscale (Gaussian filtering $\ell_s = 19$ here, see also fig. 11), the nearly scale invariant form; on the half-degree scale ($\ell_s = 269$ here), where the standard recombination spectrum is a maximum; with no smoothing at all, with the shapes defined entirely by the destructive interference that occurred across the photon decoupling region. For early-reionization, the shapes in the 60° NR map are also the naturally occurring ones, since there is no power left at $\ell_s \sim 269$ to artificially filter.

4.4 2D spectra with tilt and a Gaussian coherence angle

A phenomenology characterized by three parameters, a broad-band power $\langle \mathcal{C}_\ell \rangle_W$, a broad-band tilt $\nu_{\Delta T}$, and a Gaussian coherence scale θ_c is often a good local approximation to $\mathcal{C}_\ell$:

$$\mathcal{C}_\ell = \langle \mathcal{C}_\ell \rangle_W \frac{Q^{\nu_{\Delta T}} e^{-\frac{1}{2} Q^2 \varpi_c^2} \mathcal{I}[\overline{W}_\ell]}{\mathcal{I}[\overline{W}_\ell Q^{\nu_{\Delta T}} e^{-\frac{1}{2} Q^2 \varpi_c^2}]}, \quad Q \equiv \ell + \frac{1}{2}, \quad \varpi_c \equiv 2 \sin(\frac{1}{2} \theta_c). \quad (106)$$

Instead of $Q^{\nu_{\Delta T}}$, it has become standard to use a form $\mathcal{U}_\ell$ which, as is shown in section 5.1, arises when the anisotropy is generated by emission from a thin shell at cosmological distance of sources described by a 3D Gaussian random field with power spectrum $\mathcal{P}(k) \propto k^{\nu_{\Delta T}}$:

$$\mathcal{U}_\ell \equiv \frac{\Gamma(\ell + \nu_{\Delta T}/2) \Gamma(\ell + 2)}{\Gamma(\ell) \Gamma(\ell + 2 - \nu_{\Delta T}/2)} = Q^{\nu_{\Delta T}} (1 + \mathcal{O}(Q^{-2})). \quad (107)$$

The band-power $\langle \mathcal{C}_\ell \rangle_W$ derived for an experiment is estimated using one or more of these functional forms, and is often quite insensitive to $\nu_{\Delta T}$ or to θ_c . Two special cases are usually analyzed:

(1) The pure power law case has zero coherence scale and $\nu_{\Delta T}$ variable. $\nu_{\Delta T}$ nearly 0 corresponds to scale invariant in ΔT ; $\nu_{\Delta T} \approx 0.15$ is an effective index appropriate for COBE-scale anisotropies for a standard initially scale-invariant CDM model with $\Omega = 1$, $h = 0.5$ and 5% baryon content [144, 89], and varying $\nu_{\Delta T}$ can model changing the primordial tilt of the spectrum: it is relatively insensitive to h and Ω_B changes for low ℓ , but see fig. 23.

(2) The other case has a coherence scale and a white noise spectral index $\nu_{\Delta T} = 2$, a Gaussian correlation function model with $C(\varpi) = (\Delta T/T)_c^2 \exp(-\frac{1}{2} \varpi^2 / \varpi_c^2)$. It describes uncorrelated blobs of size $\sim \theta_c$, and is similar to the spectrum for a shot noise distribution of blobs with Gaussian profiles (section 5.3.1, also see fig. 7) and so is a reasonable form to try. Note that

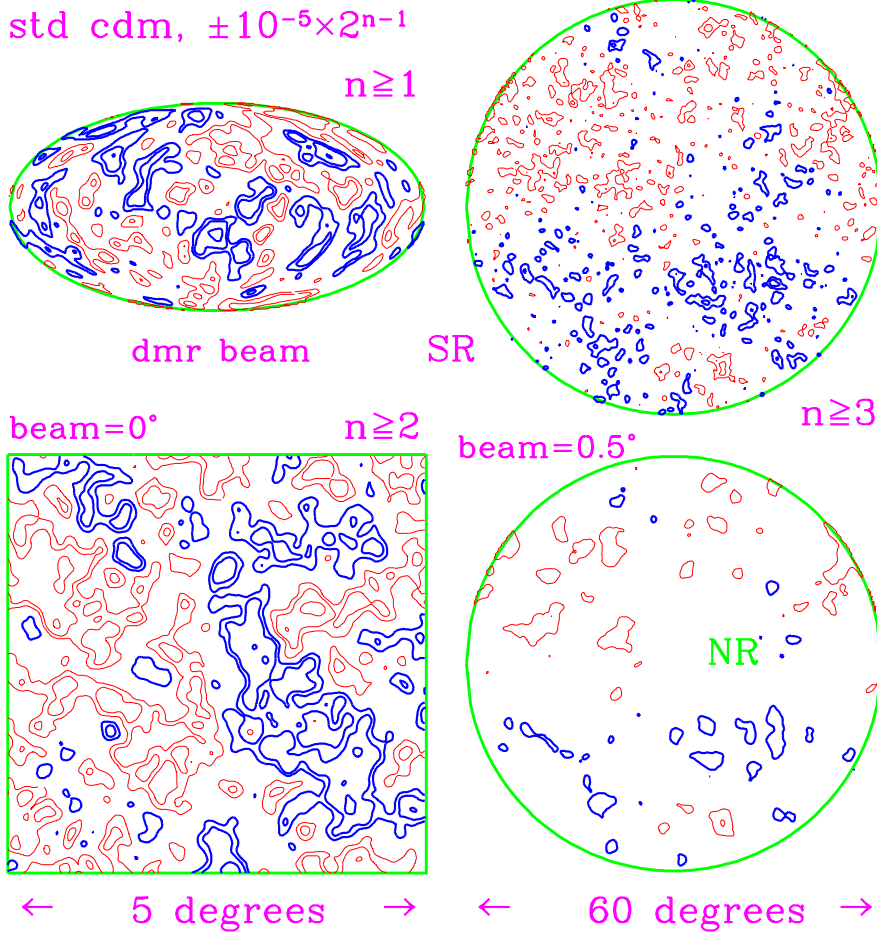


Figure 9: How a CDM model normalized to COBE varies with resolution. The contours begin at $109 \mu\text{K}$ in the half-degree smoothing cases, $54.5 \mu\text{K}$ in the no-smoothing case, $27.3 \mu\text{K}$ in the all-sky aitolff projection map. Positive contours are heavy, negative are light. SR denotes standard recombination, NR denotes very early reionization, so there is no Doppler peak. The hills and valleys in the 5° SR (60° NR) map are naturally smooth: mapping them will give a direct probe of the physics of how the photon decoupling region at redshift ~ 1000 (200) damped the primary signal.

the form $\sim \ell^2 \exp[-\ell^2(\varpi_c^2/2 + \varpi_s^2)]$ when the beam smoothing is included is not very dissimilar from the form of the $\overline{W}_\ell$ of a 2-beam experiment, again not unreasonable. However, the $(\Delta T/T)_c$ versus θ_c plots that became the standard way of representing experimental data until the band-power representation was developed are somewhat misleading: even if there is no information at all in the experiment on the shape of the spectrum, $(\Delta T/T)_c$ will have a minimum at a scale corresponding to where the filter function has its maximum; by contrast, the band-power is largely θ_c -independent unless one has a mapping experiment with very broad filters; i.e., is sensitive to the shape.

4.5 Experimental band-powers: past and present

The story of the experimental quest for anisotropies is a heroic one. The original Penzias and Wilson (1965) discovery paper quoted angular anisotropies below 10%, but by the late sixties 10^{-3} limits were reached [123, 124]. As calculations of baryon-dominated adiabatic and isocurvature models improved in the seventies, with the seminal work of Peebles and Yu [130], Doroshkevich, Sunyaev and Zeldovich [131] and Wilson and Silk [132, 133], the theoretical expectation was that the experimentalists just had to get to 10^{-4} . And they did march from 10^{-3} down to 10^{-4} in the seventies, with results from Boynton and Partridge [125], among others. The only signal that was found was the dipole, hinted at by Conklin and Bracewell in 1973, but found definitively in Berkeley and Princeton balloon experiments. Throughout the 1980s, the upper limits kept coming down, punctuated by a few experiments widely used by theorists to constrain models: the Uson and Wilkinson [126] limit of 5×10^{-5} with a $4.5'$ switch (and therefore not much primary signal); the 1987 OVRO limit [94] of 2×10^{-5} on $7'$ scales (also below the coherence scale); the 6° limit of Melchiorri's group [127]; early versions of the Tenerife experiment [129]; the 7° -beam Relict 1 satellite limit [128]; Lubin and Meinhold's 1989 half-degree South Pole limit [95], on an angular scale which was optimal for testing CDM-like models.

These data were used to rule out adiabatic baryon-dominated models, but by then the dark matter dominated universes had come to the rescue to lower the theoretical predictions by about an order of magnitude [134, 135]. Many groups developed codes to solve the perturbed Boltzmann–Einstein equations when dark matter was present e.g., [134, 135, 215, 136, 137, 138, 139], and, post-COBE [140, 141, 142, 143, 159, 302, 306]. With the results of the pre-COBE computations, a number of otherwise interesting models fell victim to the data: scale invariant isocurvature cold dark matter models [215]; large regions of parameter space for isocurvature baryon models [216, 218]; many broken scale invariant inflation models with enhanced power on large scales [233, 192]; CDM models with a decaying ($\sim \text{keV}$) neutrino if its lifetime was too long ($\gtrsim 10\text{yr}$) [233, 252]; constraints on Ω_B , Ω and Λ in CDM models [243, 136]. For all of these the strategy was to normalize the anisotropy predictions using the clustering properties predicted by the model, in particular by σ_8 .

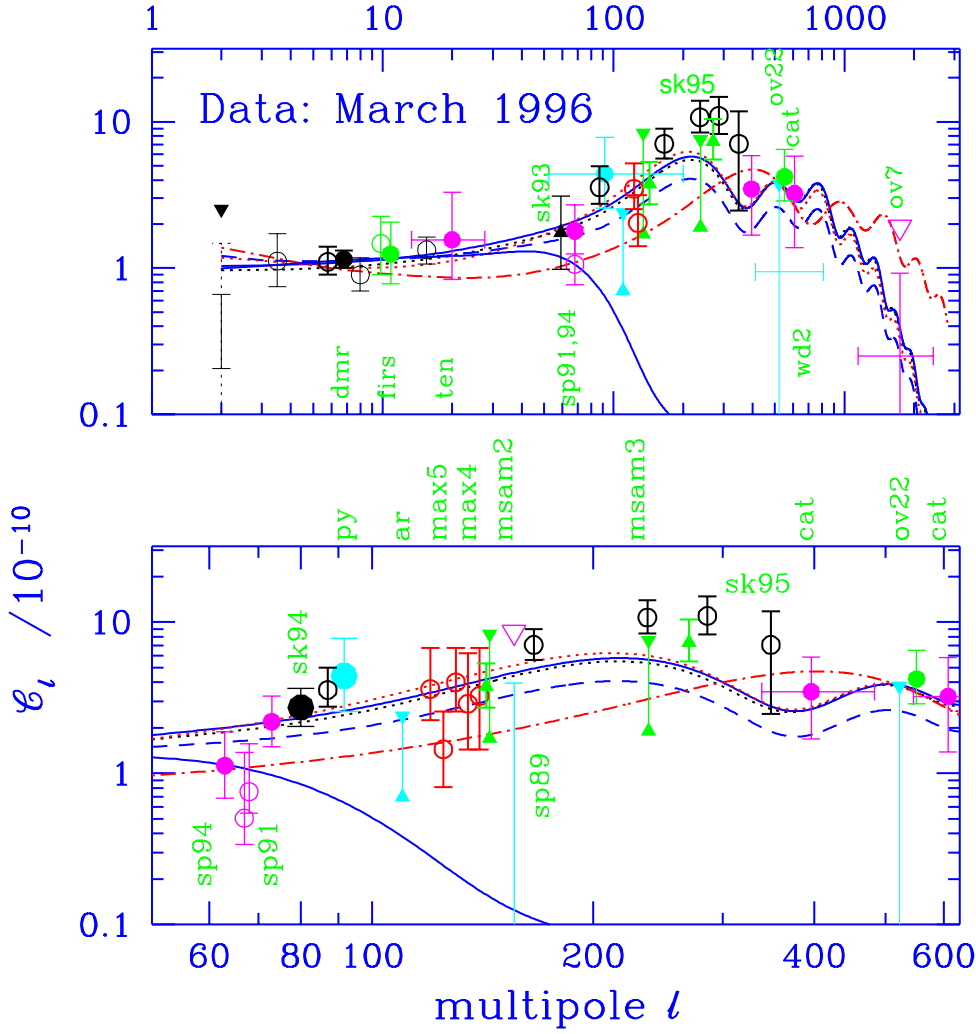


Figure 10: Band-power estimates derived for the anisotropy data up to March 1996. The lower panel is a closeup of the first “Doppler peak” region. The theoretical curves are those of the filter figure, normalized to the *dmr4* data: the standard SR CDM model, the nearly degenerate one with $\Omega_\nu = 0.2$ in light massive neutrinos, the NR CDM model, the $H_0 = 75$ vacuum-dominated model (upper), the slightly tilted CDM model with a gravity wave contribution, and the $H_0 = 60$ open model. All are for universes with age 13 Gyr. Although current data broadly follows the inflation-based expectations, the band-powers shown may have signals from systematic effects such as sidelobe contamination or Galactic effects such as bremsstrahlung and dust, as well as the cosmological primary and secondary anisotropy signals.

Now that we have detections, we normalize spectra to the COBE anisotropy level, and can now use the data to rule out theories from below as well as from above. In fig. 10, I use the band-power estimates with their error bars to give a snapshot of the current data at this time and use it as a vehicle for discussing the associated experiments. To determine band-powers for an experiment [5, 89], a local model of $\mathcal{C}_\ell$ is constructed, assumed to be valid over the scale of the experiment’s average filter $\bar{W}_\ell$. I usually choose eq. (106) with zero coherence angle. The once popular $\nu_{\Delta T}=2$ coherence angle form is rapidly disappearing from the scene, but it is also easy to transform such results into the band-power language. As we learn more, a shape that fits the data will be the preferred form [89].

Because there are so many detections now, fig. 10 is split into two panels for clarity, the upper giving the overview, the lower focussing on the crucial first Doppler peak region. Data points either denote the maximum likelihood values for the band-power and the error bars give the 16% and 84% Bayesian probability values (corresponding to $\pm 1\sigma$ if the probability distributions were Gaussian) or are my translations of the averages and errors given by the experimental groups to this language. Aspects of these statistical techniques are described in section 4.6. Upper and lower triangles denote 95% confidence limits unless otherwise stated. The horizontal location is at $\langle \ell \rangle_W$ and the horizontal error bars denote where the filters have fallen to $e^{-0.5}$ of the maximum (with fig. 6 providing a more complete representation of sensitivity as a function of ℓ). Only wavelength-independent Gaussian anisotropies in $\Delta T/T$ are assumed to be contributing to the signals, but nonprimary sources (e.g., dust, synchrotron) may contribute to these $\mathcal{C}_\ell$ ’s (as can unknown systematic errors of course). Either it has been shown that the frequency spectrum is compatible with the CMB and incompatible with expected contaminating foregrounds or some attempt at cleaning the observations of residual signals in almost all of these cases. With residual contaminants, one generally expects the underlying primary $\mathcal{C}_\ell$ to be lower than the values shown, but it can be higher because of “destructive interference” among component signals. In the following, considerable space is devoted to the *dmr* data since the definitive 4-year data set has been released and it is so important for normalizing spectra. (A closeup view of the large angle band-powers is given in fig. 23.)

The $\ell = 2$ power uses the 4-year quadrupole value [85, 86], determined from high Galactic latitude data. It is the multipole most likely to have a residual Galactic signal contaminating it, possibly destructively, and the “systematic” error, the dashed addition to the statistical error bar (solid), reflects this. In determining the *dmr* band-power, it is therefore wise to assume that in addition to any primordial signal, unknown monopole and dipoles, there is an unknown Galactic quadrupole contamination. Further, the Galactic plane should be cut out, and other regions of known large Galactic contamination should be removed as well. I used to take $|b| > 25^\circ$ as a safe cut [146], but have now adopted the $|b| > 20^\circ$ customized cut used by the *dmr* team in its analysis of the 4-year data [85], which also removes regions found to be high when

correlated with the 140 μm DIRBE map, in particular Ophiucus and Orion. The two heavy points at $\ell \sim 7$ are band-powers derived for the 4-year *dmr* 53+90+31 GHz “*A+B*” maps, using the methods of section 4.6 [146, 162], the solid point assuming a $\nu_{\Delta T} = 0$ spectrum, the open marginalizing over all possible $\nu_{\Delta T}$. (The latter is lower by 6% if a Galactic quadrupole component is not allowed for.) A good fit as a function of $\nu_{\Delta T}$ to the band-power is $\langle \mathcal{C}_\ell \rangle_{dmr}^{1/2} \approx [0.82 + 0.26(1 - \frac{\nu_{\Delta T}}{2})^{2.8}] \times 1_{-0.06}^{+0.07}$, which is reasonably insensitive to modest variations in $\nu_{\Delta T}$. Band-powers derived for the “*A+B*” maps as a function of frequency are in excellent agreement: the functional form with the 2.8 power fits well, but the coefficients and errors are slightly different: 53 GHz: 0.88, 0.24, $1_{-0.07}^{+0.08}$; 90 GHz: 0.95, 0.15, $1_{-0.10}^{+0.11}$; 31 GHz: 0.90, 0.08, $1_{-0.14}^{+0.16}$. The band-powers are also remarkably insensitive to making signal-to-noise cuts in the data (i.e., filtering it). The “*A-B*” maps are consistent with no signal.

DMR had enough coverage in ℓ -space that one can estimate the spectral index $1 + \nu_{\Delta T}$ from the data as well by Bayesian means (marginalizing over $\langle \mathcal{C}_\ell \rangle_{dmr}^{1/2}$): $1.07_{-0.29; .58}^{+.27; .52}$ for 53+90+31 GHz (agreeing with the *dmr* team’s result using this method [87] and also with the $1.05_{-0.28; .57}^{+.27; .52}$ I determine for the Ecliptic (as opposed to Galactic) coordinate versions of the maps); $0.97_{-0.32; .64}^{+.30; .57}$ for 53+90 GHz; $1.15_{-0.33; .66}^{+.31; .59}$ for 53 GHz; $1.27_{-0.43; .86}^{+.40; .77}$ for 90 GHz. (First errors are 1σ , second are 2σ .) Notice the preferred index actually goes down when the 53 GHz and 90 GHz are added. For the 31 GHz map, there is a residual bremsstrahlung contribution that lowers the index determination to 0.3, with large errors.

The approximate relation between $\nu_{\Delta T}$ and the primordial tilt ν_s for the standard CDM model over the *dmr* band is $\nu_{\Delta T} \approx 0.15 + \nu_s$, hence the values of $1 + \nu_{\Delta T}$ are quite compatible with the simplest inflation expectations, $0.6 \lesssim n_s \lesssim 1$. Indeed when standard CDM models with tilts ranging from 0.5 to 1.5 are considered, the index marginalized over σ_8 for the summed map is $n_s = 1.02_{-0.25; .52}^{+.23; .41}$ for the case with no gravity waves ($\nu_t = 0$), and $n_s = 1.02_{-0.18; .46}^{+.23; .40}$ with gravity waves ($\nu_t = \nu_s$). Band-powers for specific ℓ ranges also show the nearly flat character of $\mathcal{C}_\ell$, as the light open points at $\ell \sim 4, 8, 16$ from [87] show.

Optimally-filtered maps show the same large scale features independent of frequency. The 53+90+31 GHz version of this in fig. 11 shows the true sky anisotropy features as revealed by COBE (cleaned of experimental noise). It is compared with a realization of a scale invariant $\Omega = 1$ dark matter dominated model which has driven so much of the theory of the last decade. The basic lesson is that there is a tremendous consistency in the 4-year *dmr* data set, with the overall band-power being very well determined and the shape moderately well determined.

Many papers estimating $Q_{rms,PS}$ (i.e., $\mathcal{C}_2$), $\nu_{\Delta T}$, σ_8 , etc. have been written using the 1-year and 2-year *dmr* data e.g., [90, 84, 256, 257, 89, 6, 146], [288, 289, 290, 291, 292] which will all now be footnotes to history except for their importance in developing the statistical techniques that have been applied to the definitive 4-year *dmr* data. A variety of different measures such as corre-

lation functions and power spectra estimated by quadratic pixel combinations, multipole modes and S/N -modes using linear pixel combinations, have been used and there is good agreement among the methods [85] and with the results given here.²

The two points at $\ell \sim 10$ are for the 170 GHz *firs* map, solid with the restriction $\nu_{\Delta T} = 0$, open with $\nu_{\Delta T}$ allowed to float. A “Galactic quadrupole” as well as a model for residual “noise” that exists in the data [89, 6] were integrated over. The band-power $\langle \mathcal{C}_\ell \rangle_{firs}^{1/2} \approx 1.15 + 0.03(1 + \nu_{\Delta T})^2 \times 1_{-.23}^{+.25}$ is compatible with the COBE value. For example, for the $n_s = 1$ standard CDM model, $\sigma_8 = 1.27_{-.29, .57}^{+.30, .62}$, compared with the 53+90+31 GHz $A+B$ *dmr* map value of $\sigma_8 = 1.20_{-.08, .15}^{+.08, .17}$. This strengthens the case for a CMB origin, extending the 31-90 GHz band to 170 GHz. The *firs* team [93] also showed a significant cross-correlation with *dmr* exists. Although the *firs* coverage in ℓ -space is large enough that a spectral index can be estimated, the value I obtain for $1 + \nu_{\Delta T}$, $1.6_{-.8}^{+.7}$, has very large errors, and the small angle residual “noise” is probably driving the higher values [89]. Using a correlation function analysis, which filters some of the residual noise, the *firs* team [93, 147] derived similar amplitudes but a smaller $\nu_{\Delta T}$, although quite compatible within the large error bars.

The Tenerife point [103] at $\ell \sim 20$ uses combined 15 and 33 GHz data, agrees with the band-power for their data at 15 GHz only, which covered a much larger region of the sky, and is rather remarkable in view of the relatively low frequency (fig. 5). The Tenerife results have also been shown to strongly correlate with the DMR maps.

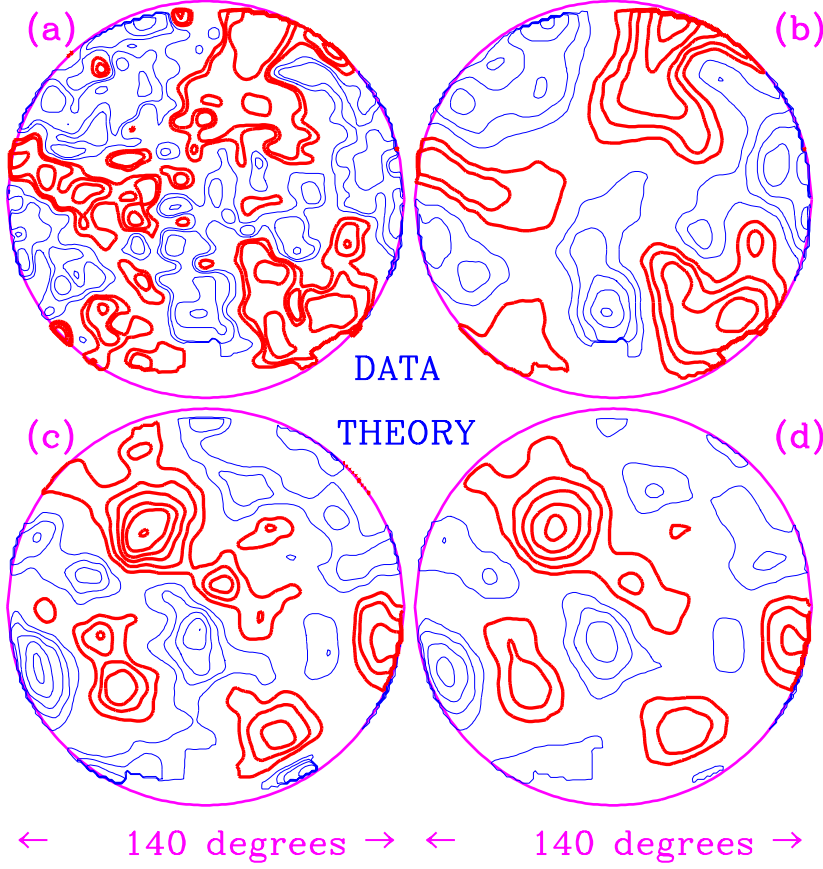
We now come to the crowded region from two degrees to half a degree. The lower open circle is from a joint 4-channel analysis of the 9 and 13 point *sp91* scans [96, 5, 89, 162] (with the individual 9 point and 13 point values given in the lower panel). The upper solid is for a simultaneous analysis of all channels of the *sp94* data [148, 162], with separate values for the Ka and Q bands (fig. 5) in the lower panel. The solid triangle in the upper panel is the *sk93* result; the big solid circle at $\ell \sim 80$ in the lower panel is the *sk93+94* result. The nearness of the *sp94*, *sk93* and *sk93+94* band-powers, and the demonstration for both experiments that the preferred frequency dependence is nearly flat in $\Delta T/T$ and many sigma away from bremsstrahlung or synchrotron, the expected contaminants in this 30-40 GHz range, lend confidence that the spectrum in the $\ell \sim 60$ –80 region has really been determined; and it looks quite compatible with the COBE-normalized CDM spectrum: *sp94* gives $\sigma_8 = 1.26_{-.27, .47}^{+.37, .85}$, and *sk93+94* gives $\sigma_8 = 1.21_{-.19, .39}^{+.24, .54}$ [162], very close to the *dmr* value given above:

²The 1-year $\langle \mathcal{C}_\ell \rangle_{dmr}^{1/2}$ and σ_8 value is rather close to the preferred 4-year estimate, e.g., only 3% lower for 53 GHz $A+B$ [89]. For this channel, the 2-year value is about 4% higher, but for 90 GHz $A+B$ it is 14%, leading to a 12% higher $\langle \mathcal{C}_\ell \rangle_{dmr}^{1/2}$ and a 14% higher σ_8 in the 53+90+31 GHz $A+B$ map. Much of this can be attributed to the customized cut: a straight $|b| > 20^\circ$ cut gives only a 4% discrepancy. The main lesson is that one should disregard the earlier values and only use the new ones with the customized cuts; further foreground corrections beyond this do not change the values by much [85].

4-yr COBE South Galactic Pole Map

UNFILTERED

OPTIMAL FILTER



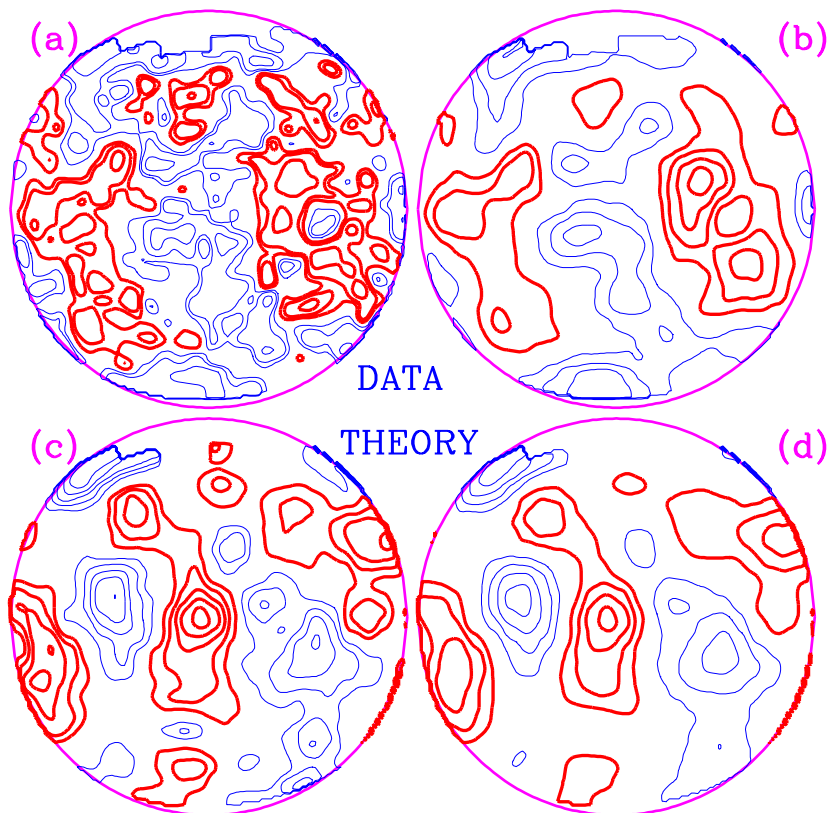
(b,c,d) contours: $\Delta T/T = \pm 0.55 \times 10^{-5} n$, $n \geq 1$
(a) contours: $\Delta T/T = \pm 0.55 \times 10^{-5} 2^{n-1}$, $n \geq 1$

Figure 11: 140° diameter maps centered on the South and North Galactic Poles are shown for a realization of a CDM C_ℓ -spectrum convolved with the *dmr* beam in (c). No noise has been added. This is how the primary sky would appear in a $n_s = 1$ CDM Universe with $\sigma_8 = 1.2$ (or in a $\Omega_{m\nu} = 0.2$ hot/cold universe with $\sigma_8 = 0.8$), the most likely values for the *dmr* data. This is contrasted with the 4-year *dmr* (53+90+31)a+b map shown in (a) and the map after the data has undergone optimal signal-to-noise filtering in (b) (using the same C_ℓ -shape and amplitude for the filter). The statistically significant features are also seen in *each* of the *dmr* channel maps after optimal filtering (which preferentially removes high angular frequencies, more so for noisier maps). Thus, to compare, (d) shows the theoretical realization after passing (c) through the same optimal filter used for (b); the average, dipole and quadrupole of the full $|b| > 20^\circ$ sky were also removed, an effective low ℓ -filter – if they stay in, the maps look similar to the unfiltered theory maps except small scale smoothing leads to loss of the higher contour levels. Note that the contours are linearly spaced at $\pm 15n \mu\text{K}$ for all but (a), for which the spacings are $\pm 15, 30, 60, 120 \mu\text{K}$. The maps have been smoothed by an additional 1.66° Gaussian filter.

4-yr COBE North Galactic Pole Map

UNFILTERED

OPTIMAL FILTER



← 140 degrees → ← 140 degrees →

(b,c,d) contours: $\Delta T/T = \pm 0.55 \times 10^{-5} n$, $n \geq 1$

(a) contours: $\Delta T/T = \pm 0.55 \times 10^{-5} 2^{n-1}$, $n \geq 1$

that this would be so is evident from the curves. The 5 heavy open circle points probing ℓ 's ranging from 60 to 400 repeated in the upper and lower panels labelled *sk95* are combined *sk93+94+95* results [150]. The estimated 14% error in the overall amplitude because of calibration uncertainties associated with Cas A are included. The large ℓ -space coverage from this one intermediate angle experiment gives a first glimpse of the ℓ -space coverage that will become standard in the next round of anisotropy experiments.

Python [104], *py*, the heavy solid curve at $\ell \sim 90$, is sensitive to a wide coverage in ℓ -space as the horizontal error bars in the top panel indicate. Argo [105], *ar*, a balloon-borne experiment, is next. The next five points in the lower panel are from the fourth and fifth flights of the MAX [100, 99], *max4,max5*, another balloon experiment. Because the filters changed with frequency, the points are placed at the average over all *max* filters. In the upper panel three *max4* scans are combined into one data point as are two *max5* scans. The lines ending in triangles at $\ell \approx 145$ and 240 denote the 90% limits for the MSAM [101] single (*msam2*) and double (*msam3*) difference configurations. A limitation on these balloon experiments is the ~ 5 hours over which data can be effectively taken. Planned long duration balloon flights that would circle Antarctica for about a week would allow extensive mapping at high precision to be done, and a number of groups have been proposing designs labelled ACE, Boomerang, Maxima, Top Hat.

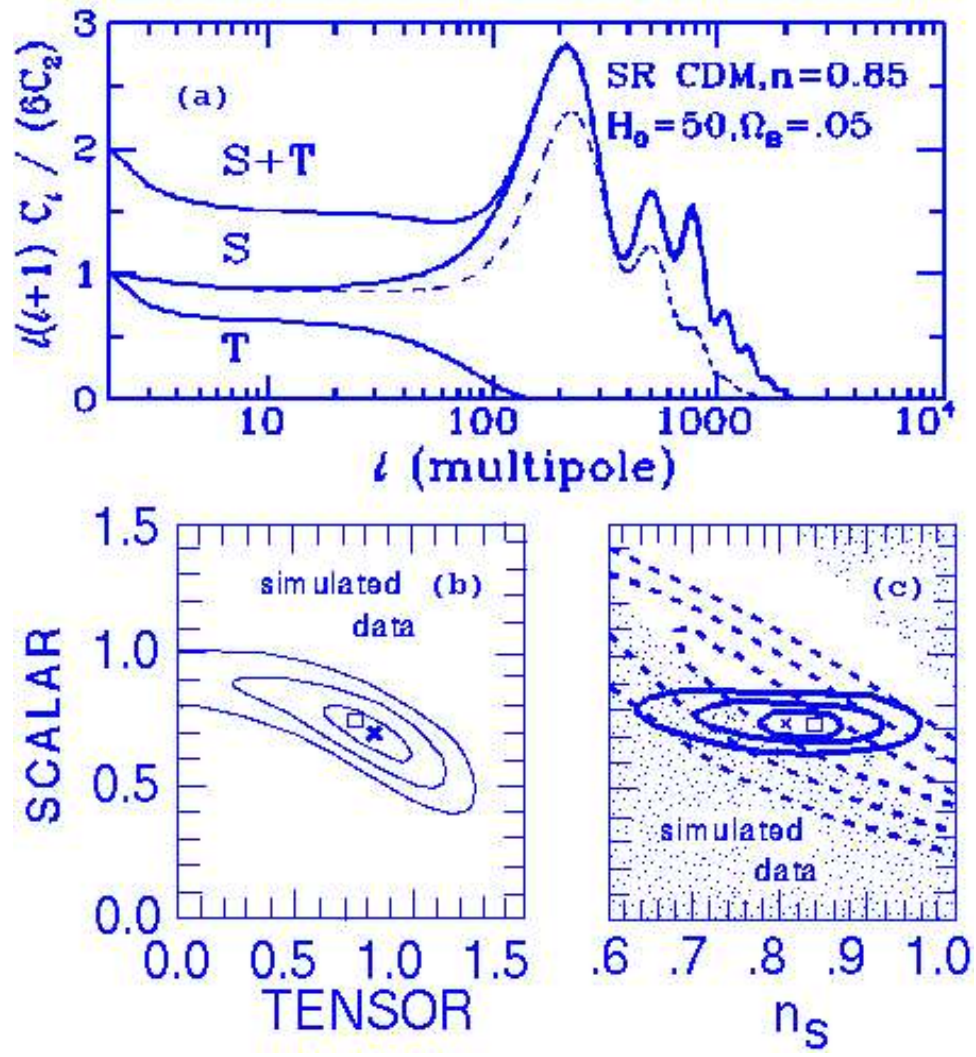
The CAT points at $\ell \approx 400$ and 600 represent a very different experimental technique, interferometry, so I will discuss the approach in some detail. CAT is a 3-element synthesis telescope, probing ~ 15 GHz frequencies with a $27'$ synthesized beam and a 2° field-of-view (the *fwhm* of the individual telescopes). It is a precursor to the larger VSA (Very Small Array), covering a wider frequency range with more telescopes and a larger (4°) FOV. Two other CMB interferometers are also planned: CBI (fig. 16) and VCA. Interferometers directly measure the Fourier amplitudes $\tilde{\Delta}_T(\mathbf{Q})$ for wavevectors $\mathbf{Q}$ associated with the baseline separation of the telescopes; $\mathbf{Q}$ rotates with the rotation of the earth and with many movable telescopes many $|\mathbf{Q}|$ can be probed as well. Analysis is most naturally done in $\mathbf{Q}$ -space with the power spectrum a direct outcome. The phase information can be used to reduce atmospheric contamination. Maps sensitive within the FOV can be made using methods such as maximum entropy reconstruction. The CAT team have done this. The low frequencies of CAT implies that radio source contamination is a problem: these are found using the higher resolution RYLE interferometer (section 5.3.4) and subtracted from the CAT data.

The *ovro* experiments are also at radio frequencies, but use single dishes. The 1987 *ovro* 7 point upper limit [94] used a 40 meter dish. A 5 meter dish has been used for the larger scale *ovro22* experiment (fig. 6) that has detections awaiting the cleaning of radio sources found with the 40 meter dish. WhiteDish [102] had a small amplitude filter function, fig. 6, a hint of a detection in the $m = 1$ mode and a 95% limit in $m = 2$ mode at $\ell \approx 520$, *wd2*. The open triangle at $\ell \approx 160$ is the (historically important) 95% credible limit for the

4.6 Measuring cosmological parameters with the CMB

In the future we will be able to strongly select the preferred theories by simultaneously analyzing experiments like these. Although combining the statistics for a number of experiments was quite effective when we just had upper limits (e.g., using *sp89* and *ov7* in [243]) and interesting when we had a mix of detections and upper limits (e.g., using *dmr1*, *sp91*, *sp89* and *ov7* [140]), it can also be quite misleading unless we are careful to include secondary backgrounds, foregrounds, instrumental systematics and calibration uncertainties as well as primary anisotropies in our model for Δ_p , or can demonstrate that they are absent. Band-power diagrams such as fig. 10 are very useful guides in the evolving progress towards a primary $\mathcal{C}_\ell$ spectrum, and help to inhibit theorists from over-interpreting the cosmological consequences of the current data. Some of the datasets have now been shown convincingly to be consistent with a CMB rather than a foreground or systematic origin, and this warrants a return to the multiresolution approach, since so much more can be determined using long baselines in ℓ -space. In this subsection I illustrate how the multiresolution approach works with two exercises. The first, fig. 4.6 taken from [140], shows what the near future looked like as seen from summer 1993, with experiments still characterized by a narrow $\overline{W}_\ell$ because of beam-to-throw constraints, like most experiments in fig. 10. Although the prognosis was good, an approximate degeneracy in the parameter space was identified that showed that apart from an overall amplitude (e.g., σ_8) only a single parameter, $\tilde{\nu}_s$ (described below), was really well determined by few-scan data [144]. The second illustration, fig. 13, shows how well all-sky high-resolution satellite experiments can do, as examples, albeit best-case ones, of what low noise maps of contiguous regions may achieve: the approximate degeneracy is broken and a large number of cosmological parameters can be simultaneously determined with relatively high accuracy. Less spectacular, in ℓ range and sky coverage, but not by a huge amount, are projections for what will be achievable in ℓ -space from long duration balloon and ground-based interferometry experiments. In between, statistical methods are briefly discussed.

In [140], simulations of data sets with assumed noise levels were constructed for experiments probing $\ell \lesssim 15$ (*dmr*), $\ell \sim 100$ (*sp91*), $\ell \sim 200$ (*sp89*) and $\ell \sim 500$ (*ov22*). The noise levels were being achieved even then and the number of scans chosen for each configuration is conservative. The input signal was a model with $r_{ts} = 1$, equal scalar and tensor quadrupole powers and tilts $\nu_s = \nu_t = -0.15$, with spectrum shown in fig. 4.6(a). The issue addressed was how well the input signal could be recovered as one progressively relaxed what was known. What fig. 4.6(b) shows is that if we happened to know ν_s and ν_t , recovery of the overall amplitude is excellent, and recovery of the ratio r_{ts} is also quite good. (With only *dmr* data, r_{ts} cannot be determined because the tensor and scalar spectra look the same apart from $\ell = 2$.) Allowing ν_s to vary



(caption next page)

Figure 12: (previous page) A vintage simulation of how the multiresolution combination of experiments can determine cosmological parameters, from Crittenden et al. (1993). (a) $\mathcal{C}_\ell/\mathcal{C}_2$ for tensor, scalar and the sum for a tilted ($\nu_s = \nu_t = -0.15$) but otherwise standard CDM model with normal recombination. Hot/cold hybrid models look quite similar. The light dashed line is an $\Omega_B = 0.01$ model. The rest of the panels show contour maps in parameter space derived from simulated large and small angle data consisting of the *dmr* correlation function (with error bars appropriate to 4 years of data), six 13 point *sp91* strips (1.5° beam, 18–27 μ K error bars for each of the 4 frequency channels), six 9 point *sp89* strips (0.5° beam, 15 μ K error bars) and one *ov22* strip (7' beam, 22' throw, 25 μ K error bars). Error bars as good or better than these are now being achieved. The mean signal input into the simulated data is denoted by the square; “**x**” denotes maximum likelihood. (b) Shows 1, 2 and 3 sigma contour lines in the scalar ($[C_2^{(S)}]^{1/2}/10^{-5}$) versus tensor ($[C_2^{(T)}]^{1/2}/10^{-5}$) amplitude plane, assuming the index 0.85 is known. (c) Shows 1, 2 and 3 sigma likelihood contours for the simulated data in $[C_2^{(S)}]^{1/2}-n_s$ space, constrained to obey the $C_2^{(T)}/C_2^{(S)} \approx -7\nu_s$ and $\nu_t = \nu_s$ relationship (solid) and with this ratio unconstrained, but $\nu_t = \nu_s$. Shading indicates the range for which CDM models are not dynamically viable based on σ_8 . Without other information such as this, one recovers well only one parameter, a combination of $[C_2^{(S)}]^{1/2}$ and n_s , while parameters orthogonal to this have wide error bars. This ambiguity increases when the space is opened up to encompass more cosmological parameters.

in fig. 4.6(c) shows that we can get one direction in parameter space very well, that corresponding to the $\tilde{\nu}_s$ variable of eq. (108), but the orthogonal one is sloppy. However, if we accept that we know the relationship between ν_t and ν_s , which can be computed for any specific inflation model, e.g., $\nu_t \approx \nu_s$, with r_{ts} then following, then recovery is excellent. Using the 4-year *dmr* data in conjunction with just the *sk94* and *sp94* data gives error bars on n_s that are similar to what this simulated exercise gave [162]. Another important point to note in fig. 4.6(c) is information on the allowed value of σ_8 taken from cluster observations tightens up the precision with which the parameters are determined. This is addressed more fully in section 7.3. However, if we open up the parameter space to include variations in Ω_Λ , h , z_{reh} , etc. – as in fig. 8 – then more ambiguity arises.

Superposed upon the spectra in figs. 7 and 8 are theoretical band-powers derived for a variety of anisotropy experiments. Figure 7 also shows 10% one-sigma error bars: with 4 years of data, the *dmr* band-power errors are 14%; to achieve this with smaller angle experiments one would need to have about the same number of pixels as COBE, but scaled to the beam size hence covering a smaller region of the Universe: that is, mapping experiments on smaller angular scales. Even if there were idealized perfect all-sky coverage with noise-free versions of the experiments of Fig. 7, there would still be cosmic variance errors on the band-powers to limit the accuracy. These go as $\langle \ell \rangle^{-1}$ [89] even for quite narrow bands, as is shown below. Thus it appears that by using (nearly perfect) CMB experiments which are sensitive to a wide range of angular scales, we might expect to distinguish even among the nearly degenerate theoretical models shown in fig. 8(f), and be able to measure the parameters that define the variations in these models.

The near-degeneracy is especially prevalent through the first Doppler peak. In [144], we showed that for small variations about the “standard” CDM model, with $n_s = 1$, $\Omega = \Omega_{nr} = 1$, $\Omega_B = 0.0125h^{-2}$ (from big bang nucleosynthesis), the height of the first Doppler peak relative to the *dmr* band-power is (within $\sim 15\%$)

$$\frac{\mathcal{C}_\ell|_{max}}{\langle \mathcal{C}_\ell \rangle_{dmr}} \approx 5 e^{3.6\tilde{\nu}_s}, \quad (108)$$

$$\tilde{\nu}_s \approx \nu_s - \frac{\ln(1 + \tilde{r}_{ts})}{3.6} - 0.5[\Omega_{nr}^{1/2}h - \frac{1}{2}] + 0.08\left(\left(\frac{\Omega_B h^2}{0.0125}\right) - 1\right) - \left(\frac{z_{reh}}{200}\right)^{3/2}.$$

The nature of the tensor reduction term is clear. The $\Omega_{nr}^{1/2}h$ term follows from a strong dependence on the redshift at equal energy densities in relativistic and non-relativistic matter or a_{eq} , eq. (69). This term shows that the height of the peak goes up as $\Omega_{vac} = 1 - \Omega_{nr}$ goes up, quite dramatically for fixed h , but not by much for models with fixed age, since h goes down, as fig. 8(d) illustrates. a_{eq} also varies with the relativistic energy density, Ω_{er} , if it is not the “standard” value with three massless neutrino species; if not, $\Omega_{nr}^{1/2}h$ should be divided by $[\Omega_{er}/(1.68\Omega_\gamma)]^{1/2}$ in eq. (108). (See [158] for small variations breaking the

simple a_{eq} degeneracy, and for another form for $\tilde{\nu}_s$.) The reionization term is simply related to the depth to Compton scattering from the re-ionization redshift z_{reh} to the present by $2\zeta_C/3.6 \propto (z_{reh}/z_{\zeta_C=1})^{3/2}$, where $z_{\zeta_C=1} \approx 10^{2.1} (\Omega_B h/0.02)^{-2/3} \Omega_{nr}^{1/3}$ (eq. 72), and so depends on $\Omega_{nr} h^2$ (and on $\Omega_B h^2$). z_{reh} must be $\lesssim 150$ to have a local maximum, as fig. 8(e) shows.

In [144], we fixed $\Omega_B h^2$ at 0.0125, but, as fig. 8(b) shows, a linear dependence in $\tilde{\nu}_s$ on Ω_B gives the variation of the peak height to sufficient accuracy. However, fig. 8(b) also shows that the relative heights of the secondary Doppler peaks are sensitive to Ω_B , so the approximate degeneracy is broken in the variable $\Omega_B h^2$. It is also broken by Ω_{tot} , since the position of the peak, determined by the angle-distance relation, changes. The formula eq. (108) shows that a model with no gravity wave contribution but $n_s \approx 0.88$ has a spectrum that is almost degenerate with the $n_s = 0.95$, $\tilde{r}_{ts} = 0.3$ spectrum, so much so that it will be difficult to tell them apart. We argued that the precision required to separately determine $n_s, \tilde{r}_{ts}, \Omega_\Lambda, \dots$ was too high for what was then the near-term future, but $\tilde{\nu}_s$ could be determined accurately, and that to separate the various contributions to $\tilde{\nu}_s$ in the near term would require other cosmological experiments, e.g., measuring the scalar perturbation shape through galaxy-galaxy power spectra and amplitude through cluster abundances or streaming velocities (section 7.3); and, in some happy future, determining H_0 definitively. In the future, NASA and ESA high precision CMB space experiments should achieve the sensitivities necessary using CMB anisotropy information alone [161, 152, 154].

Fig. 13 gives a closeup view of how very fine differences in the theoretical $\mathcal{C}_\ell$ can be measured using detector sensitivities and the long observing times appropriate for satellite experiments feasible with present technology [152, 153, 154].

To discuss how cosmic variance, experimental noise, and sky coverage affect the estimation of the predicted band-powers, it is worthwhile to make a brief excursion into statistical analysis. For the CMB data sets that have been obtained up to now, including COBE, it has been possible to do a relatively complete Bayesian statistical analyses [155] if the primary anisotropies are Gaussian and the non-Gaussian Galactic foregrounds are not large. The goal is to determine the best error bars on the parameters of a target set of theories with correlation matrices $C_{Tpp'}$, by first determining the likelihood function $\mathcal{L}$ for each theory, and then comparing the likelihoods as a function of the parameters. To give preferred values and errors for a specific cosmological parameter of interest such as the Hubble parameter, one often integrates (marginalizes) over the other parameters, such as σ_8 and n_s , assuming a prior probability distribution, which can be a statement of *a priori* maximal ignorance, or take into account constraints from other information such as large scale structure observations, as is done in section 7.3.

A useful method for likelihood determination is to expand in signal-to-noise eigenmodes [146], those linear combinations of pixels which diagonalize

the matrix $C_n^{-1/2} C_T C_n^{-1/2}$, where the noise correlation matrix $C_n = C_D + C_{res}$ consists of the pixel errors C_D and the correlation of any unwanted residuals C_{res} , whether of known origin such as Galactic or extragalactic foregrounds or unknown extra residuals within the data.³

The S/N -mode basis facilitates the many $N_{pix} \times N_{pix}$ matrix inversions of $C_t \equiv C_n + C_T$ involved in evaluating the likelihood function,

$$\ln \mathcal{L}(\sigma_{th}) = -\frac{1}{2} \bar{\Delta}^\dagger (C_n + C_T)^{-1} \bar{\Delta} - \frac{1}{2} \text{Trace} \ln(C_n + C_T) - N_{pix} \ln \sqrt{2\pi}, \quad (109)$$

as a function of an overall amplitude $\sigma_{th} \propto [\text{Trace}(C_T)]^{\frac{1}{2}}$ (e.g., σ_8 , $C_2^{1/2}/10^{-5}$, $\langle \mathcal{C}_\ell \rangle_{dmr}^{1/2}/10^{-5}$). Here $\dagger$ denotes transpose. (This form of the likelihood function assumes a Gaussian distribution of errors and that the target signal and residuals are also Gaussian-distributed. To derive it, integrate $\delta^{(N_{pix})}(\bar{\Delta} - \Delta)$ over each Δ_A probability distribution, $A = n, T, res$. The total Δ coming in to the detector is modelled as $\Delta = \Delta_n + \Delta_T + \Delta_{res}$, each with a distribution $\exp[-\frac{1}{2} \Delta_A^\dagger C_A^{-1} \Delta_A] / ((2\pi)^{N_{pix}} \det[C_A])^{\frac{1}{2}}$. If the target signal or any of the residuals has a non-Gaussian distribution, the integrations cannot usually be done and Monte Carlo treatments of the statistics becomes necessary.)

Constraints such as averages, gradients (dipoles, quadrupoles) and known templates, which may be frequency dependent (e.g., IRAS or DIRBE dust maps) can also be modelled in the total Δ , as “nuisance variables” to be integrated (marginalized) over. Denoting each constraint c on pixel p by $\kappa_c \Upsilon_{pc}$, where the template for constraint c is Υ_{pc} (e.g., the $\mathcal{F}_{p,1m}$ and $\mathcal{F}_{p,2m}$ of eq. (91) for the dipole and quadrupole) and the amplitude is κ_c , we need only replace $\bar{\Delta}_p$ in eq. (109) by $\bar{\Delta}_p - \sum_c \Upsilon_{pc} \kappa_c$, then integrate, assuming a prior probability distribution for the amplitudes κ_c . This is most easily done if we assume the κ_c are also Gaussian-distributed with a very broad distribution reflecting our ignorance of its values (or if we know its likely range, incorporating that as prior information in the Gaussian spreads). The integration over κ_c then yields

$$\begin{aligned} \ln \mathcal{L}_{+C} = \ln \mathcal{L} + \frac{1}{2} \bar{\Delta}^\dagger [\Upsilon^\dagger C_t^{-1}]^\dagger (K^{-1} + \Upsilon^\dagger C_t^{-1} \Upsilon)^{-1} [\Upsilon^\dagger C_t^{-1}] \bar{\Delta} \\ - \frac{1}{2} \text{Trace} \ln(I + K \Upsilon^\dagger C_t^{-1} \Upsilon), \end{aligned} \quad (110)$$

³The N_{pix} modes $\xi_k = \sum_{p=1}^{N_{pix}} (R C_n^{-1/2})_{kp} (\Delta T/T)_p$, having the “dimensions” of signal-to-noise, can be expanded into noise n_k , signal s_k , residual “noise” res_k not accounted for by C_n , and any further “constraints” c_k (residual dipoles, quadrupoles, etc.): $\xi_k = s_k + n_k + c_k + res_k$ [5, 89, 6, 146]. Here R is a rotation matrix. In this basis, the noise and signal have diagonal correlations: $\langle n_k n_{k'} \rangle = \delta_{kk'}$, $\langle s_k s_{k'} \rangle = \mathcal{E}_{TR,k} \delta_{kk'}$. The great simplification of orthogonality, i.e., no mode-mode correlations, is destroyed somewhat by off-diagonal terms in the $\langle c_k c_{k'} \rangle$ and $\langle res_k res_{k'} \rangle$ (if they are not fully modelled by C_n). The modes are sorted in order of decreasing S/N -eigenvalues, $\mathcal{E}_{TR,k}$, so low k -modes probe the theory in question best. This expansion is a complete (unfiltered) representation of the map. In S/N -filtering, only restricted ranges in this k -space are kept. The sum of ξ_k^2 over bands in S/N -space defines a S/N -power spectrum which gives a valuable picture of the data and shows how well the target theory fares [89, 146, 162].

where $K_{cc'} = \langle \kappa_c \kappa_{c'} \rangle$ is the assumed prior variance for the constraint amplitudes. Evaluating this involves only $N_C \times N_C$ matrix inversions, where N_C is small compared with N_{pix} . Taking into account constraints with amplitudes that are not linear multipliers times the template is much more complex.

An equivalent expression to eq.(110) for $\ln \mathcal{L}_{+C}$ takes the form eq.(109) but with C_n replaced by $\tilde{C}_n \equiv C_n + \Upsilon K \Upsilon^\dagger$. The constraint portion of the matrix is just $\langle \sum_c \kappa_c \Upsilon_{pc} \sum_{c'} \kappa_{c'} \Upsilon_{p'c'} \rangle$. The span of the templates Υ_{pc} defines a subspace in the data. As the eigenvalues of K become very large, the effect of the constraint matrix is to project onto the data subspace orthogonal to that spanned by Υ_{pc} . Although one can directly use the likelihood equation in this projection limit (using $\delta^{(N_C)}(\kappa)$ for the constraint prior), it is computationally simpler to use the Gaussian prior.

Which form of the likelihood to use depends upon the application: using eq.(110), one can vary the number of constraints to include without recomputing the S/N modes associated with $C_n^{-1/2} C_T C_n^{-1/2}$ (e.g., allowing for a Galactic quadrupole contamination in the *dmr* data or not); for a fixed but large number of constraints, the eq.(109) form is better, using S/N modes associated with $\tilde{C}_n^{-1/2} C_T \tilde{C}_n^{-1/2}$.

Many of the determinations of the band-powers and their error bars shown in Fig. 10 were facilitated by S/N mode expansions. This technique has the highest sensitivity to the data, but a byproduct is that it is also sensitive to low level residuals. If these exist, they can sometimes be removed by signal-to-noise filtering, getting rid of modes that are very insensitive to the class of theories being tested.

A strong indication of the robustness of the *dmr* data set is the insensitivity of the band-powers to the degree of signal-to-noise filtering and to which frequencies are probed (section 4.5). This S/N -filtering is a form of data compression: when the eigenmodes are rank-ordered by decreasing eigenvalues, one usually finds that only the moderate to high S/N -modes (e.g., $\sim 10\%$ for COBE) probe the target theory well and the rest must be consistent with noise [89, 6, 146, 162]; and if they are not, filtering out the high S/N -modes leaves offending residuals whose nature can then be explored. [89, 6, 146].

Filtering using S/N -modes has a long history in signal processing where it is called the Karhunen-Loeve method [156], and it is now being widely adopted for analysis of astronomical databases.

When the number of pixels becomes too large, statistical compromises are necessary because the eigenvectors of the full S/N matrices cannot be determined. An all-sky experiment with $10'$ resolution will have more than a million pixels per frequency channel, and long duration balloon experiments will have tens of thousands of pixels. Exploring how to best estimate power spectra and cosmological parameters given computational constraints by first projecting the data onto well chosen smaller subsets is sure to become a very active area. This happy day of too many pixels is now upon us.

The optimal (Wiener) filtering shown in fig. 11 is an immediate byproduct

of the S/N -eigenmode expansion [89, 157, 146]: given observations $\overline{\Delta}_p$, the mean value and variance matrix of the desired signal Δ_{Tp} are [232]

$$\langle \Delta_T | \overline{\Delta} \rangle = C_T C_t^{-1} \overline{\Delta}, \quad \langle \delta \Delta_T \otimes \delta \Delta_T | \overline{\Delta} \rangle = C_T C_t^{-1} (C_t - C_T). \quad (111)$$

The mean field, $\langle \Delta_{Tp} | \overline{\Delta} \rangle$, is the optimally-filtered map. The operator multiplying $\overline{\Delta}$ is the Wiener filter. If the map is very sensitive to the assumed C_T or if the fluctuation, $\delta \Delta_{Tp} = \Delta_{Tp} - \langle \Delta_{Tp} | \overline{\Delta} \rangle$, of the signal about the mean is large in some region of space or on some resolution scale, then this tells us that the data there are not yet good enough in the optimally-filtered maps to identify real structures on the sky with this method. (Marginalization over the constraints is incorporated into the mean field and variance by adding $\Upsilon K \Upsilon^\dagger$ to C_n and thus C_t [146].)

To get an idea of how experimental noise and sky coverage affect the estimation of the predicted band-powers, we consider an experiment with noise matrix $C_{Dpp'} = \sigma_{pix}^2 \delta_{pp'}$, with the per-pixel error σ_{pix} independent of the pixel position (i.e., homogeneous uncorrelated noise). Suppose first that the pixels are sufficiently separated that $C_{Tpp'} \approx 0$ for $p \neq p'$, i.e., that only $\overline{W}_\ell$ is an effective probe of $\mathcal{C}_\ell$. For large N_{pix} , the 1-sigma uncertainty in the experimental value of the band-power about the maximum likelihood value, $\langle \mathcal{C}_\ell \rangle_{B, maxL}$, is [89]

$$\Delta \langle \mathcal{C}_\ell \rangle_B = \sqrt{2/N_{pix}} \left[\langle \mathcal{C}_\ell \rangle_{B, maxL} + \sigma_{pix}^2 / \mathcal{I}[\overline{W}_\ell] \right]. \quad (112)$$

For large N_{pix} , the observed maximum likelihood will fluctuate from $\langle \mathcal{C}_\ell \rangle_{B, th}$, the theoretical quantity we want, but the error bars of eq. (112) include these realization-to-realization fluctuations (thus $\sqrt{2}$ appears, not 1). To get 10% error bars as in fig. 7 requires low experimental noise and $N_{pix} \sim 200$ “independent” pixels, i.e., a mapping experiment.

In a mapping experiment, the pixels will be adjacent and off-diagonal correlations in $C_{Tpp'}$ are very important, but for a large enough contiguous region and simple observing strategies this can be adequately treated with an expansion in the $a_{\ell m}$ (or Fourier) modes. With uniform weighting and all-sky coverage, the S/N -modes are just the independent $Re(a_{\ell m})$ and $Im(a_{\ell m})$. For each ℓ , there is a $(2\ell + 1)$ degeneracy, an effective pixel number for ℓ -modes. If only a fraction f_{sky} of the sky is covered, then for high ℓ , so that the angular scale ℓ^{-1} is small compared with the patch probed, the effective pixel number scales by f_{sky} . Thus, for each ℓ , we have

$$\begin{aligned} \Delta \mathcal{C}_{T\ell} &\approx \frac{\sqrt{2}}{\sqrt{(2\ell + 1)f_{sky}}} (\mathcal{C}_{T\ell} + \mathcal{C}_{res, \ell} + \mathcal{C}_{D\ell} \mathcal{B}_\ell^{-2}), \\ \mathcal{C}_{D\ell} &\equiv \frac{\ell(\ell + 1)\sigma_\nu^2}{2\pi}, \quad \sigma_\nu \equiv \sigma_{pix} \varpi_{pix}. \end{aligned} \quad (113)$$

Thus the cosmic variance for each ℓ goes as $Q^{-1/2}$, where, as usual, $Q \equiv \ell + \frac{1}{2}$. The filter function associated with the beam is $\mathcal{B}_\ell$. It has been divided out

to show that the effective noise level in $\mathcal{C}_\ell$ determination picks up enormously above $\ell_s \sim (0.425\theta_{fwhm})^{-1}$. For fixed experimental parameters, the combination σ_ν remains the same as the pixel size is varied.

Figs. 7 and 8 show that the variation in $\mathcal{C}_{T\ell}$ with cosmological parameters is quite smooth so we can broaden the band-power filters to encompass more than a single ℓ . In fig. 13, the errors shown are those appropriate for logarithmic binning of width $\pm \frac{1}{2}\Delta \ln \ell$ about $\ln \ell$, with $\frac{1}{2}\Delta \ln \ell = 0.05$. This gives a better feeling for how well parameter estimation can occur. The variance is

$$\Delta \mathcal{C}_{T\ell} \approx \frac{[(\mathcal{C}_{T\ell} + \mathcal{C}_{res,\ell})^2 + 2\gamma_1(\mathcal{C}_{T\ell} + \mathcal{C}_{res,\ell})\mathcal{C}_{D\ell}\mathcal{B}_\ell^{-2} + \gamma_2\mathcal{C}_{D\ell}^2\mathcal{B}_\ell^{-4}]^{1/2}}{\sqrt{Qf_{sky}} \sqrt{\cosh(\Delta \ln \ell)[1 + Q\sinh(\Delta \ln \ell)]}}. \quad (114)$$

The factors γ_1 and γ_2 are nearly unity if $\Delta \ln \ell$ is small. There is a crossover point at which $\Delta \mathcal{C}_{T\ell}$ from cosmic variance goes from the usual $Q^{-1/2}$ dependence to a Q^{-1} dependence.

(The derivation evaluates the likelihood function within the (integer-spaced) $[e^{-\Delta \ln \ell/2}\ell] \leq L \leq [e^{\Delta \ln \ell/2}\ell]$ interval. The cosmic variance term is just the sum of $\ell + \frac{1}{2}$ over the bin. The γ_1 and γ_2 terms are estimated by expanding in noise-to-signal, $\mathcal{C}_{D\ell}/[(\mathcal{C}_{T\ell} + \mathcal{C}_{res,\ell})\mathcal{B}_\ell^2]$, up to second order, grouping terms to preserve the basic form of $(\Delta \mathcal{C}_{T\ell})^2$ in eq. (113). If $\nu_{\Delta T}$ is the local slope of $(\mathcal{C}_{T\ell} + \mathcal{C}_{res,\ell})$, then

$$\begin{aligned} \gamma_1 &\approx \frac{\sinh[(2 - \nu_{\Delta T}/2 + (Q\varpi_s)^2)\Delta \ln \ell]}{(2 - \nu_{\Delta T}/2 + (Q\varpi_s)^2)\sinh(\Delta \ln \ell)}, \\ \gamma_2 &\approx 4\gamma_1^2 - \frac{3\sinh[(3 - \nu_{\Delta T} + (Q\varpi_s)^2)\Delta \ln \ell]}{(3 - \nu_{\Delta T} + (Q\varpi_s)^2)\sinh(\Delta \ln \ell)}. \end{aligned} \quad (115)$$

If ϖ_s is small and for a flat $\nu_{\Delta T} = 0$, $\gamma_1 \approx \cosh(\Delta \ln \ell)$ and $\gamma_2 \approx 1$. For example, although fig. 13 includes the full corrections of eq. (115), the result without them is indistinguishable for the $\Delta \ln \ell = 0.1$ chosen.)

The all-sky uniform-noise assumption was used to model the *dmr* correlation function errors before the 1-year data were released, as in [5, 140] and Fig. 4.6. The uniform-noise assumption for regular connected patches covering a fraction f_{sky} of the sky has been used recently to address the ultimate accuracy in measuring cosmological parameters that satellite and balloon experiments might achieve if foreground contamination (i.e., $\mathcal{C}_{res,\ell}$) is ignored [159, 160, 161, 154, 163, 164, 165]. That application will be sketched here, following the treatment in [164], since it represents a nice exercise for working with the likelihood formula, eq.(109), is being widely used, and it allows us to focus on the two forthcoming satellite experiments.

We shall use current specifications for MAP and COBRAS/SAMBA (now called Planck), although these may well evolve. In fig. 13, parameters roughly suitable for the NASA mission MAP [152] and the higher resolution COBRAS/SAMBA [154] are shown. Of the 5 HEMT channels for MAP, we

shall assume the 3 highest frequency channels, at 40, 60 and 90 GHz, will be dominated by the primary cosmological signal, and adopt *fw* beams of 24, 17 and 14 arcminutes, respectively. We shall take the noise power to be $C_{D\ell} = 2.3 \times 10^{-15}$ for each channel (i.e., $25 \mu K$ per $18'$ pixel) for a 2 year mission. For COBRAS/SAMBA, which has both HEMTs and bolometers, we take the 4 best bolometer CMB channels at 100, 150, 220 and 350 GHz to be the primary cosmological ones, with *fw* beams 15, 10, 7, 4 arcminutes and noise the remarkable $C_{D\ell} = \{2.8, 1.5, 0.5, 35\} \times 10^{-17}$ (3.5, 3.5, 3.3, and 444 μK per *fw* pixel), respectively, corresponding to a 14 month mission. We shall also assume $f_{sky} = 0.65$ will be usable, the same as the fraction used in the analysis of the 4-year *dmr* data.

Consider a class of cosmological models with Gaussian-distributed temperature anisotropies defined by a parameter set $\{y_A\}$. For definiteness here we shall use the parameter space $\{\Omega_{tot}, h, \Omega_B h^2, \nu_s, r_{ts}, \Omega_{vac}, \langle \mathcal{C}_\ell \rangle_B, \zeta_C\}$, with the residual energy density, $\Omega_{tot} - \Omega_{vac} - \Omega_B$, assumed to be in cold dark matter, $\langle \mathcal{C}_\ell \rangle_B$ the total bandpower for the experiment and ζ_C the Compton optical depth from a reheating redshift z_{reh} to the present. This is similar to the space used in [154], except that $\mathcal{C}_2$ was used instead of $\langle \mathcal{C}_\ell \rangle_B$ (which does not change the results much); [161] added 2 more parameters, while [164] added 7 more.

For illustration we shall assume that the correct underlying theory is an untilted standard CDM one (the hot/cold model has a very similar power spectrum). After integrating the 8 parameter probability distribution over all other parameters but one, sample results for COBRAS/SAMBA are ν_s to ± 0.006 , $\Omega_B h^2$ to 0.6%, H_0 to 2%, Ω_{vac} to ± 0.05 , Ω_{tot} to ± 0.007 , while for MAP they are ± 0.04 , 5%, 11%, ± 0.28 , ± 0.04 . Here Ω_{tot} is forced to be zero if Ω_{vac} is not, and vice versa, because there is an approximate degeneracy in the $\mathcal{C}_\ell$ spectrum for cosmologies with different $\Omega_{tot}, \Omega_{vac}$ but the same angle-distance relationship which maps the post-recombination spatial structure to angular structure at high ℓ . COBRAS/SAMBA has such high sensitivity that it could even determine $\Omega_{m\nu}$ to ± 0.02 . Of course, between the experimental data and these wonderful numbers many complications, especially foreground removal, must be overcome.

We now sketch the method used for this analysis. Choose the parameter set $\{y_{Am}\}$ which approximately maximizes the likelihood (e.g., using quadratic estimators to determine the power spectrum from the data and fitting it with $\mathcal{C}_\ell(y_{Am})$). Expand $\ln \mathcal{L}$ to quadratic order in $\delta y_A \equiv y_A - y_{Am}$. Adjusting the y_{Am} so that the linear term $\partial \ln \mathcal{L} / \partial y_A$ vanishes, thereby yielding a Gaussian approximation with zero mean to the likelihood,

$$\mathcal{L} \sim \mathcal{L}_m \exp\left[-\frac{1}{2} \sum_{AB} \bar{S}_{AB} \delta y_A \delta y_B\right], \quad \bar{S}_{AB} = \sum_{\ell} \frac{\partial \mathcal{C}_{T\ell}}{\partial y_A} (\Delta \mathcal{C}_{T\ell})^{-2} \frac{\partial \mathcal{C}_{T\ell}}{\partial y_B}. \quad (116)$$

Here the parameter derivatives $\partial \mathcal{C}_{T\ell} / \partial y_A$ are evaluated at y_{Am} as is the $\mathcal{C}_{T\ell}$ in the variance $(\Delta \mathcal{C}_{T\ell})$ given by eq.(113).⁴

⁴The derivation is most easily done using S/N eigenmodes, which, as a result of the

Just as was done for the constraints, it is convenient to choose a Gaussian *prior* probability for the fluctuations δy_A , with covariance matrix H_{AB} . The limit of very large eigenvalues of H corresponds to no prior information on the δy_A . The final probability for the parameter fluctuation δy_A is then a Gaussian with mean zero and variance $(\bar{S} + H^{-1})^{-1}$. If we are interested in the error bars on δy_A irrespective of the values of the other variables, we would marginalize over these. The 1-sigma error is then $\pm \sqrt{(\bar{S} + H^{-1})_{AA}^{-1}}$, the numbers quoted above.

Generally the errors in the parameters will be correlated through nondiagonal components of $(\bar{S} + H^{-1})^{-1}$. There are linear combinations of the parameters which are uncorrelated, namely $(\bar{S} + H^{-1})^{1/2} \delta y$. When the eigenvalues of $(\bar{S} + H^{-1})$ are rank ordered, from high to low, the variable combinations corresponding to the top ones will be very accurately determined, while those for the lowest may be very poorly determined, representing the degenerate directions in parameter space. Such parameter eigenmode combinations are therefore a natural generalization of the degeneracy parameter $\tilde{\nu}_s$ of eq.(108).

These idealized studies do not take into account the issue of separating the many components expected in the data, in particular Galactic and extragalactic foregrounds. As can be seen from fig. 7, the effects of Sunyaev-Zeldovich fluctuations on the power spectrum are not large. However the power comes from the clusters and so non-Gaussian aspects of this “foreground” are important (fig. 16). Little is known about high redshift extragalactic sources in the sub-mm. The shape of the power spectrum will have a $\sim \ell^2$ part just from the source counts, and could also have a tail into lower ℓ associated with clustering, as shown in fig. 7. By contrast, much is known about the abundance of extragalactic radio sources as a function of flux at long wavelengths. However extrapolations to higher frequencies are required, some poorly known fraction of the sources have flat ($p_s \sim 0$) spectra, and it is not known how much of

homogeneous noise assumption, are spherical harmonics for all-sky coverage or Fourier modes for smaller patches. If we expand about parameters $\{y_{A*}\}$ for which the likelihood is not necessarily a maximum, we have

$$\begin{aligned} \ln \mathcal{L} &= \ln \mathcal{L}_* + \sum_A F_A \delta y_A - \frac{1}{2} \sum_{AB} (\bar{S}_{AB} + \delta S_{AB}) \delta y_A \delta y_B, \\ F_A &= \frac{1}{2} \sum_k \frac{[\bar{\xi}_k^2 - (1 + \mathcal{E}_{*k})]}{(1 + \mathcal{E}_{*k})^2} (\partial \mathcal{E}_k / \partial y_A), \quad \bar{S}_{AB} = \frac{1}{2} \sum_k \frac{(\partial \mathcal{E}_k / \partial y_A)(\partial \mathcal{E}_k / \partial y_B)}{(1 + \mathcal{E}_{*k})^2}, \\ \delta S_{AB} &= \frac{1}{2} \sum_k \frac{[\bar{\xi}_k^2 - (1 + \mathcal{E}_{*k})]}{(1 + \mathcal{E}_{*k})^2} \left\{ 2 \frac{(\partial \mathcal{E}_k / \partial y_A)(\partial \mathcal{E}_k / \partial y_B)}{(1 + \mathcal{E}_{*k})} - (\partial^2 \mathcal{E}_k / \partial y_A \partial y_B) \right\}, \end{aligned}$$

where $\delta y_A = y_A - y_{A*}$ and the S/N eigenvalue $\mathcal{E}_k$ and its derivatives are evaluated at $\{y_{A*}\}$. The appropriate (linear) adjustment to the maximum likelihood parameters is $y_{Am} - y_{A*} = (\bar{S} + \delta S)^{-1} F$. The $F_A(y_A - y_{Am})$ term then vanishes, where now $\delta y_A = y_A - y_{Am}$, leaving $\ln \mathcal{L} = \ln \mathcal{L}_m - \frac{1}{2} \sum_{AB} (\bar{S}_{AB} + \delta S_{AB}) \delta y_A \delta y_B$. The matrix δS_{AB} contains the fluctuations in the S/N power spectrum, $\bar{\xi}_k^2$, about its mean value $(1 + \mathcal{E}_{*k})$. If $\mathcal{E}_{*k}$ is the correct theory, the ensemble average of δS_{AB} vanishes and it is usually ignored – as was done for the specific numbers given.

a problem this will be in the optimal 50-150 GHz observing window for the CMB. The $\sim \ell^2$ Poisson part should strongly dominate.

There is currently some optimism that the Galactic foregrounds may not be a severe problem. The individual warm and cold clouds in the standard three phase ISM model are quite small (see, e.g., [50] for an inventory), and the observed structure of the far-infrared emission, dominated by the dust-laden Cirrus clouds discovered by the IRAS satellite, is actually rather filamentary with a power spectrum rising towards low ℓ with $\nu_{\Delta T} \sim -1$ [107]. Galactic bremsstrahlung also has a $\nu_{\Delta T} \sim -1$ power spectrum, found using the *dmr* data [108]. This is extremely important since it suggests that in the $\ell \sim 100 - 500$ range, especially in the frequency range around 90 GHz, these backgrounds will not overly contaminate high precision experiments. Complications will arise however, the most important being the non-Gaussian concentration of power and the possible multicomponent nature of the dust. Because of this rise to low ℓ , the quadrupole is more contaminated than higher multipoles, which results in a large (70%) systematic error in its value from Galactic modelling uncertainties [85], hence the large error bars in fig. 10. An important exercise is to see how well we can do parameter estimation as we vary experimental mapping strategies and sky-coverage when realistic non-Gaussian foregrounds are included in simulated data sets. In [154], optimal-filtering techniques (eq. 111) were used on simulations of primary and secondary signals and realistic frequency-dependent foregrounds to show that a well-designed high resolution experiment with good frequency coverage (e.g., the COBRAS/SAMBA set of channels in fig. 5) should be able to accurately recover the primary signal.

5 Primary and secondary sources of anisotropy

The radiative transfer solution involves a line-of-sight integral of a 3D random source-field $\mathcal{G}(q, \hat{q}, \mathbf{r}, \tau)$ through some region, projecting it onto the 2D sky through the action of the Green function, eq. (18), on the source. General features of the $\mathcal{C}_\ell$'s can be understood from projections of simplified forms for $\mathcal{G}$, which is the exercise undertaken in section 5.1. Since this section also introduces some of the typical mathematical manipulations used to treat transport, it is reasonable on a first pass to just read the introductory overview, and then go directly to section 5.2 which describes the sources for linear primary anisotropies and to section 5.3 which describes some nonlinear secondary anisotropy sources.

5.1 Angular power spectra from 3D random source-fields

A general selection function or visibility $\mathcal{V}(\tau)$ is taken out of $\mathcal{G}$. It “illuminates” the portion of the 3D random source-field we are to look at. In this subsection, general formulae are derived for the multipole coefficients $a_{\ell m}$ in terms of fluctuations in $\mathcal{G}(q, \hat{q}, \mathbf{k}, \tau)$ (eq. (123)) and for the associated $\mathcal{C}_\ell$ in terms of the 3D

- 1) beam=20', 20' pix; noise per pix = $20\mu\text{K}$
- 2) beam=10', 10' pix; noise per pix = $6\mu\text{K}$

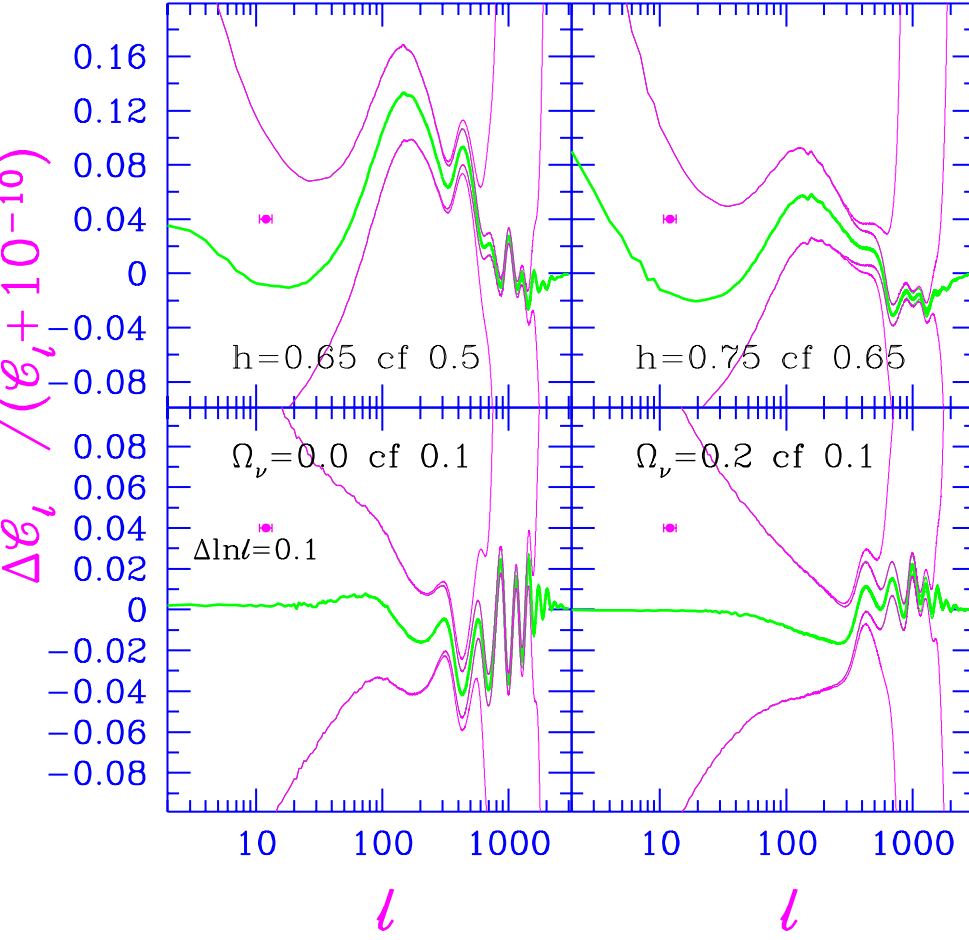


Figure 13: This shows the ability of satellites to measure cosmic parameters to high accuracy. The relative difference of the power spectrum in question from a comparison spectrum (both normalized to the 4-year dmr (53+90+31)A+B COBE maps) are shown so that the few percent deviations can be clearly seen over the entire ℓ range. The lighter lines are 1- σ error bars for all-sky coverage (averaged over the smoothing width shown, with $\frac{1}{2}\Delta \ln \ell = 0.05$) and include cosmic variance (dominant at low ℓ) and pixel noise at $20\mu\text{K}$ or $6\mu\text{K}$ (dominant at high ℓ), with the very rapid growth relative to the theory curve at high ℓ coming from the finite beam-size (with the $fwhm$ indicated, corresponding to a Gaussian filter in multipole space of $\ell_s = 404$ and $\ell_s = 809$ respectively). The first choice corresponds to the NASA satellite experiment MAP, the second choice to the ESA mission COBRAS/SAMBA, if the entire sky were usable (errors scaling $\propto f_{sky}^{-1/2}$). The ultimate accuracy achievable will depend upon the decontamination of the primary signal of non-Gaussian Galactic synchrotron, bremsstrahlung and dust signals. The models shown all have a uniform age of 13 Gyr, $\Omega_{cdm} + \Omega_{m\nu} + \Omega_{\Lambda} + \Omega_B = 1$, $\Omega_B h^2 = 0.0125$, $n_s = 1$ and no gravity wave contribution. Notice the scale change for the hot/cold model panels. (One species of massive neutrino was adopted for these two cases.)

power spectra for $\mathcal{G}$ (eq. (125)). Seeing what happens in special cases is quite instructive: limiting cases for high ℓ (eq. (126)) and the relation to the Fourier transform approximation for 2D maps (eq. (129)); narrow and broad visibility limits (section 5.1.5); simplified 3D spectra which allow analytic evaluations (section 5.1.6). The latter tells us in what limits the phenomenological spectra of section 4.4 are realized by the transport of physical 3D fields; in particular, 3D power law spectra with narrow visibilities lead to the 2D formula eq. (107); white noise spectra with Gaussian coherence-filtering lead to the 2D “Gaussian correlation function” model for a narrow visibility function, but the form is modified for broad visibilities.

These results are applied to a treatment of SZ and dust-emission secondary anisotropies in section 5.3 to give an understanding of why the spectra for these anisotropies look as they do in fig. 7. All that is needed from this section, section 5.1, is the broad visibility high ℓ limit. For secondary anisotropies, the back action of the radiation field on the fluctuations in $\mathcal{G}$ is usually ignorable, but $\mathcal{G}$ is determined by the nonlinear physics of cosmic structure evolution – and subject to the inevitable approximations the treatment of that entails.

A look at the dominant source fields for primary scalar anisotropies is given in section 5.2, which relies upon the results quoted in section 3.1 for $\mathcal{G}$ for Thomson scattering and the Sachs–Wolfe effect. Among other things it describes how one arrives at the $\Delta T/T \sim \Phi_N/3$ “naive Sachs–Wolfe formula,” with Φ_N the gravitational potential, and what it neglects. This section will be easier to follow in conjunction with section 6 which gives a full treatment of perturbation theory and the primary scalar (section 6.5.1) and tensor (section 6.5.2) anisotropies. We shall see that the 3D source fields are highly coupled to ΔT and to each other so we can expect analytic forms for $\mathcal{G}$ to be only approximate.

5.1.1 Simple sample sources

The source function $\mathcal{G}(q, \hat{q}, \mathbf{r}, \tau)$ can be expanded in powers of $\hat{q}$. For all sources we need to consider, the expansion contains at most terms of quadratic order in $\hat{q}$. The quadratic terms for scalar and tensor modes come from the Sachs–Wolfe gravitational redshift sources of eq. (22) and some of the subdominant Thomson scattering sources. However the dominant Thomson scattering source terms and all of the secondary sources have only terms of zeroth and first order in $\hat{q}$, i.e., monopole and dipole sources; further momentum space transformations can put the scalar Sachs–Wolfe terms into this form.

For the illustrations in this section, we shall consider only monopole and dipole sources, and further simplify the dipole by assuming it is a gradient:

$$\begin{aligned} \mathcal{G} &\approx \mathcal{V}(\tau) [\mathcal{G}_0(\mathbf{r}) - \hat{q} \cdot \nabla \mathcal{G}_1(\mathbf{r}) + \cdots], \\ \frac{\Delta T}{T} &\approx \mathcal{T}(\tau) [\Delta_{\mathcal{G}_0} + \Delta_{\mathcal{G}_1} + \cdots]. \end{aligned} \quad (117)$$

The secondary anisotropy sources accompanying distortions are of the scalar $\mathcal{G}_0$ form: eq. (31) for Compton upscattering, eq. (46) for dust emission and

eq. (38) for bremsstrahlung. The asymmetry in Thomson scattering from the flow of electrons contributes a dipole term, $\sigma_T \bar{a} \hat{q} \cdot [n_e \mathbf{v}_e]$, to $\mathcal{G}$. The current $n_e \mathbf{v}_e$ can be expanded in terms of the gradient of a scalar potential, thus of the $\mathcal{G}_1$ form, and the curl of a divergenceless vector potential. For primary scalar perturbations, the curl vanishes, leaving only a $\mathcal{G}_1$ -type term, but when the gas is nonlinear and clumpy, ∇n_e will not be aligned with $\mathbf{v}_e$, generating a curl source which can sometimes exceed the anisotropy driven by the gradient term.¹ For primary scalar fluctuations there are also $\mathcal{G}_0$ -type terms.

For mathematical convenience, a (differential) “visibility function” $\mathcal{V}$ has been removed from the sources and a “transparency function” $\mathcal{T}$ from $\Delta T/T$. This is useful to do if there is a reasonably strong concentration of the “emissivity” in redshift space. For Thomson scattering, the transparency is $\mathcal{T} = e^{-\zeta_C}$. Depending upon which sources we are interested in, $\mathcal{V}$ will either be the differential Thomson visibility function, $\mathcal{V}_C = n_e \sigma_T \bar{a} e^{-\zeta_C}$, of fig. 3 or the integrated visibility $e^{-\zeta_C}$. For the Sunyaev–Zeldovich effect from clusters, we would take both $\mathcal{T}$ and $\mathcal{V}$ to be unity. For dust emission from primeval galaxies, $\mathcal{V}$ can be chosen to define an (angle-averaged) emission shell and $\mathcal{T}$ to be unity, as we described in BCH2 [42].

5.1.2 Angular power spectra for simple sample sources

We now manipulate and solve the transfer equation in a manner which shows how one passes to $\mathcal{C}_\ell$ from the 3D power spectra for the random source fields,

$$\mathcal{P}_{\mathcal{G}_A \mathcal{G}_B}(k; \bar{\chi}, \Delta\chi) = \frac{k^3}{2\pi^2} \langle \mathcal{G}_A(\mathbf{k}, \tau_+) \mathcal{G}_B^*(\mathbf{k}, \tau_-) \rangle, \quad A, B = 0, 1, \quad (118)$$

where $\bar{\chi} = (\chi_+ + \chi_-)/2$, $\Delta\chi = \chi_+ - \chi_-$, $\chi_\pm = \bar{\chi} \pm \frac{1}{2}\Delta\chi$. We shall assume that the sources are statistically homogeneous and isotropic, so that the 3D power spectra are functions only of $|\mathbf{k}|$. If the sources are Galactic, for example, this will not be correct.

The easiest way to deal with the gradient terms is to rewrite the transfer equation as

$$\begin{aligned} \frac{\partial}{\partial \tau} \left(\frac{\Delta T}{T} \right)' + \hat{q} \cdot \nabla \left(\frac{\Delta T}{T} \right)' &= \mathcal{G}', \\ \left(\frac{\Delta T}{T} \right)' &= \frac{\Delta T}{T} + \mathcal{V} \mathcal{G}_1, \quad \mathcal{G}'(q, \mathbf{r}, \tau) = \mathcal{V}(\tau) \mathcal{G}_0 + \frac{\partial}{\partial \tau} \mathcal{V} \mathcal{G}_1, \end{aligned} \quad (119)$$

where $\mathcal{G}'$ now has no $\hat{q}$ dependence. If we assume that there is initially no anisotropy, then the solution (in a flat background cosmology) is

$$\left(\frac{\Delta T}{T} \right)'(q, \hat{q}, \text{here, now}) = \int_0^\infty d\chi \int \frac{d^3 \mathbf{k}}{(2\pi)^3} \delta \mathcal{G}'(q, \mathbf{k}, \tau) e^{-i\mathbf{k} \cdot \hat{q} \mathcal{R}(\chi)},$$

¹Nonlinear Thomson scattering in a flowing plasma is responsible for the moving cluster effect generated at relatively low redshift from ionized gas in groups and clusters, and for the Vishniac effect, which has quadratic nonlinearities included to correct scalar primary anisotropies in a baryon isocurvature example with early reionization.

$$\mathcal{R}(\chi) = \chi = \tau_0 - \tau. \quad (120)$$

In open or closed universes, the mean curvature precludes making a Fourier transform expansion, but a prescription for small angles using a modified $\mathcal{R}(\chi)$ is described in eq. (130) below.

For secondary anisotropies, there is a nonzero angle-averaged part to the random function $\mathcal{G}(q, \hat{q}, \mathbf{r}, \tau)$, which gives average spectral distortions:

$$\left\langle \frac{\Delta T}{T}(q, \mathbf{r} = 0, \tau_0) \right\rangle_{angle} = \int_0^\infty \int \frac{d\Omega_{\hat{r}}}{4\pi} \mathcal{V}(\tau) \mathcal{G}_0(q, r = \chi, \hat{r}, \tau) d\chi \quad (121)$$

For primary Thomson scattering anisotropies, this term vanishes.

To go from eq. (119) to the anisotropy pattern on the sky as embodied in the multipole coefficients $a_{\ell m}$ of eq. (82), we make use of the plane wave expansion

$$e^{-i\mathbf{k} \cdot \hat{q}\chi} = \sum_{\ell m} (-i)^\ell 4\pi Y_{\ell m}^*(\hat{k}) Y_{\ell m}(\hat{q}) j_\ell(k\chi). \quad (122)$$

(Recall that $(2\ell+1)P_\ell(\hat{q} \cdot \hat{k}) = \sum_m 4\pi Y_{\ell m}^*(\hat{k}) Y_{\ell m}(\hat{q})$.) Denoting the contribution of mode $\mathbf{k}$ to the anisotropy $a_{\ell m}$ by $\tilde{a}_{\ell m}(\mathbf{k})$, we have

$$a_{\ell m} = \int \frac{d^3\mathbf{k}}{(2\pi)^3} \tilde{a}_{\ell m}(\mathbf{k}), \quad (123)$$

$$\begin{aligned} \tilde{a}_{\ell m}(\mathbf{k}) &= \int_0^\infty d\chi (-i)^\ell 4\pi Y_{\ell m}^*(\hat{k}) j_\ell(k\chi) \delta\mathcal{G}'(q, \mathbf{k}, \tau) \\ &= \int_0^\infty d\chi \mathcal{V}(\chi) (-i)^\ell 4\pi Y_{\ell m}^*(\hat{k}) [\mathcal{G}_0 j_\ell(k\chi) + k\mathcal{G}_1 j'_\ell(k\chi)] \end{aligned} \quad (124)$$

using an integration by parts on the $\tilde{\mathcal{G}}_1$ term. The statistical homogeneity and isotropy assumption implies in particular that the correlation of $\langle \tilde{a}_{\ell m}(\mathbf{k}) \tilde{a}_{\ell m}^*(\mathbf{k}') \rangle$ is zero unless $\mathbf{k}' = \mathbf{k}$, so the 2D radiation power spectrum, $\mathcal{C}_\ell = \ell(\ell+1) \langle |a_{\ell m}|^2 \rangle / (2\pi)$, is

$$\begin{aligned} \mathcal{C}_\ell &= 2\ell(\ell+1) \int d\ln k \int d\bar{\chi} d\Delta\chi \mathcal{V}(\bar{\chi} + \tfrac{1}{2}\Delta\chi) \mathcal{V}(\bar{\chi} - \tfrac{1}{2}\Delta\chi) \\ &\quad \sum_{A,B=0,1} k^{A+B} \mathcal{P}_{\mathcal{G}_A \mathcal{G}_B}(k; \bar{\chi}, \Delta\chi) j_\ell^{(A)}(k\chi_+) j_\ell^{(B)}(k\chi_-) \end{aligned} \quad (125)$$

with $j_\ell^{(0)} = j_\ell$, $j_\ell^{(1)} = j'_\ell$ and $A, B = 0, 1$. Thus to understand how the $\mathcal{C}_\ell$ will look, we must get into the arcana of how products of spherical Bessel functions behave. We show below that the product $j_\ell^{(A)}(k\chi_+) j_\ell^{(B)}(k\chi_-)$ can be written as $j_\ell^{(A)}(k\bar{\chi}) j_\ell^{(B)}(k\bar{\chi}) (\cos(\bar{k}_\parallel \Delta\chi)$ plus a fast-oscillation term which will often average to zero). In that case, the $\Delta\chi$ integration reduces to a Fourier transform defining a function $p(\bar{k}_\parallel; \bar{\chi}, A, B)$ which encodes information

about variations in the visibility about the average longitudinal distance $\bar{\chi}$:

$$\begin{aligned} \frac{d\mathcal{C}_\ell}{d\ln k} &\approx 2\ell(\ell+1) \\ &\times \int d\bar{\chi} \mathcal{V}^2(\bar{\chi}) \sum_{A,B} p(\bar{k}_\parallel) \mathcal{P}_{\mathcal{G}_A \mathcal{G}_B}(k; \bar{\chi}, 0) \langle j_\ell^{(A)} j_\ell^{(B)} \rangle(k\bar{\chi}) k^{A+B}, \\ p(\bar{k}_\parallel) &\equiv \int d\Delta\chi \frac{\mathcal{V}(\chi_+) \mathcal{V}(\chi_-)}{[\mathcal{V}(\bar{\chi})]^2} \frac{[\mathcal{P}_{\mathcal{G}_A \mathcal{G}_B} + \mathcal{P}_{\mathcal{G}_B \mathcal{G}_A}](k; \bar{\chi}, \Delta\chi)}{[\mathcal{P}_{\mathcal{G}_A \mathcal{G}_B} + \mathcal{P}_{\mathcal{G}_B \mathcal{G}_A}](k; \bar{\chi}, 0)} \cos(\bar{k}_\parallel \Delta\chi), \end{aligned} \quad (126)$$

where $\bar{k}_\parallel \equiv k_\parallel(\bar{\chi})$, $k_\parallel(\chi) \equiv \sqrt{k^2 - Q^2/\chi^2}$, $Q \equiv \ell + \frac{1}{2}$. Here $\langle j_\ell^{(A)} j_\ell^{(B)} \rangle$ is either a direct product of the Bessel functions evaluated at $k\bar{\chi}$ or an approximation to it given by eq. (128) below. For example, the main feature of $\langle j_\ell j_\ell \rangle(x)$ at high ℓ is that it is nearly zero for $x < Q$, falling from a finite maximum as $(2x\sqrt{x^2 - Q^2})^{-1}$ for $x > Q$.

Examples of $d\mathcal{C}_\ell/d\ln k$ for primary anisotropies are shown in fig. 14 for the standard CDM model for a variety of ℓ 's. At $\ell = 4, 10$ the oscillatory behavior of the products of j_ℓ 's is apparent. For $\ell = 59, 121$, smoothing over nearby ℓ 's has been done, with the sharp rise in k (at $\sim \ell/\bar{\chi}$) and power law decline a characteristic shape for averages of j_ℓ products, as we now describe.

5.1.3 Products of Bessel functions

In this technical subsection, we make use of standard Bessel function asymptotics to define approximations for $\langle j_\ell^{(A)} j_\ell^{(B)} \rangle(x)$:

$$\begin{aligned} j_\ell(k\chi) &\sim \frac{1}{\sqrt{k_\parallel k\chi}} \left(\cos(k_\parallel \chi - Q \arcsin(k_\parallel/k) - \frac{\pi}{4}) + \mathcal{O}(Q^{-1}) \right), \quad k_\parallel > 0, \\ j_\ell(k\chi) &\sim \frac{1}{2} \left(\frac{k}{|k_\parallel| + \sqrt{|k_\parallel|^2 + k^2}} \right)^{\ell+1} \frac{\exp[|k_\parallel|\chi]}{\sqrt{|k_\parallel|k\chi}}, \quad k_\parallel \text{ imaginary}, \\ j_\ell(k\chi) &\approx \frac{\sin(\pi/3)\Gamma(1/3)}{3^{2/3}2^{1/6}\sqrt{\pi}} \frac{1}{(k\chi)^{5/6}} \left(1 - \sqrt{2} \frac{k_\parallel}{k} + \mathcal{O}\left(\frac{k_\parallel}{k}\right)^2 \right), \quad k_\parallel \approx 0. \end{aligned}$$

For high ℓ , the $(k\chi/Q)^{\ell+1}$ behavior in the $k\chi < Q$ “imaginary- $k_\parallel$ ” regime ensures that it is almost zero.

To evaluate products of form $j_\ell^{(A)}(k\chi_+) j_\ell^{(B)}(k\chi_-)$, with $A, B = 0, 1$ as in eq. (125), expand in $\Delta\chi/\bar{\chi}$:

$$j_\ell^{(A)}(k\chi_+) j_\ell^{(B)}(k\chi_-) \sim \langle j_\ell^{(A)} j_\ell^{(B)} \rangle(k\bar{\chi}) (\cos(\bar{k}_\parallel \Delta\chi) + \text{fast}). \quad (127)$$

“Fast” denotes a cosine or sine term with a large argument consisting of terms like $\bar{k}_\parallel \bar{\chi}$ which average to zero. For the $\langle j_\ell^{(A)} j_\ell^{(B)} \rangle(k\bar{\chi})$, one can either take the

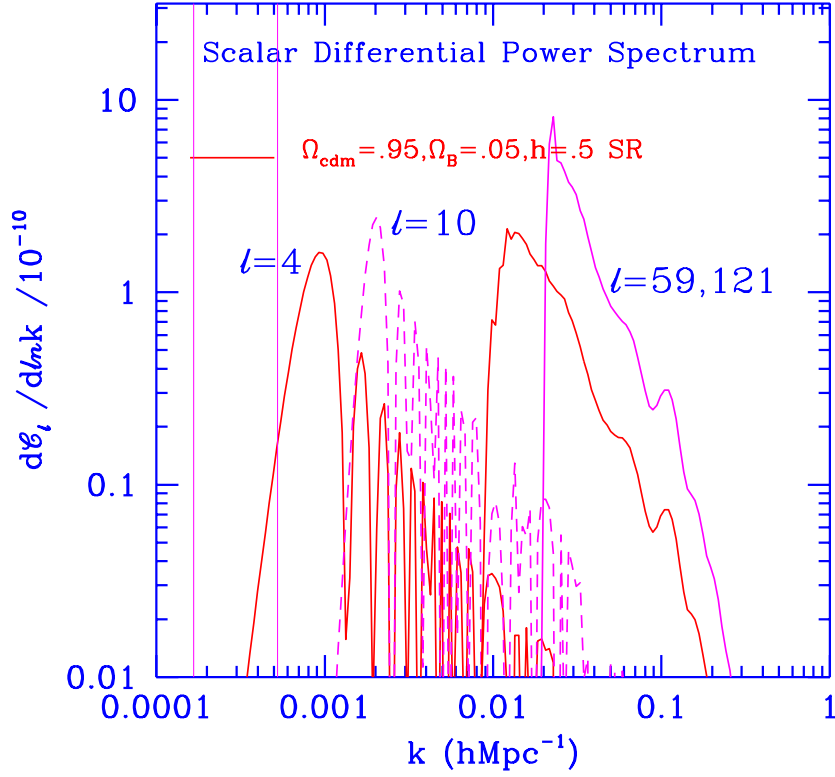


Figure 14: $d\mathcal{C}_\ell/d\ln k$ for the scale invariant CDM model shown demonstrates the basic $j_\ell^2(k\chi_{dec})$ oscillatory behavior for low ℓ . For the two higher ℓ cases shown, smoothing over nearby ℓ 's has been done to damp the fast oscillations, and the result basically follows the limiting high ℓ behavior of the products of j_ℓ and/or j'_ℓ . The vertical lines are defined by $k^{-1} = 2cH_0^{-1}$ and $\pi k^{-1} = 2cH_0^{-1}$, when half a wavelength equals the horizon size.

product $j_\ell^{(A)}(k\bar{\chi})j_\ell^{(B)}(k\bar{\chi})$, or use an average based on the high- Q asymptotics [88]:

$$\begin{aligned}
\langle j_\ell^{(A)} j_\ell^{(B)} \rangle(k\bar{\chi}) &= 0 \quad \text{for } k < Q/\bar{\chi}; \text{ and for } k > Q/\bar{\chi}, \\
\langle j_\ell^{(0)} j_\ell^{(0)} \rangle(k\bar{\chi}) &= \min \left\{ \frac{1}{2k\bar{k}_\parallel \bar{\chi}^2}, j_\ell^2(Q) \left(\frac{Q}{k\bar{\chi}} \right)^{5/3} \right\}, \quad j_\ell(Q)Q^{5/6} \approx 0.59, \\
\langle j_\ell^{(1)} j_\ell^{(0)} \rangle &= -\min \left\{ \frac{(k^2 + \bar{k}_\parallel^2)}{(2k\bar{\chi})^2 \bar{k}_\parallel \bar{\chi} \bar{k}_\parallel^2}, \frac{5}{6Q} j_\ell^2(Q) \left(\frac{Q}{k\bar{\chi}} \right)^{8/3} \right\}, \\
\langle j_\ell^{(1)} j_\ell^{(1)} \rangle &= \min \left\{ \frac{1}{2k\bar{k}_\parallel \bar{\chi}^2} \left(\frac{\bar{k}_\parallel^2}{k^2} + \frac{(k^2 + \bar{k}_\parallel^2)^2}{4(k\bar{\chi})^2 \bar{k}_\parallel^4} \right), \frac{25}{36Q^2} j_\ell^2(Q) \left(\frac{Q}{k\bar{\chi}} \right)^{11/3} \right\}.
\end{aligned} \tag{128}$$

The apparent singularities as $\bar{k}_\parallel \rightarrow 0$ are avoided by saturating the $\langle j_\ell^{(A)} j_\ell^{(B)} \rangle$ at their values for $\bar{k}_\parallel = 0$, as indicated by the minimum function. For low Q , the drop on the imaginary- $\bar{k}_\parallel$ side is not rapid enough to use this approximation, but the expansion in $\Delta\chi$ may still be good: the direct product $j_\ell^{(A)} j_\ell^{(B)}$ should then be used. It is by dropping the “fast” term that we get eq. (126).

5.1.4 Fourier derivation of the simple sample spectra at high ℓ

With the form for $\langle j_\ell^{(A)} j_\ell^{(B)} \rangle(k\bar{\chi})$ valid for high ℓ , we encounter the first of a class of “small angle approximations” that have been used over the years to simplify the calculations of $C(\theta)$ and $\mathcal{C}_\ell$. They turn out to be reasonably good provided we are not interested in low multipoles and do not need the answer to high precision for higher multipoles. A useful exercise to guide understanding of eqs. (125), (126) is to first derive the correlation function $C(\varpi)$ for an isotropic source source-field of the $\mathcal{G}_0$ form, then calculate its 2D Fourier transform (using the notation of eq. (88), and splitting the wavenumber into components $k_\parallel$ along the average line of sight and $\mathbf{k}_\perp$, perpendicular to it):

$$\begin{aligned}
\mathcal{C}_\ell &\approx \int d\bar{\chi} d\Delta\chi \mathcal{V}(\chi_+) \mathcal{V}(\chi_-) \\
&\int_0^\infty dk_\parallel \left(\frac{Q}{\mathcal{R}(\bar{\chi})} \right)^2 e^{-ik_\parallel \Delta\chi} \frac{1}{k^3} \mathcal{P}_{\overline{\mathcal{G}\mathcal{G}}}(k; \bar{\chi}, \Delta\chi), \\
k^2 &\approx (Q/\mathcal{R}(\bar{\chi}))^2 + k_\parallel^2, \quad Q \equiv \ell + \frac{1}{2}, \quad \mathcal{R}(\bar{\chi}) = \bar{\chi}; \\
\int_0^\infty (\dots) dk_\parallel &= \int_{Q/\mathcal{R}(\bar{\chi})}^\infty (\dots) k^2 dk \frac{1}{k \sqrt{k^2 - (Q/\mathcal{R}(\bar{\chi}))^2}}, \\
\mathcal{V}(\chi_+) \mathcal{V}(\chi_-) \mathcal{P}_{\overline{\mathcal{G}\mathcal{G}}}(k; \bar{\chi}, \Delta\chi) &\equiv \int \frac{d\Omega_{\hat{q}}}{4\pi} \frac{k^3}{2\pi^2} \langle \mathcal{G}(q, \hat{q}, \mathbf{k}, \tau_+) \mathcal{G}^*(q, \hat{q}, \mathbf{k}, \tau_-) \rangle.
\end{aligned} \tag{129}$$

Equation (129) allows one to turn the $k_\parallel$ integral into one over k ; hence, with $p(k_\parallel)$ defined as a Fourier transform over $\Delta\chi$, we regain eq. (126) with the

eq. (128) approximation. Note that the $k > Q/\bar{\chi}$ restriction is just a consequence of positive $k_{\parallel}$.

As a further approximation when the source-fields are not isotropic, but have more complicated angular dependence, e.g., $\mathcal{G}(q, \hat{q}, \mathbf{k}, \tau)$, an isotropized power spectrum for the source fields has been used; for the $\mathcal{G} = \mathcal{V}(\mathcal{G}_0 - i\hat{q} \cdot \hat{k} k \mathcal{G}_1)$ case of this section, $\mathcal{P}_{\bar{\mathcal{G}}\bar{\mathcal{G}}} = \mathcal{P}_{\mathcal{G}_0\mathcal{G}_0} + \mathcal{P}_{\mathcal{G}_1\mathcal{G}_1} k^2/3$: cross-correlation terms do not appear.²

These small-angle $\mathcal{C}_{\ell}$ formulas can also be applied to open (or closed) universes for multipoles on angular scales $\ell \gg 2|1 - \Omega|^{1/2}\Omega^{-1}$ if we replace $\mathcal{R}(\chi) = \chi$ by

$$\begin{aligned} \mathcal{R}(\chi) &= d_{curv} \sinh((\tau_0 - \tau)/d_{curv}) \quad \text{if } \Omega < 1, \quad (\Omega_{curv} \equiv 1 - \Omega), \\ d_{curv} &= H_0^{-1} \Omega_{curv}^{-1/2}, \quad \Omega_{curv} \equiv 1 - \Omega, \quad \tau_0 = f_{\tau} 2H_0^{-1} \Omega_{nr}^{-1/2}, \\ f_{\tau} &= \left(\frac{\Omega_{nr}}{\Omega_{curv}} \right)^{1/2} [\ln(\Omega_{curv}^{1/2} + 1) - \ln(\Omega_{nr}^{1/2})], \quad (\Omega = \Omega_{nr} < 1), \end{aligned} \quad (130)$$

where τ_0 is the conformal time now and f_{τ} is a factor that must sometimes be computed, e.g., in universes with sizable vacuum energies. For closed universes, $f_{\tau} = (\Omega_{nr}/|\Omega_{curv}|)^{1/2} \arcsin[(|\Omega_{curv}|/\Omega_{nr})^{1/2}]$ and “sinh” is replaced by “sin”. In flat universes, $d\mathcal{C}_{\ell}/d\ln k$ is concentrated around $Q \sim k\bar{\chi}$, in open universes Q is pushed higher than $k\bar{\chi}$ and in closed universes Q is pushed lower: hence features in $\mathcal{C}_{\ell}$ will be at smaller angular scales in open models (as fig. 18(c,d) shows) and larger scales in closed models than in flat ones.

5.1.5 Narrow and broad visibilities

A few other nonessential approximations are useful to get a simple analytic form for $p(\bar{k}_{\parallel})$: e.g., absorb the leading growth terms in $\mathcal{P}_{\mathcal{G}_A\mathcal{G}_B}$ in the visibility product so the remaining weak dependence upon $\Delta\chi$ can be ignored – e.g., the linear growth factors $D(\chi_+)D(\chi_-)$ may describe the dependence. To get a nice result for discussion of limiting cases, it is useful and is often not even a bad approximation to assume that the product of visibilities looks like a Gaussian in $\Delta\chi$, $\mathcal{V}(\chi_+)\mathcal{V}(\chi_-) \approx \mathcal{V}^2(\bar{\chi}) e^{-(\Delta\chi/R_{\mathcal{V}})^2/4}$.

$$p(\bar{k}_{\parallel}) \approx 2\sqrt{\pi} R_{\mathcal{V}} e^{-\bar{k}_{\parallel}^2 R_{\mathcal{V}}^2}, \quad R_{\mathcal{V}}^{-2} \equiv -\frac{\partial^2 \ln \mathcal{V}}{\partial \bar{\chi}^2}. \quad (131)$$

The function $R_{\mathcal{V}}(\bar{\chi})$ will generally be dependent upon $\bar{\chi}$ and could also depend upon $A, B = 0, 1$. A case of some interest where it is constant is a Gaussian

²Replacing $\langle \mathcal{G}(q, \hat{q}, \mathbf{k}, \tau_+) \mathcal{G}^*(q, \hat{q}', \mathbf{k}, \tau_-) \rangle$ by $\langle \mathcal{G}(q, \hat{q}_P, \mathbf{k}, \tau_+) \mathcal{G}^*(q, \hat{q}_P, \mathbf{k}, \tau_-) \rangle$, where $\hat{q}_P \approx (\hat{q} + \hat{q}')/2$ – without the averaging – is a “DSZ” approximation [131]; [132, 134, 88] exploited the isotropized form. These methods have been applied to $C(\theta)$ and $\mathcal{C}_{\ell}$ estimations for primary anisotropies (section C.3.2) and to the secondary $\delta n_e \mathbf{v}_e$ nonlinear source-field ([216] and section 5.3.6).

visibility. We showed in section 3.4.2 that this was a reasonable approximation for $\mathcal{V}_C$ for standard recombination, with $R_{\mathcal{V}_C}(\bar{\chi}) = R_{\mathcal{V}_C, dec}$. It is used in section 5.2. There are also a few interesting limiting cases:

- 1 *Broad visibility.* When the selection function is broad, $R_{\mathcal{V}} \rightarrow \infty$, $p(\bar{k}_{\parallel}) \rightarrow 2\pi\delta(\bar{k}_{\parallel})$. This is the approximation used in [2, 42] and eq. (140) in section 5.3.1 below for anisotropy power spectra from primeval dust and the SZ effect. The broad limit has a nice interpretation: the “column depth” across the visibility surface is $\Sigma_0(\hat{x}) = \int d\chi \mathcal{V}(\chi) \mathcal{G}_0(\mathbf{x} = \hat{x}\chi, \tau)$. The radiation correlation function $C(\theta)$ is the correlation function of the column depths along lines of sight separated by angle θ . For SZ anisotropies, the “column depth” is the Compton y -parameter, and for dust-emission anisotropies it is the dust optical depth (for a constant redshifted dust temperature).
- 2 *Narrow visibility.* When the selection function is narrow, $R_{\mathcal{V}} \rightarrow 0$, $p(\bar{k}_{\parallel}) \rightarrow 2\sqrt{\pi}R_{\mathcal{V}}$, and the power spectrum is integrated over the unobserved $\bar{k}_{\parallel}$, the projection of the 3D spectrum onto 2D. Only for primary anisotropies with normal recombination is one ever really close to this limit and even then damping of the spectrum for $\ell \gtrsim \bar{\chi}/R_{\mathcal{V}}$ due to the “fuzziness” of the visibility surface is important to include, as encoded in $p(k_{\parallel})$. The surface is perpendicular to the photon path to us, so the spatial oscillations are across the surface, giving destructive interference from *both* peaks and troughs for waves with $kR_{\mathcal{V}} > \pi$. There is no destructive interference if the photons are only received from *either* peaks or troughs, but not both, the case if oscillations are along the surface, or if the wavenumbers are small.

5.1.6 Power-law spectra with coherence scales in 3D and 2D

There are a few power spectra for which there is an analytic integral over the Bessel function products. To deal with these we shall restrict ourselves to sources of the $\mathcal{G}_0$ form. The first case we shall consider has a thin shell source concentrated at z_s . For the power law 3D spectrum

$$\begin{aligned} \mathcal{P}_{\mathcal{G}_0\mathcal{G}_0}(k; \chi_s, 0) &= \frac{2\sigma_{\mathcal{G}_0}^2}{2^{\nu_{\mathcal{G}_0}/2}\Gamma(\nu_{\mathcal{G}_0}/2)} (kR_{coh})^{\nu_{\mathcal{G}_0}} e^{-(kR_{coh})^2/2}, \\ \mathcal{C}_{\ell} &= \frac{\Gamma(2 - \nu_{\mathcal{G}_0})}{2^{2-\nu_{\mathcal{G}_0}}\Gamma^2\left(\frac{3-\nu_{\mathcal{G}_0}}{2}\right)} \pi \frac{2\sigma_{\mathcal{G}_0}^2}{2^{\nu_{\mathcal{G}_0}/2}\Gamma(\nu_{\mathcal{G}_0}/2)} \left(\frac{R_{coh}}{\bar{\chi}_s}\right)^{\nu_{\mathcal{G}_0}} \\ &\quad \times \frac{\Gamma\left(\ell + \frac{\nu_{\mathcal{G}_0}}{2}\right)\Gamma(\ell + 2)}{\Gamma(\ell)\Gamma\left(\ell + 2 - \frac{\nu_{\mathcal{G}_0}}{2}\right)}. \end{aligned} \tag{132}$$

This formula is valid as long as the angular scale ℓ^{-1} is large compared with the coherence angle $R_{coh}/\mathcal{R}(\bar{\chi}_s)$.

If the visibility has finite extension, the integral over k can still be done, but in terms of a hypergeometric function which is not very useful. What can be done easily once again is the Fourier calculation for large ℓ :

$$\mathcal{C}_\ell \approx \int d\bar{\chi} \mathcal{V}^2(\bar{\chi}) \pi \frac{2\sigma_{\mathcal{G}_0}^2}{2^{\nu_{\mathcal{G}_0}/2} \Gamma(\nu_{\mathcal{G}_0}/2)} \left(\frac{Q R_{coh}}{\mathcal{R}(\bar{\chi})} \right)^{\nu_{\mathcal{G}_0}} e^{-\frac{1}{2}(Q R_{coh}/\bar{\chi})^2} \times \frac{R_{coh} \sqrt{2} R_{\mathcal{V}}}{\sqrt{Q^2(R_{coh}^2 + 2R_{\mathcal{V}}^2)/\bar{\chi}^2 + (3 - \nu_{\mathcal{G}_0})}}. \quad (133)$$

The $(3 - \nu_{\mathcal{G}_0})$ term in the denominator is an approximation based on a first order expansion in $(k_{\parallel} \bar{\chi}/Q)^2$. With the visibility concentrated at $\bar{\chi}_s$, eq. (133) is just the 2D power law equation, eq. (106), with $\varpi_c = R_{coh}/\mathcal{R}(\bar{\chi}_s)$ and $\nu_{\Delta T} = \nu_{\mathcal{G}_0}$.

For the special case of a 3D white noise distribution, $\nu_{\mathcal{G}_0} = 3$, the exact result including the Gaussian coherence length R_{coh} can be expressed in terms of a modified Bessel function:

$$\mathcal{P}_{\mathcal{G}_0 \mathcal{G}_0} = (2/\pi)^{1/2} \sigma_{\mathcal{G}_0}^2 (k R_{coh})^3 e^{-(k R_{coh})^2/2},$$

$$\mathcal{C}_\ell = \int d\bar{\chi} d\Delta\chi \mathcal{V}(\chi_+) \mathcal{V}(\chi_-) \sqrt{2\pi} \sigma_{\mathcal{G}_0}^2 \ell(\ell+1) \times e^{-(\Delta\chi)^2/(2R_{coh}^2)} e^{-(\chi_+\chi_-)/R_{coh}^2} I_Q(\chi_+\chi_-/R_{coh}^2). \quad (134)$$

If the coherence scale of the blobs is small compared with the cosmological distance at which the visibility is concentrated, the asymptotic expansion of the modified Bessel function, $I_{\ell+1/2}(z) \sim \frac{1}{\sqrt{2\pi z}} e^z (1 - \ell(\ell+1)/(2z) + \dots)$, can be used to simplify the expression:

$$\mathcal{C}_Q \approx \int d\bar{\chi} \mathcal{V}^2(\bar{\chi}) \sqrt{2\pi} \sigma_{\mathcal{G}_0}^2 \left(\frac{Q R_{coh}}{\mathcal{R}(\bar{\chi})} \right)^2 e^{-\frac{1}{2}(Q R_{coh}/\mathcal{R}(\bar{\chi}))^2} \frac{R_{coh} \sqrt{2} R_{\mathcal{V}}}{\sqrt{R_{coh}^2 + 2R_{\mathcal{V}}^2}}.$$

This is the $\nu_{\mathcal{G}_0} \rightarrow 3$ limit of eq. (133), as expected, and if the visibility is concentrated at $\bar{\chi}_s$, it is the $\nu_{\Delta T} = 2$, $\varpi_c = R_{coh}/\mathcal{R}(\bar{\chi}_s)$ version of the 2D law eq. (106). If the visibility is broad, there is a distribution of coherence angles contributing so the final result is only roughly of the Gaussian form.

It might be thought that clouds in our Galaxy could be modelled by such a blob spectrum with no long range correlations, but this is not so. As we saw in section 4.6, $\mathcal{C}_\ell$ for dust-emission and Galactic bremsstrahlung apparently rise as $\ell \ell^{-1}$ [107, 108] at the resolutions they have been observed.

5.2 The primary primary anisotropy effects

For primary scalar anisotropies, we describe here the leading terms associated with the Sachs–Wolfe and Thomson scattering sources, eqs. (25), (26), using the $\nu, \varphi, \Psi_\sigma$ notation for the metric perturbations introduced there. This is just a preview to show which fields are “illuminated” by the visibility of Thomson scattering. The terminology and manipulations of this section will become more familiar after reading section 6.

5.2.1 Sachs–Wolfe, photon-bunching and Doppler sources

By manipulation of eqs. (25), (26), they can be cast into the $\mathcal{G}_{0,1}$ form for a modified field $\tilde{\Delta}_t^{(S)}$ (and for a flat Universe):

$$\begin{aligned} \left[\frac{\partial}{\partial \tau} + \hat{q}^i \frac{\partial}{\partial x^i} \right] \mathcal{T}_C \tilde{\Delta}_t^{(S)} &= \mathcal{V}_C \left[\frac{1}{4} \tilde{\delta}_\gamma - \hat{q}^i \frac{\partial}{\partial x^i} \bar{a}^{-1} \tilde{\Psi}_{v,B} + (\dots) \right] \\ &+ \mathcal{T}_C \frac{\partial}{\partial \tau} [\nu - \varphi + \partial_\tau (\bar{a}^{-1} \Psi_\sigma)], \\ \tilde{\Delta}_t^{(S)} &\equiv \Delta_t^{(S)} + \nu + \frac{\partial \bar{a}^{-1} \Psi_\sigma}{\partial \tau} - \hat{q}^i \partial_i \bar{a}^{-1} \Psi_\sigma, \quad \mathbf{v}_B = -\bar{a}^{-1} \nabla \Psi_{v,B}, \end{aligned} \quad (135)$$

$$\mathcal{G}_{nSW,0} + \mathcal{G}_{\delta_\gamma,0} \equiv \frac{1}{4} \tilde{\delta}_\gamma \equiv \frac{1}{4} \delta_\gamma + \nu + \frac{\partial}{\partial \tau} \bar{a}^{-1} \Psi_\sigma \equiv F_\Phi(k, \tau) \Phi_N / 3, \quad (136)$$

$$\mathcal{G}_1 \equiv \bar{a}^{-1} \tilde{\Psi}_{v,B} \equiv \bar{a}^{-1} (\Psi_{v,B} + \Psi_\sigma) \equiv F_v(k, \tau) \tau \Phi_N / 3, \quad (137)$$

$$\mathcal{G}_{iSW,0} = \frac{\partial^2}{\partial \tau^2} \bar{a}^{-1} \Psi_\sigma + \dot{\nu} - \dot{\varphi} = F_{\dot{\Phi}}(k, \tau) \Phi_N / 3. \quad (138)$$

Here δ_γ is the photon energy density perturbation, $\mathbf{v}_B$ is the baryon velocity and $\Psi_{v,B}$ is the baryon velocity potential. The $(\dots)$ refer to source terms driven by the quadrupole and by polarization terms. These are subdominant and can be ignored in rough treatments. There are three types of terms multiplied by $\mathcal{V}_C$. The “Doppler” source is $\mathcal{G}_1$. The $\mathcal{G}_0 = \tilde{\delta}_\gamma/4$ term has two parts, metric terms which give the “naive” Sachs–Wolfe effect, in particular the famous $\Phi_N/3$, where Φ_N is the gravitational potential, if Φ_N is constant – which it rarely is. The $\delta_\gamma/4$ term describes the amount of “photon bunching” when decoupling releases the photons, which gives the equally famous $\frac{1}{3} \delta \rho_B / \bar{\rho}_B$ term if the entropy per baryon is constant – which it is not. The source term multiplied by $\mathcal{T}_C$ is known as the integrated Sachs–Wolfe or Rees–Sciama effect. While the Doppler and integrated Sachs–Wolfe sources do not depend upon the gauge which is chosen, the relative amount which is attributed to photon bunching and the naive Sachs–Wolfe effect does, although the $\tilde{\delta}_\gamma/4$ combination is gauge-invariant.

All of the effects have been normalized to $\Phi_N(k, \tau_0)/3$ through form factors $F_\Phi(k)$, $F_v(k)$ [2] and $F_{\dot{\Phi}}$, to emphasize that the magnitude of Φ_N controls the magnitude of $\Delta T/T$ – although it is the explicit form of these order-unity form factors that define the bumps and wiggles of the spectral shape. The statistical distribution of $\Delta T/T$ is also completely determined by the statistical distribution of the stochastic field Φ_N , with the spatial Fourier transforms $\{F_\Phi, F_v, F_{\dot{\Phi}}\}(\mathbf{x}, \tau)$ defining nonstochastic time-dependent fields which are convolved with $\Phi_N(\mathbf{x}, \tau_0)$ to give the sources.

5.2.2 Longitudinal and synchronous pictures of the Sachs–Wolfe effect

The longitudinal gauge has $\Psi_\sigma = 0$. The metric is characterized by ν_L , which is the closest analogue to the perturbed Newtonian potential Φ_N ; and φ_L goes

to $-\Phi_N$ once anisotropic pressure forces can be neglected, which it can after $\bar{a}_{eq}$ and $\bar{a}_{dec}$. In the regime in which nonrelativistic (nr) matter dominates the evolution Φ_N is constant (in linear perturbation theory). The velocity potential for baryons in that gauge is $\Psi_{v,BL} \equiv \tilde{\Psi}_{v,B}$ and the velocity potential for cold dark matter is $\Psi_{v,cdmL} \equiv \tilde{\Psi}_{v,cdm}$. In the nr -dominated regime, $\bar{a}^{-1}\tilde{\Psi}_{v,cdm} = \frac{1}{3}\Phi_N\tau$. Compton drag stops the baryons from following the nr dark matter flow, but once the photons do let go, $\mathcal{G}_1$ also approaches $\frac{1}{3}\Phi_N\tau$. For normal recombination, there is by this time no differential visibility left; the $\mathcal{G}_1$ -field is determined by the earlier baryon physics, i.e., the transition through the optical-depth-unity regime of tight-coupling to damped-streaming. In universes with early reionization, much of the “visible” region can be after the Compton drag lets go and $\mathcal{G}_1 \approx \frac{1}{3}\Phi_N\tau$ can be a good approximation.

In the synchronous gauge, ν is set to zero and the constant time surfaces chosen to be those on which cold dark matter is at rest; the synchronous gauge metric variable $\Psi_{\sigma S}$ is then just $\tilde{\Psi}_{v,cdm}$, and the metric part of $\mathcal{G}_{nSW,0} + \mathcal{G}_{\delta\gamma,0}$ gives $\frac{1}{3}\Phi_N$ in the nr -dominated regime, i.e., the classic naive Sachs–Wolfe term. This suggests we define the photon bunching source $\mathcal{G}_{\delta\gamma,0}$ to be $\frac{1}{4}\delta\gamma_S$, which is then a gauge-invariant term, with the remainder of eq. (136) defining the naive Sachs–Wolfe source $\mathcal{G}_{nSW,0}$. In the oft-used longitudinal gauge, the correct $\Phi_N/3$ behavior is obtained only when a piece of $\delta_{\gamma L}/4$, photon bunching as viewed in this gauge, is added to $\nu_L = \Phi_N$: we show in section 6 that $\delta_{\gamma L}/4 = \delta_{\gamma S}/4 - \bar{H}\Psi_\sigma$, which becomes $\delta_{\gamma S}/4 - \frac{2}{3}\Phi_N$ in the nr -dominated regime.

The integrated Sachs–Wolfe term at late times becomes $\mathcal{T}_C 2\dot{\Phi}_N$. Thus $F_{\dot{\Phi}} = 6\dot{\Phi}_N/\Phi_N$. In the nr -dominated regime, it vanishes for linear perturbations. Nonlinear clustering generates nonzero $\dot{\Phi}_N$. Even though the potential change may not be very great, the factor of 6 enhances the impact on the CMB. When the equation of state changes from nr -dominance, $\dot{\Phi}_N$ no longer vanishes. This is the source for the relative upturn in $\mathcal{C}_\ell$ in the vacuum-dominated model in fig. 7 [110]. It is also rarely true that a_{eq} is so much less than a_{dec} that changes in Φ_N around recombination can be ignored. In that case, we can absorb it into a $\mathcal{V}_C$ style source by replacing $F_{\dot{\Phi}}$ by $F_{\dot{\Phi}} + \tau_C F_{\dot{\Phi}}$, where $\tau_C^{-1} = \bar{n}_e \sigma_T \bar{a} = \mathcal{V}_C/\mathcal{T}_C$.

5.2.3 Differential power spectrum and form factors

We can apply the machinery leading to eq. (126) to get the power spectrum. Let us ignore late-time integrated Sachs–Wolfe effects associated with nonzero Λ , etc. so we can use just the visibility $\mathcal{V}_C$ sources. Since we are also interested in low ℓ , we use the Bessel function product $j_\ell^{(A)} j_\ell^{(B)}$ for $\langle j_\ell^{(A)} j_\ell^{(B)} \rangle$:

$$\frac{d\mathcal{C}_\ell}{d\ln k} = 2\ell(\ell+1) \frac{1}{9} \mathcal{P}_{\Phi_N}(k, \tau_0) \int d\bar{\chi} \mathcal{V}_C^2(\bar{\chi}) p(\bar{k}_\parallel) [(F_\Phi(k, \bar{\tau}) + \tau_C F_{\dot{\Phi}}(k, \bar{\tau})) j_\ell(k\mathcal{R}(\bar{\tau})) + k\bar{\tau} F_v(k, \bar{\tau}) j'_\ell(k\mathcal{R}(\bar{\tau}))]^2. \quad (139)$$

We have seen that the visibility $\mathcal{V}_C$ is roughly a Gaussian in conformal time with width $R_{\mathcal{V},dec}$ for normal recombination. If the 3D source functions do not

change much over $\mathcal{V}_C$, the form factors can be evaluated at $\bar{\tau}$ and even at τ_{dec} ; otherwise an average over the shell is needed, defining effective form factors which can also absorb $\tau_C F_\Phi$ and the last scattering surface fuzziness damping associated with $p(\bar{k}_\parallel)$.

The goal of analytic approaches is to use approximate form factors like these to understand the physics defining the basic features of the spectra and to provide a tool for rapid estimation of $\mathcal{C}_\ell$. It has been developed in various approximations in [2, 143, 269, 270, 271]. Hu and Sugiyama [271] have included the most effects, in particular the time variation of Φ_N that arises because a_{eq} is not very far from a_{dec} ; by doing so they obtain remarkably good reproductions of the spectra derived using full Boltzmann transport codes, within about 10% or so even at high ℓ . Here I shall just use a simple analytic result [2, 201] to illustrate the physics that determines the nature of the oscillations that translate into the $\mathcal{C}_\ell$ peaks and troughs, but caution that the more elaborate scheme of [271] is needed if one wants a quantitative tool. (It was, for example, used in [161] to rapidly calculate $\mathcal{C}_\ell$ for a large parameter set to assess how well parameters could be determined in idealized all-sky satellite experiments.)

Earlier than decoupling, the photons and baryons are so tightly coupled by Thomson scattering that they can be treated as a single fluid with shear viscosity $(4/(15f_\eta))\bar{\rho}_\gamma \bar{a}\tau_C$, zero bulk viscosity, thermal conductivity $\kappa_\gamma = (4\rho_\gamma/(3T_\gamma))\bar{a}\tau_C$ and sound speed $c_{s,(\gamma+B)} = (c/\sqrt{3})[1 + 3\bar{\rho}_B/(4\bar{\rho}_\gamma)]^{-1/2}$, lowered over the $(c/\sqrt{3})$ for a pure photon gas because of the inertia in the baryons. Here f_η is a parameter which depends upon the approximations that are made to treat the photons: it is 3/4 if all effects are included, 9/10 for unpolarized photons and 1 if the angular dependence in the Thomson cross section is also ignored. These results are derived in Appendix C.3.1.

Let us assume constant Φ_N and $\bar{\rho}_B \ll \bar{\rho}_\gamma$ through decoupling. The WKB solution of the tight coupling fluid equations is, for $\tau < \tau_{dec}$,

$$F_\Phi \approx e^{-\frac{1}{2}(\sigma_D k\tau)^2} (3\bar{c}_s^2)^{1/4} \cos(k\bar{c}_s\tau),$$

$$F_v \approx e^{-\frac{1}{2}(\sigma_D k\tau)^2} (3\bar{c}_s^2)^{1/4} \frac{\sin(k\bar{c}_s\tau)}{k\bar{c}_s\tau}.$$

Also $F_\dot{\Phi} \approx 0$. σ_D is a parameter describing Silk damping. In this tiny baryon number limit, the sound speed is $\bar{c}_s = (c/\sqrt{3})$, but for finite $\bar{\rho}_B$, $\bar{c}_s$ is a suitable time-average of $c_{s,(\gamma+B)}$ and there is also a weak amplitude-diminisher, $(3\bar{c}_s^2)^{1/4}$.

The WKB solution for $\delta_{\gamma S}/(-\Phi_N) = 1 - F_\Phi$ shows $\delta_{\gamma S}$ growing outside the horizon like τ^2 ; the horizon is “entered” for photons when $k\bar{c}_s\tau \approx \pi/2$; and thereafter oscillations spaced equally in $k\bar{c}_s\tau$ should be expected in the evolution of individual k -modes. Some examples of this behavior for different k ’s are shown in fig. 19. By contrast, the view of the density fluctuations in the longitudinal gauge is $\delta_{\gamma L}/(-\Phi_N) \approx 5/3 - F_\Phi$, dominated by the constant 2/3 part which swamps the rising part. This emphasizes the care that must be taken in choosing which variables to integrate – no matter what the initial gauge choice.

The phase of the waves as they hit the narrow recombination band, $k\bar{c}_s\tau_{dec}$, determines the oscillations in $\mathcal{C}_\ell$ that appear in fig. 7. The combination of viscous and fuzziness damping diminishes the amplitude of the Doppler peaks. Because the oscillations are in both δ_γ and $\tilde{\Psi}_{v,B}$, both contribute to the detailed structure.

In section 5.1.2, a high ℓ form of $\mathcal{C}_\ell$ was given, eq. (129), and a further simplification associated with isotropizing the total source power spectrum was described. For the limiting WKB case, the isotropized source-power evaluated at $\tau = \tau_{dec}$ is $\frac{1}{9}\mathcal{P}_{\Phi_N}e^{-(\sigma_D k\tau)^2}(3\bar{c}_s^2)^{1/2}(1 + \frac{3\bar{\rho}_B}{4\bar{\rho}_\gamma}\sin^2(k\bar{c}_s\tau))$. This illustrates that in the instantaneous recombination limit with no damping and tiny baryon abundance, $\mathcal{P}_{\Phi_N}/9$, the naive Sachs–Wolfe effect, is recovered. But this is obviously not what one sees in the figures. It is in the finite Ω_B effects, the time dependence of Φ_N and even the differences between j_ℓ and j'_ℓ that the dramatic hills and valleys of $\mathcal{C}_\ell$ owe their origin – and it is with just those factors that the $\mathcal{C}_\ell$ -landscape can be estimated accurately.

5.2.4 Damping

The parameter σ_D is an integral of the damping rate involving the shear viscosity and thermal conductivity. In the WKB limit, it is given by eq. (357) in section C.3.1: for CDM models with low Ω_B , σ_D is roughly 0.02–0.03 with polarization included, which enhances damping, and is about 10% lower if the radiation is assumed to be unpolarized. With $\Omega_B h^2 = 0.0125$ preferred by Big Bang nucleosynthesis, $\sigma_D \approx 0.02$, and the same rough value is obtained in the limit of large Ω_B . Of course the tight coupling equations break down as the radiation passes through decoupling, so it is better to treat σ_D as a phenomenological factor, but matching to numerical results for Silk damping in baryon dominated models also gives the 0.02 estimate for CDM-model parameters [2]. The damping acting on $\Delta T/T$ due to fuzziness of the last scattering surface is $e^{-(k_\parallel R_{\mathcal{V}_C,dec})^2/2}$, while that from Silk damping is $e^{-(k\sigma_D\tau_{dec})^2/2}$. From eq. (69), we have $R_{\mathcal{V}_C,dec}/(\sigma_D\tau_{dec}) \approx \sigma_{a,dec}/(2\sigma_D)$ which is ~ 2 for the examples of fig. 3(b).

The fuzziness damping acts only on $k_\parallel$, while the WKB viscous damping acts on k . Effective isotropized fuzziness filters are found by expanding in $k_\parallel R_{\mathcal{V}_C,dec}$ and angle-averaging [2], which reduces the effective filter to $R_{\mathcal{V}_C,dec}/\sqrt{3}$; this makes the WKB and fuzziness damping values similar in magnitude. The WKB tight-coupling solution does in fact calculate a version of fuzziness damping acting on δ_γ , along with other transport effects, but the $k_\perp$ – $k_\parallel$ asymmetry is obscured by the truncation of the ℓ -hierarchy at such low ℓ : up to $\approx \tau_{dec}$, higher moments are strongly damped, but as the photons pass through $\zeta_C = 1$, fuzziness damping in this “scattering atmosphere” occurs. At decoupling, τ_C is only 5% of τ_{dec} .

5.2.5 Early reionization form factors

If we assume early reionization, and a decoupling redshift (where the visibility peaks) in the nr -dominated regime and below the epoch at which Compton drag lets up, $\lesssim 200$, then we have $F_v = 1$. For small $k\tau$ and adiabatic perturbations, we expect to have $F_\Phi = 1$ in this nr -regime, damping as $k\tau$ increases, but not exponentially once $\tau_C \sim \bar{a}^2$ grows to a point where tight-coupling breaks down. What one does get is a photon density gradient responding to the residual Compton drag; a converging baryon flow increases $\tilde{\delta}_\gamma/4$, a diverging one diminishes it: the net effect for large $k\tau$ is $F_\Phi \approx -\tau/\tau_C$, which falls like $\bar{a}^{-3/2}$ for a fully ionized medium. Thus in reionized adiabatic models, one expects a normal Sachs–Wolfe behavior at small ℓ , with a velocity-induced extra piece pushing it up a bit at larger ℓ , both being diminished by an overall large fuzziness factor, typically with $R_{\nu_C} \sim 0.3\tau$ (section 3.4.2). The high- $k\tau$ part of F_Φ has been shown to augment the velocity-induced term by an order-unity factor [217, 218].

5.2.6 The isocurvature effect on low multipoles

If the perturbation mode is isocurvature rather than adiabatic, the fluctuations are initially perturbations in the entropy (per CDM particle for isocurvature CDM perturbations or per baryon) without accompanying curvature perturbations. For these, there is another effect which amplifies F_Φ to 6, the isocurvature effect. Let $\delta_s \equiv \frac{3}{4}\delta_\gamma - \delta_x$ denote the relative perturbation in the entropy per x -particle, where $x = cdm, B$. To have no energy density perturbation in the $k\tau \rightarrow 0$ limit and yet have a nonzero δ_s , we must have $\delta_{\gamma S} \approx \frac{4}{3}\delta_s(1 + \frac{4}{3}\bar{\rho}_{er}/\bar{\rho}_{nr})^{-1} \approx -(\bar{\rho}_{nr}/\bar{\rho}_{er})\delta_{xS}$, where $\bar{\rho}_{nr}$ is the density of non-relativistic particles, $\bar{\rho}_{er}$ is the density of relativistic particles and it has been assumed for this illustration that all nr and er particles will have the same relative density perturbations, δ_{xS} and $\delta_{\gamma S}$, respectively. At very early times, $\delta_{\gamma S}$ is tiny, with the entropy perturbation being carried by the x -particles, but as $\bar{\rho}_{nr}/\bar{\rho}_{er} = \bar{a}/\bar{a}_{eq}$ grows from unity to $\sim 10^4$, the perturbation is primarily carried by the radiation. When a given wave enters the horizon, δ_{xS} ceases declining, and begins to grow after τ_{eq} via the usual Jeans instability. This diminishment of δ_x at low k translates to a sharp bend in the isocurvature CDM transfer function at $k \sim \tau_{eq}^{-1}$, falling as k^2 at low k , but being unity at high k . The reciprocal impact of this on $\delta_{\gamma S}$ gives the isocurvature effect.

It is easiest to see why $F_\Phi=6$ using the equation for $\tilde{\delta}_\gamma(k, \tau)/4$ in the $k\tau \lesssim 1$ limit, the angle-average of the $\tilde{\Delta}_t$ transfer equation (and eq. (356) of Appendix C.3.1). Since $\delta_{\gamma S}$ is initially nearly zero, we have $\tilde{\delta}_\gamma(k, \tau)/4 \approx (\nu - \varphi + \partial_\tau[\bar{a}^{-1}\Psi_\sigma])(k, \tau)$. Both sides turn out to be gauge-invariant and so the right-hand side is the quantity $\nu_L - \varphi_L$; but ν_L is Φ_N by definition, and $\varphi_L \approx -\Phi_N$, which becomes exact if there are no anisotropic pressure terms; i.e., $\tilde{\delta}_\gamma(k, \tau)/4 \approx 2\Phi_N$ for low k : i.e., $F_\Phi \approx 6$ [215, 174]. The isocurvature effect can also be expressed in terms of the initial entropy fluctuation,

$2\delta_s/5 = \delta_s/3 + \Phi_N/3$, where the gravitational potential is related to the initial entropy perturbation by $\Phi_N = \delta_s/5$.

The $F_\Phi = 6$ factor is so large that isocurvature theories with nearly scale invariant spectra are strongly ruled out by the observations, [215] although there is still some room for them to contribute at a subdominant level to adiabatic scalar perturbations [233] for dark matter dominated models. For isocurvature baryon perturbations, there are significant features in the density transfer function for scales $k\tau_{eq} \sim 1$ [131, 248, 216]. Although the isocurvature effect does lead to nearly scale invariant spectra being strongly ruled out, isocurvature baryon models with arbitrary spectra (which are steep with $n_s \approx 0$ corresponding to seed models) and arbitrary ionization histories are usually adopted; with this freedom, the case against them is strong but not yet definitive [216, 218, 219, 243]. (See also fig. 18).

5.3 Secondary anisotropies

Secondary anisotropies are non-Gaussian, with the spectral power concentrated in hot and/or cold spots on the sky, rather than being democratically distributed as it is for Gaussian anisotropies. The power spectra are instructive since they do tell us what the best scales to probe are, but they are far from the whole story. Examples of $\mathcal{C}_\ell$ for the ambient SZ effect from clusters and dust-emission from primeval galaxies are shown in fig. 7. The dual nature of the power spectra for dust anisotropies can be understood using the methods of section 5.1, as sketched in section 5.3.1. The power spectrum can be used to calculate rms anisotropies. In the mid to late 80s, estimates made on the strength of the SZ effect from clusters concentrated on the distribution of y and the rms variations as a function of beam, using simplified peak or Press–Schechter based models to calculate the abundance and clustering of the sources [2, 112, 113]. This work showed that for inflation-based CDM-like models, the rms SZ fluctuations would be quite small, well below 10^{-5} . Similar techniques were applied to estimating the rms anisotropies in the far infrared and sub-mm from dusty primeval galaxies [81, 2, 42]. More powerful methods were developed in the nineties [117, 118, 120, 121, 122] to address how non-Gaussian the signals would actually be (e.g., figs. 15, 16).

5.3.1 Sample secondary anisotropy power spectra

For secondary anisotropies, the broad visibility case of section 5.1.5 is the appropriate limit for the angular power spectrum. For high $Q \equiv \ell + \frac{1}{2}$,

$$\mathcal{C}_\ell \approx \int d\bar{\chi} \mathcal{V}^2(\bar{\chi}) (Q/\mathcal{R}(\bar{\chi}))^{-1} \pi \mathcal{P}_{\mathcal{G}_0 \mathcal{G}_0}(k = Q/\mathcal{R}(\bar{\chi}); \bar{\chi}, 0). \quad (140)$$

It is often suitable to adopt a shot noise model for the distribution of the random source-field $\mathcal{G}_0$ [2, 42]: this consists of (1) a class of objects defined by parameters $\mathbf{C}$ (e.g., mass, luminosity, X-ray temperature) whose positions are

specified by a random point process $n_{C*}(\mathbf{r}) = \sum_{j \in C} \delta^{(3)}(\mathbf{r} - \mathbf{r}_j)$ with the sum over points j satisfying the conditions C ; and (2) profiles for $\mathcal{G}_0$ centered at each point, $g(\mathbf{r}|C, \tau)$. n_{C*} is a comoving density if $\mathbf{r}_j$ are comoving positions. The points C could define galaxies, clusters, N -body groups, the centers of cosmic explosions, ..., and the profiles g may be asymmetrical (e.g., filaments, pancakes). In the “peak-patch picture” of [2, 68, 117, 120], the shots are equated with specially selected peaks of the smoothed linear density field.

The source function $\mathcal{G}_0$ for a shot noise model is the sum of convolutions:

$$\begin{aligned} \mathcal{G}_0(\mathbf{r}, \tau) &= \sum_C \int d^3\mathbf{r}' g(\mathbf{r} - \mathbf{r}'|C) n_{C*}(\mathbf{r}, \tau), \\ \tilde{\mathcal{G}}_0(\mathbf{k}, \tau) &= \sum_C \tilde{g}(\mathbf{k}|C, \tau) \tilde{n}_{C*}(\mathbf{k}, \tau), \quad \tilde{g}(\mathbf{k}|C, \tau) \equiv g_c(\tau|C) V_C \mathcal{F}_g(\mathbf{k}|C). \end{aligned} \quad (141)$$

We have separated $\tilde{g}(\mathbf{k})$ into a central value g_c , a weighted volume of the region $V_C \equiv \int g(\mathbf{r}) d^3\mathbf{r}/g_c$, and a form factor $\mathcal{F}_g(k)$ which is dimensionless and equal to unity at $k = 0$ by construction. Although g can be considered to be a random field as well, it is usual to just assume fixed profile shapes. An example of some interest is a truncated spherical β -profile,

$$g(\mathbf{r}) = g_c (1 + r^2/r_{core*}^2)^{-3\beta/2} \vartheta(R_{C*} - r),$$

with core radius r_{core} and truncation radius R_C ; the $*$ denotes comoving quantities (e.g., $r_{core*} = \bar{a}^{-1} r_{core}$). This form is widely used to model the gas density in clusters and thus g for the SZ effect if the cluster is isothermal. Fits to X-ray profiles give $\beta \approx 2/3$. An approximate form factor which roughly takes the truncation into account is

$$\mathcal{F}_g(k|C) \approx \frac{\exp[-kr_{core*}]}{((kR_{C*})^2 + 1)^{\frac{3}{2}(1-\beta)}}. \quad (142)$$

For small k , $\mathcal{F}_g(k|C) \approx 1$ and $g_c(\tau|C) V_C$ is $\propto \langle T_e \rangle B_C$ for the SZ effect from clusters, where B_C is the cluster baryon number, while for dust at fixed temperature, it is $\propto$ the mass of dust in the galaxy.

The power spectrum for $\mathcal{G}_0$ can be written in terms of the cross-correlation power spectra for the shots:

$$\begin{aligned} \mathcal{P}_{\mathcal{G}_0\mathcal{G}_0}(k) &= \sum_{C_1} |\tilde{g}(k|C_1)|^2 \frac{\bar{n}_{C_1*} k^3}{2\pi^2} \\ &+ \sum_{C_1 C_2} \tilde{g}(k|C_1) \tilde{g}^*(k|C_2) \bar{n}_{C_1*} \bar{n}_{C_2*} \mathcal{P}_{C_1 C_2}^{cc}(k), \end{aligned} \quad (143)$$

where the tilde denotes Fourier transform. The shot correlation power has been decomposed into a Poisson contribution $\delta_{C_1 C_2} (k \bar{n}_{C_1}^{-1/3})^3 / (2\pi^2)$ describing the self-correlation of the discrete objects and a continuous correlation piece

$\mathcal{P}_{C_1 C_2}^{cc}$ describing the clustering of the objects. In a linear biasing approximation, we would have $\mathcal{P}_{C_1 C_2}^{cc}(k, \tau) = b_{C_1}(\tau) b_{C_2}(\tau) \mathcal{P}_{\rho\rho}(k, \tau)$, where the $b_{C_j}(\tau)$ are biasing factors and $\mathcal{P}_{\rho\rho}(k, \tau)$ is an underlying mass density power spectrum. Even if such a relation were to hold for low k one would expect considerable modification at high k .

For the Poisson piece, the contribution from an object which subtends an angle $\theta_C \sim R_{C*}/\mathcal{R}(\bar{\chi})$, whose core subtends $\theta_{core} \sim r_{core*}/\mathcal{R}(\bar{\chi})$, is

$$\sim Q^2(1 + Q^2\theta_C^2)^{-3(1-\beta)} e^{-2Q\theta_{core}}, \quad (144)$$

i.e., white noise (Poissonian) for small Q , $Q^{3\beta-1}$ for $Q > \theta_C^{-1}$, with an exponential suppression at very high Q . For the continuous clustering contribution, the overall amplitude is usually lower and the shape is multiplied by $Q^{n_{\rho,eff}(Q/\bar{\chi})}$, where $n_{\rho,eff}(k)$ is the local index of $\mathcal{P}_\rho$ (i.e., $\sim k^{3+n_{\rho,eff}}$). For angular scales $> \theta_C$, it can often dominate, $\sim Q^{2+n_{\rho,eff}(k)}$, cf. the $\sim Q^2$ Poisson term.

Notice that if we use a Gaussian profile for the shots and have a narrow visibility at redshift z_s , the $\mathcal{C}_\ell$ we get from the Poisson piece is a Gaussian coherence spectrum with coherence angle $\theta_c \approx \varpi_c = \sqrt{2}(1+z_s)R_C/\mathcal{R}(\chi_s)$, i.e., eq. (106) with $\nu_{\Delta T} = 2$.

5.3.2 Anisotropy power from dusty primeval galaxies

The BCH2 [42] spectra shown in fig. 15 show the basic features: a Gaussian radial profile of scale $R_G = 10$ kpc for the dust in galaxies defines the cutoff at high ℓ , the amplitude is determined by the galaxy (shot) density, here chosen to be that of bright galaxies $n_{G*} = 0.02 \text{ (h}^{-1} \text{ Mpc)}^{-3}$. The continuous clustering piece dominates at lower ℓ .³

The spectra clearly show that to get the maximum signal one would like to probe the shot noise power, i.e., have a small beam. This is misleading because a small beam may be unlikely to capture a galaxy. Large beams have too many galaxies in them to give much shot-noise anisotropy. Clustering dominates the signal there. Figure 15 emphasizes how useful very small angle anisotropy experiments can be for detecting high redshift dust emission from primeval galaxies even for cases which fall well below the FIRAS bounds. The dust maps in fig. 15 were constructed using the peak-patch method to identify the high redshift galaxies [82]. The most promising instrument coming on line for this is SCUBA on JCMT [114], with $12''$ resolution, and the ability to probe a number of frequencies, in particular an atmospheric window around $850\mu\text{m}$. The peak model shown in the figure gives rms anisotropies $(\Delta\nu I_\nu)_{rms}$ of $0.2S$ – as measured in units of $10^{-6} \text{ erg cm}^{-2} \text{ s}^{-1} \text{ sr}^{-1}$. Assuming galaxies

³The particular model chosen hybridized a biased linear density power spectrum shape for small k and a nonlinear power law contribution for high k , with the two joined at k_{NL} where the power is unity. The shape change in the graph is a result of this rough approximation. The maximum occurs where $(Q/\mathcal{R}(\bar{\chi}))^{-1} \pi \mathcal{P}_{\mathcal{G}_0 \mathcal{G}_0}(k = Q/\mathcal{R}(\bar{\chi})) \sim Q^{2+n_{\rho,eff}(k)}$ is maximum, at $n_{\rho,eff}(k) \sim -2$, which occurs at $\sim 0.5 \text{ h}^{-1} \text{ Mpc}$ for the CDM spectrum, and on somewhat larger scales for adiabatic models with nonzero Λ .

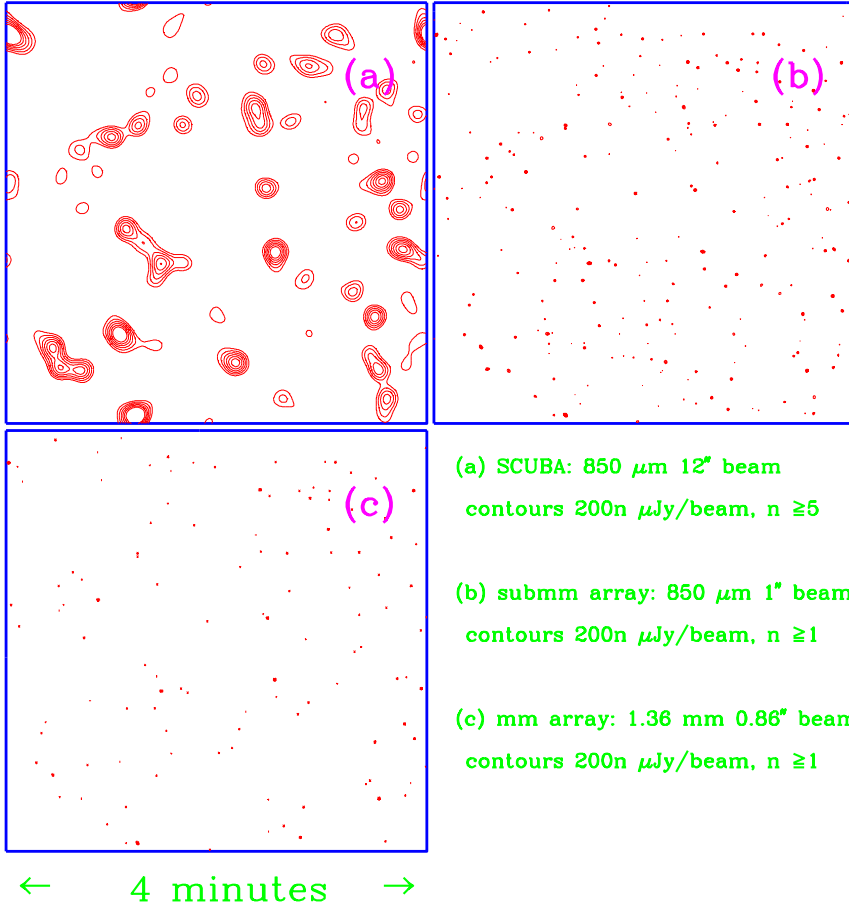


Figure 15: Illustration of what the sub-mm emission might look like from primeval galaxies in a $\sigma_8 = 0.7$ CDM model. (a) A $4' \times 4'$ contour map for dust-emission from primeval galaxies at $z \sim 5$ convolved with a 12'' beam appropriate for the 855 μ 37-pixel SCUBA array. The minimum contour is $1000S \mu\text{Jy}/\text{beam}$ and subsequent contours increase linearly in $250S \mu\text{Jy}/\text{beam}$ steps. SCUBA has a $2' \times 2'$ FOV and is expected to achieve $470 \mu\text{Jy}/\text{beam}$ at the 1σ level in just one hour of integration. (b) Shows the same map seen with a 1'' beam with $250S \mu\text{Jy}/\text{beam}$ contours for an 800 μ sub-mm array. (c) Shows the map with a 0.86'' beam with $200S \mu\text{Jy}/\text{beam}$ contours for a 1.36 mm array. S is a scaling factor which is 1 if all “bright galaxies” have Arp 220 luminosities down to redshift 4. To satisfy the FIRAS bound (Fixsen et al. 1996), apparently either $S \lesssim 0.1$ is required, or $\lesssim 0.1$ of the sources present could be bursting. (1 Jansky $\equiv 10^{-26} \text{W m}^{-2} \text{Hz}^{-1}$, hence $I_\nu = (\lambda/3 \mu\text{m})(\delta E(\lambda)/E_{\text{cmb}}) \text{MJy sr}^{-1} \text{sr}^{-1}$)

with a density of $\sim 0.02 (\text{h}^{-1} \text{ Mpc})^{-3}$ (the current density of “bright” galaxies) do all of the emission in a biased CDM model for the other (BCH2) models given in table 1, the rms anisotropies that SCUBA would see (assuming a double differencing mode) would be quite large: 1, 4, 4, 0.3 in the above units, corresponding to $\Delta T/T$ of $(7, 20, 12, 2) \times 10^{-5}$ for BCH2 models 8, 11, 14, 13, respectively. These anisotropies should be compared with the current 800μ JCMT 95% C.L. upper limit for an $18''$ beam of 3.4×10^{-3} [115, 42], and the 1300μ IRAM millimeter 95% C.L. upper limit for an $11''$ beam of 2.4×10^{-4} [116, 42]. (The signal would also have fallen off from the 800μ value by 1300μ .) Of course the map fig. 15 also demonstrates that the rms emission is somewhat misleading for SCUBA since it is concentrated in bright patches; and it is totally misleading for the interferometric arrays.

5.3.3 SZ and nonlinear Thomson scattering from clusters

The most direct way to make maps of secondary anisotropies is to do hydrodynamical simulations, then calculate the line-of-sight integrals of $\mathcal{G}$ through the computational volume. It is difficult to make large enough simulations and still get the resolution needed to treat the structure in the objects. A good example of the current state of the art is given in [122], in which SZ maps were constructed by using many hydrodynamical simulations. The “peak-patch picture” that Steve Myers and I developed [68] allows us to determine the spatial distribution and properties of rare events in the medium such as clusters over very large volumes of space by identifying them with carefully selected peaks of the linear density field [68, 117, 120]. Peak-patch catalogues accord well with N -body cluster and group catalogues, both statistically and spatially, reproduce well the gross internal properties such as mass and internal energy, and do reasonably well at getting the bulk flow of the rare events [68]. The maps in fig. 16 were constructed in this way, finding all clusters and groups in a $16^\circ \times 16^\circ$ patch over a region extending out to redshift 1.5 for a $\sigma_8 = 0.7$ CDM model. A truncated β profile was used with $\beta = 2/3$ to give the gas density distribution, the core radius was calibrated with X-ray observations, and the gas was assumed to be isothermal. σ_8 controls the overall abundance of rich clusters: maps such as these look dramatically different with even small variations. The shape of the power spectrum controls the poor-to-rich cluster ratio. The σ_8 was chosen so the cluster abundances as a function of temperature roughly agreed with X-ray observations. A COBE-normalized CDM model has $\sigma_8 \sim 1.2$ (eq. (222)) and far too many large clusters, but, for example, a COBE-normalized $\Omega = 1$ model with a mixture of hot and cold dark matter (see section 7.3) has $\sigma_8 \sim 0.7$, fits the X-ray data reasonably well, and has a similar appearance to the CDM model shown here, albeit with a smaller poor-to-rich cluster ratio [120].

The SZ, moving cluster and primary maps of fig. 16 have the following minima, maxima, mean offsets, and rms, in units of 10^{-6} : (a) $(-47, 0, -2, 3)C_{SZ}$; (c) $(-8, 6, -0.04, 0.4)C_V$; (d) $(-53, 48, -0.06, 18)$. Thus the SZ effect is com-

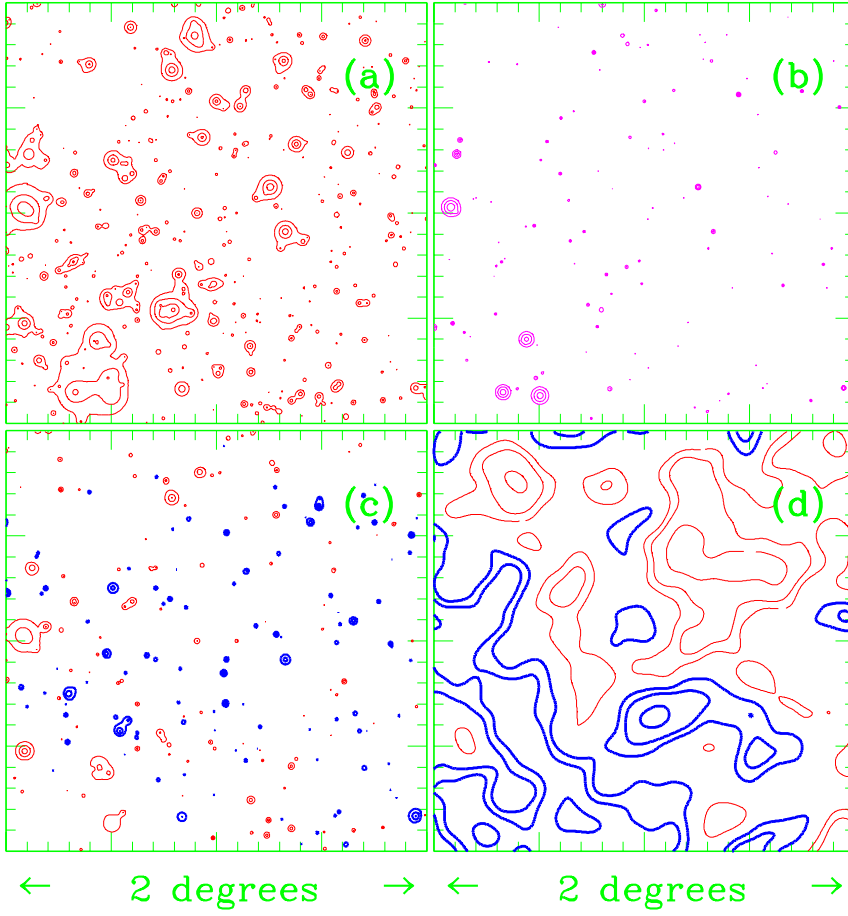


Figure 16: $2^\circ \times 2^\circ$ maps for a $\sigma_8 = 0.7$ CDM model that could be probed by the Cosmic Background Imager (CBI) being built by Caltech: an 8 small-dish interferometer to map scales from $\sim 2'-20'$, with optimal sensitivity $\gtrsim 5'$, using HEMTs to cover frequencies 30–40 GHz, with a 15 GHz channel to help to remove contamination. (a) Shows the SZ effect for 30 GHz, with contours $-5 \times 10^{-6} C_{SZ} \times 2^{n-1}$; (b) the associated ROSAT map (0.1–2.4 keV), with contours $10^{-14} C_X \times 2^{n-1} \text{ erg cm}^{-2} \text{ s}^{-1}$, so the minimum contour level is similar to the ROSAT 5σ sensitivity for long exposure pointed observations; (c) the Thomson scattering anisotropy induced by the bulk motion of the clusters, with contours now $\pm 1.25 \times 10^{-6} C_V \times 2^{n-1}$, $C_V \approx 1.2$; (d) primary anisotropies, with contour levels at $\pm 10^{-5} \times 2^{n-1}$. Negative contours are light and dotted. The C_{SZ} , C_X and C_V are order-unity correction factors.

petitive with the much larger primary anisotropies expected in this model only in the cores of clusters; and the moving-cluster anisotropies are disappointingly small, even when nonlinear corrections are included. For the X-rays, the map flux characteristics are (b) $(0, 12, 0.05, 0.2) \times 10^{-14} C_X \text{ erg cm}^{-2} \text{ s}^{-1}$. Using the information in a deep field cluster catalogue such as (b) will clearly be invaluable for separating SZ from primary. Even so, since the true sky will be the sum of (a), (c), and (d) – plus Galactic and extragalactic synchrotron and bremsstrahlung sources for CBI, and plus dust for SAMBA, some possibly cold – separation using many well-spaced frequency bands will be essential and also quite difficult.

5.3.4 Single-cluster observations of the SZ effect

There has been dramatic improvement in observations of the SZ effect from individual clusters in the last few years, with the promise of much more to come. The effect has now clearly been seen in more than a dozen clusters at between the 5 and 10 sigma level [202, 203], with redshifts ranging from 0.023 for COMA to 0.55. The immediate implication of these sort of observations is that the CMB comes from a redshift > 0.55 .

The SZ effect involves pressure integrals along the line of sight through the cluster. Even in a medium with gas in states of mixed cooling, so with large density and temperature inhomogeneities, the pressure tends to uniformize on a sound crossing time into a distribution defined by the gravitational potential. By contrast, X-ray emission – involving line of sight integrals of $n_e^2 T_e^{1/2}$ for bremsstrahlung and more complicated temperature and abundance dependences for recombination and line cooling – is very sensitive to clumping. Because the SZ effect is proportional to n_e rather than n_e^2 , it can probe the intracluster medium out to much larger radial distance than the X-ray emission, especially when sensitivities in the few times 10^{-6} range can be achieved. A further advantage is that $\Delta T/T$ for clusters does not diminish with redshift for a nonevolving cluster population, whereas cluster X-ray fluxes drop off dramatically. We have evidence from X-ray observations that there is strong evolution of the cluster population, and this is expected theoretically as well. Even so, we should eventually be able to probe clusters at $z \sim 1$ with the SZ effect. An illustration of this is fig. 17: a smooth particle hydrodynamics calculation in a CDM model of a rare-event region that grows into a cluster of COMA-like mass $10^{15} h^{-1} M_\odot$ by $z = 0$ after a major merger at $z \sim 0.05$ is seen in $\Delta T/T$ at $z = 0.7, 0.5, 0.2$. (SPH and other hydro calculations of the SZ effect for individual clusters were pioneered by Evrard [213] but have not been exploited much to date. The SPH example shown, from Bond and Wadsley [214], evolved from peak-patch initial conditions for a spherical region $30 h^{-1} \text{ Mpc}$ across constrained to give a cluster of the final state mass, as described in [68]. The cosmology is a standard CDM model of fig. 7 with $\sigma_8 = 1$ (20% lower σ_8 than COBE-normalization would give). The simulation used 65247 gas and 65247 dark particles, and a 128^3 multigrid Gauss-Seidel

gravity solver with particle-particle corrector forces to improve short distance resolution, by a factor of 20 or so. The gravitational and SPH smoothing were matched, which means that about 40 neighbors were required to be within the softening radius. As such the resolution achieved, $30 \text{ h}^{-1} \text{ kpc}$ at the final stage, is a factor of 5 or so better than X-ray core radii, just good enough for the X-ray calculations. Calculations that are optimized for X-rays with an order of magnitude more particles are easily feasible on current workstations and are currently being done by a number of groups.)

Combining the SZ and X-ray observations is one of the main paths to H_0 (and in principle q_0). Because $(\Delta T) \propto \int n_e T_e \bar{a} d\chi$ and the X-ray surface brightness $\Sigma_X \propto \bar{a}_{cl}^4 \int n_e^2 T_e^{1/2} \bar{a} d\chi$ for bremsstrahlung, the proper linear scale of the cluster, $R_{cl} \propto (\bar{a}_{cl}^2 \Delta T)^2 / (T_e^{3/2} \Sigma_X)$, can be measured. But $R_{cl} = 2H_0^{-1}(1 - \sqrt{\bar{a}_{cl}})\bar{a}_{cl}\theta_{cl}$ (for flat universes, with a weak q_0 -dependence for nearby clusters), so combining the SZ and X-ray observations with the angular size θ_{cl} and the redshift of the cluster z_{cl} allows H_0 to be estimated. In practice the data are used in a more sophisticated way than this, but even so, spherical symmetry is assumed: cluster elongation along the line of sight pushes H_0 down, clumpiness pushes it up. These and other effects make the values of H_0 derived from SZ/X observations uncertain and it is difficult to set realistic error bars.

So far the SZ effect has only been observed in massive clusters where the effect is quite large. Birkinshaw [204, 205, 206], using the single 40-m OVRO dish at 1.5 cm (*ovro* filter, fig. 6), observed the SZ effect in 3 clusters, 0016+16, Abell 665 (the richest cluster in Abell's catalogue), and Abell 2218, with observed central decrements $-(1.8, 1.5, 1.5) \times 10^{-4}$. These were used to estimate Hubble parameters for the latter two of 51 ± 18 and 65 ± 25 . Nearby clusters, such as Coma at $z = 0.023$ [207] and three other X-ray luminous ones [208], have recently been detected using the 5-m OVRO dish (*ovro22* filter, fig. 6). H_0 values obtained with this data vary, with 74 ± 29 for COMA, similarly large values for Abell 2256 and 2142, but a significantly smaller one for Abell 478. The Ryle telescope is a 5-km interferometer array with 8 13-m dishes and receivers operating at 2 and 6 cm. [209] showed the first SZ image of a cluster, Abell 2218, at $z = 0.17$ (with $H_0 = 38 \pm 17$ cf. [206]) and the Ryle team have now imaged the SZ effect in a dozen clusters, including 0016+16, Abell 665, 1722 and 773. Since some clusters give low H_0 , some high, it is unclear what conclusion to draw: although we can be confident that the statistical error bars will shrink with new technological advances, systematic error bars may never be reliable because clusters are decidedly not idealized spherical distributions. However, the distribution of H_0 determinations for a well-selected sample of clusters may help to reduce these biases.

Other interferometers are also being applied to this problem. Because clusters at moderate redshift subtend a reasonably large angle, high resolution instruments with long baselines such as the VLA are not effective. On the other hand, smaller dishes in a close-packed configuration are quite promis-

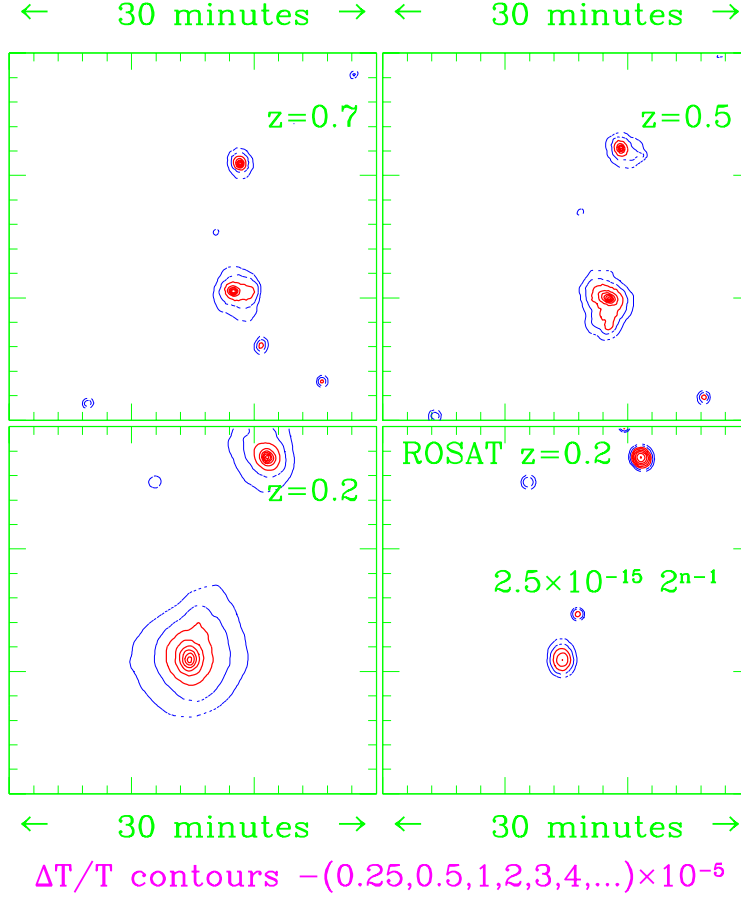


Figure 17: SZ contour maps derived from a cosmological hydrodynamics simulation for a cluster which becomes massive, hot and large at redshift zero after a major merger at $z \sim 0.05$. The SZ image ($\Delta T/T = -2y$ here) is contrasted with the ROSAT image at redshift 0.2. The ROSAT contours correspond to deep-pointing mode. The upper panels show that the SZ effect in the pre-merge pieces at $z = 0.7, 0.5$ is reasonably large on a few arcminute scales. The solid SZ contour levels $\geq 10^{-5}$ are experimentally accessible now and the dotted $< 10^{-5}$ contours should eventually be feasible. The contours roughly scale with Ω_B , here chosen to be 0.05 for this $\Omega_{nr} = 1$ CDM model.

ing: the OVRO millimeter-wave interferometer at 32 GHz was used to observe Abell 773 and 0016+16 [212]; the Australian compact telescope array, ACTA, is another example of a compact array being applied to this problem [211].

SuZIE uses bolometers (operating at 300 mK, at wavelengths 1.2 and 2.2 mm) on the Caltech sub-mm telescope (1.4' beam, 2' separation). The SZ effect in Abell 2613 has been observed, with a large $-(2.6 \pm 0.6) \times 10^{-4}$ decrement. The advantage here is that one can straddle the SZ sign change (fig. 5). It is hoped that one can use this to extract the moving cluster effect from the SZ effect, at least for cases when the cluster is moving mostly forward or away from us (cf. fig. 16) at a high speed.

5.3.5 The maximum entropy nature of Gaussian anisotropies

One of the fundamental features of these secondary maps is that they have their power concentrated in hot and/or cold spots: they are decidedly non-Gaussian. The fundamental characterization of Gaussian fluctuations is described by the following lemma [117, 120]: *Consider a general statistical distribution functional $\mathcal{P}[\Delta_t(\hat{q})]\mathcal{D}\Delta_t(\hat{q})$ giving the probability of an anisotropy configuration $\Delta_t(\hat{q})$ of the random field Δ_t . Define the entropy of this probability to be*

$$\text{Entropy}[\mathcal{P}] = - \int \mathcal{P}[\Delta_t(\hat{q})] \ln (\mathcal{P}[\Delta_t(\hat{q})]) \mathcal{D}\Delta_t(\hat{q}). \quad (145)$$

Among all of the distributions with a specified spectrum $\mathcal{C}_\ell$, the Gaussian one is the one which maximizes this entropy. Thus the Gaussian statistics of the primary anisotropy maps displayed in fig. 9 show maximally random distributions of the power available. The best observing strategy is then to concentrate the observing time on just a dozen or so patches of the sky because you are bound to hit something. For non-Gaussian fluctuations, with power more concentrated around the “hot” or “cold” spots than in the Gaussian case, a better observing strategy is to sample many patches at lower sensitivity to look for the regions of high power concentration. Because we now expect that the observed anisotropy will be a sum of many component signals, Galactic as well as cosmological, most of which will be source-like non-Gaussian ones, it is really essential to sample very many patches: i.e., to make large maps.

5.3.6 Quadratic nonlinearities in Thomson scattering

As noted in section 5.1, quadratic nonlinearities in Thomson scattering can sometimes dominate over the first-order anisotropies if the latter are strongly damped and there is early ionization. Their importance was originally suggested by Vishniac [109], and calculations have been done by Efstathiou and I [216, 218, 243], Dodelson and Jubas [142] and Hu and Scott [221]. Even if there is early reionization in nearly scale invariant models, there is generally not sufficient power on small length scales for the Vishniac effect to be important. Thus it can usually safely be ignored in inflation-based models.

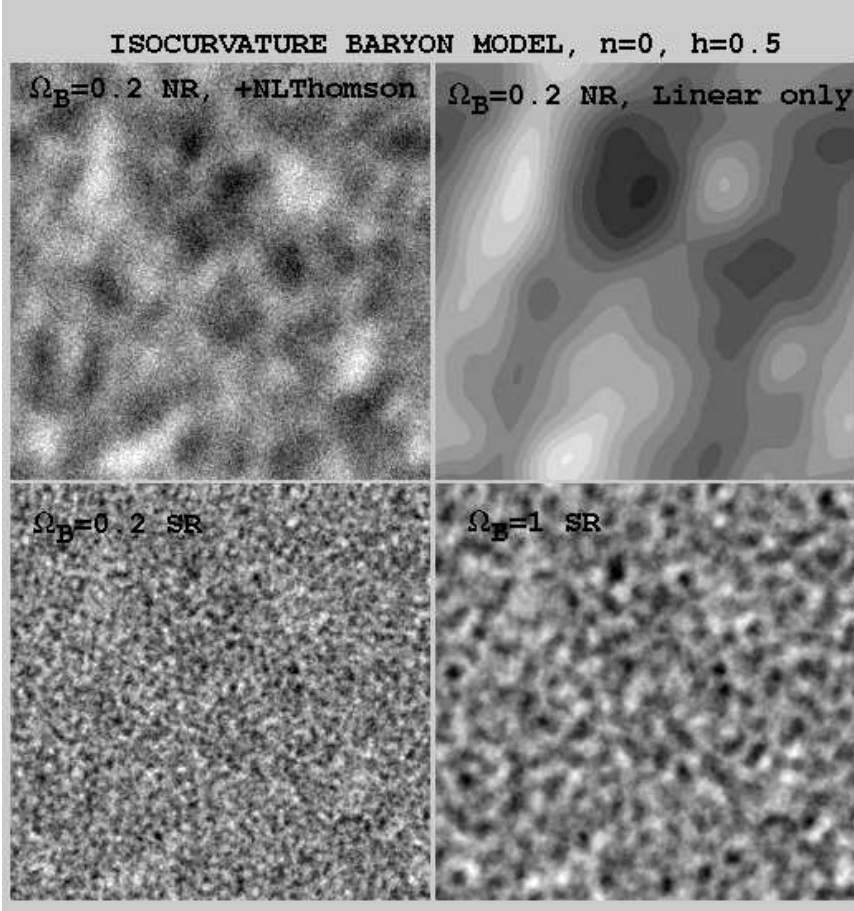


Figure 18: $10^\circ \times 10^\circ$ maps, with grey scale extending from $-4\sigma_{map}$ to $4\sigma_{map}$, for an $n_{is} = 0$, $\Omega = \Omega_B = 0.2$ isocurvature baryon model with (b) no recombination, calculated using only linear perturbation theory, (a) with quadratic nonlinearities added as well (and assuming the superposition of many sources along each line of sight is sufficient to make the nonlinear contributions Gaussian-distributed, which should be reasonable). (c) and (d) show standard recombination models, (d) with $\Omega_B = 1$ and (c) with $\Omega = \Omega_B = 0.2$, illustrating how the changed geometry concentrates the signal to smaller angular scales. The total power $\mathcal{I}[\mathcal{C}_\ell]$ for (b) is $(3 \times 10^{-5})^2 \sigma_8^2$, rising to $\approx 10^{-8} \sigma_8^2$ for (a) with the nonlinearities. The SR model (c) has total power $10^{-8} \sigma_8^2$.

This is not so for isocurvature baryon models in which the initial spectral index n_{is} is considered a free parameter, as the maps in fig. 18 adapted from [222] demonstrate, using power spectra taken from [243] calculated using the methods of [88, 216]. The nonlinear source-field is $\mathcal{G} = \mathcal{V}_C(\tau)\delta_e\hat{q} \cdot \mathbf{v}_e$, where $\delta_e = \delta n_e/\bar{n}_e$ is the perturbation to the electron density and $\mathbf{v}_e$ is the electron velocity. Subdominant nonlinearities $\propto (\delta n_e)^2$ have been ignored. The electron and baryon velocities can be taken to be the same, but $\delta_e = \delta_B + \delta Y_e/\bar{Y}_e$ can have a piece associated with fluctuations in the ionization fraction, although it is usually ignored. The calculation uses some of the techniques of section 5.1, in particular the Fourier transform expression for $\mathcal{C}_\ell$, eq. (129), using a time-averaged isotropized power spectrum for the nonlinear source-field given by [216]

$$\begin{aligned}\mathcal{C}_\ell &\sim Q^2 \int_{Q/\bar{\mathcal{R}}}^\infty \frac{d \ln k}{k\bar{\mathcal{R}} \sqrt{(k\bar{\mathcal{R}})^2 - Q^2}} I_1 \bar{\mathcal{P}}_{\bar{\mathcal{G}}\bar{\mathcal{G}}}, \\ I_1 &= \int d\bar{\chi} \mathcal{V}_C^2(\bar{\chi}) \frac{(D\dot{D})^2}{(D_1\dot{D}_1)^2} \tau_{\zeta_C=1}, \\ \bar{\mathcal{P}}_{\bar{\mathcal{G}}\bar{\mathcal{G}}} &\approx \frac{\pi}{32} \left(\frac{1}{9} \mathcal{P}_{\Phi_N}(k, \tau_{\zeta_C=1}) \right)^2 (k\tau_{\zeta_C=1})^5 I_2(k),\end{aligned}\quad (146)$$

$$I_2(k) = \int_0^\infty dy \int d\mu \frac{y(1-\mu^2)(1-2\mu y)^2}{(1+y^2-2\mu y)^{3/2}} \frac{\mathcal{P}_{\Phi_N}(k(1+y^2-2\mu y)^{1/2})}{\mathcal{P}_{\Phi_N}(k)} \frac{\mathcal{P}_{\Phi_N}(ky)}{\mathcal{P}_{\Phi_N}(k)}.$$

The gravitational potential fluctuations are assumed to be time-independent appropriate to nr -dominated evolution; $D(\tau) \sim \tau^2$ is the linear growth factor for the density fluctuations and $\dot{D} = dD/d\tau$ is the linear growth factor for the velocity fluctuations. The normalization time, $\tau_{\zeta_C=1}$, is chosen to be when the optical depth is unity, i.e., at the maximum of $\mathcal{V}_C/(\bar{H}\bar{a})$. The Vishniac effect actually comes from a broad redshift range because of the $D\dot{D}$ growth factor. $\bar{\mathcal{R}}$ is an average cosmological distance.

The spectral index in the maps has been chosen to correspond to the Poisson seed model (although Gaussian initial conditions were assumed). Phenomenologically it seems n_{is} between -1 and 0 is preferred. Since there is a great deal of power at short distances in these models, star formation is expected to occur very early, hence it seems likely that the Universe would have been photoionized shortly after the usual epoch of recombination at $z \sim 1000$. Thus no-recombination (NR) models provide a more realistic description of the anisotropies expected in isocurvature baryon models, although our limited ability to deal with star formation in the early universe gives us arbitrary freedom in designing an ionization history. The isocurvature effect due to photon bunching at large scales is augmented by Thomson scattering anisotropies from the flow of baryons during decoupling, giving an enhanced signal around $\ell \sim 200$ for the NR $\Omega_B = 0.2$ model. Although primary small angle anisotropies are

diminished if there is NR, the quadratic nonlinearities in the scattering induce a significant anisotropy in, e.g., the *ovro* window of fig. 6, and even more so for experiments with filters like “VLA”, especially for high n_{is} models. With σ_8 about unity (the conventional value for these models), large regions in Ω_B - n_{is} space are ruled out by the observations of current small and medium angle observations [243], but exactly how much depends upon one’s assumptions about ionization history. Nonlinearities beyond quadratic order could also obscure this result. There is also uncertainty in how to extrapolate the spectrum to the curvature scale, so it is unclear how such a model is to be “COBE-normalized”.

5.3.7 The influence of weak gravitational lensing on the CMB

Another nonlinear effect is gravitational lensing which bends, focusses and defocusses the CMB photons as they propagate from decoupling through the clumpy medium to us. Of course lensing is a mature subject in astronomy (e.g., [272, 273]). There was a flurry of activity in the late 80s assessing whether or not lensing would significantly decrease anisotropies by taking photons from a high $\Delta T/T$ region and dispersing them into lower $\Delta T/T$ regions [274, 275, 276, 277, 278]. Given the difficulties that astronomers have had detecting lensing, with the best observations coming from clusters of galaxies, it may seem obvious that the effect on the $\sim 10'$ coherence scale typical for primary CMB anisotropies is likely be quite small; and this is what these papers found. However there is an effect on sub-arcminute scales that may affect some types of secondary anisotropies.

An important issue to re-address is how weak-lensing from late-time linear and nonlinear structure development may complicate the interpretation of the primary anisotropy power spectrum even if it is very well determined. In the post-COBE era, the CMB lensing issue has been picked up again by [279, 280]. In particular, while the earlier papers emphasized the influence of lensing on the correlation function, Seljak [280] has shown what its impact will be on C_ℓ . To show the effects on both, in Appendix C.5 I apply the Boltzmann transport equation formalism of Appendix B to lensing of primary anisotropies, in particular to $C^{\text{lens}}(\theta)$ and C_ℓ^{lens} in eqs. (368), (370).

The critical quantity to determine is the statistical distribution of the extra displacement between two photons due to lensing relative to their unlensed separation; i.e., the component of geodesic deviation driven by curvature. At small separations, the displacement defines a 2×2 shear tensor. The surface upon which the radiation pattern is constructed and upon which the separation is measured should be well after decoupling so that the distribution function for the here and now is just a direct map of the initial distribution function on the post-decoupling hypersurface by the action of a Green function that now fully incorporates the bending geodesic trajectories.

The total angular power is conserved: i.e., $C^{\text{lens}}(0) = C^{\text{no-lens}}(0)$. However, at finite separation θ , $C^{\text{lens}}(\theta)$ is a smoothed version of $C^{\text{no-lens}}(\theta)$, with smoothing scale $\sim \varepsilon\theta$ where ε is basically an *rms* shear. Since $C(\theta)/C(0)$ gives

the statistically-averaged profile about a point, this means that lensing smooths out hot and cold spots e.g., [275, 276, 277], but it does so in such a way as to preserve the overall power in a map. How much spreading occurs depends upon how outrageous the structure formation model is, but the consensus of the papers on this subject is that the effect is not very large for primary anisotropies. Seljak [280] has used realistic power spectra that include nonlinear corrections, which enhances the role of lensing at sub-arcminute scales, to translate the correlation function decline into its effect on $\mathcal{C}_\ell$: the effect on the primary spectra of fig. 7 is to smooth the Doppler peaks; the typical range in ℓ over which the power is spread in $\Delta\ell/\ell$ is basically the weak-lensing shear, about 10% to 20% or so at a few arcminutes, depending upon the model [280] – in agreement with the levels estimated by people advocating using the influence of weak-lensing on the ellipticities of faint galaxy images to determine the mass density power spectrum e.g., [273].

6 Perturbation theory of primary anisotropies

6.1 Overview of fluctuation formalism

A generic fluctuation variable $D(\mathbf{x}, t)$ can be expanded in terms of modes $\mathcal{M} \in \{\text{adiabatic scalar, isocurvature scalar, vector or tensor; growing or decaying}\}$:

$$D(\mathbf{x}, t) = w \sum_{\mathbf{k}\mathcal{M}} \left\{ u_{\mathbf{k}\mathcal{M}}^{(D)}(t) Q_{\mathbf{k}\mathcal{M}}(\mathbf{x}) a_{\mathbf{k}\mathcal{M}} + u_{\mathbf{k}\mathcal{M}}^{(D)*}(t) Q_{\mathbf{k}\mathcal{M}}^*(\mathbf{x}) a_{\mathbf{k}\mathcal{M}}^\dagger \right\}$$

$$w = 1/2 \text{ classical, } w = 1 \text{ quantum.} \quad (147)$$

For classical fluctuations, $a_{\mathbf{k}\mathcal{M}}$ is a random variable and $a_{\mathbf{k}\mathcal{M}}^\dagger$ its complex conjugate, while for quantum fluctuations, $a_{\mathbf{k}\mathcal{M}}$ is an annihilation operator for the mode $\mathbf{k}\mathcal{M}$ and $a_{\mathbf{k}\mathcal{M}}^\dagger$ is the creation operator. The $u_{\mathbf{k}\mathcal{M}}^{(D)}(t)$ are mode functions which describe the evolution (and, for now, include polarization effects, e.g., for gravitational waves). The spatial dependence of the modes is given by eigenfunctions $Q_{\mathbf{k}\mathcal{M}}(\mathbf{x})$ of the Laplacian of the background geometry. For a flat background of most relevance to inflation models, it is simply a plane wave, $Q_{\mathbf{k}\mathcal{M}}(\mathbf{x}) = e^{i\mathbf{k}\cdot\mathbf{x}}$, labelled by a comoving wavevector $\mathbf{k}$. For curved backgrounds, the eigenfunctions are more complex.

The power spectrum of D associated with mode $\mathcal{M}$ is the fluctuation variance per log wavenumber and can be expressed in terms of the statistics of $a_{\mathbf{k}\mathcal{M}}$ and $a_{\mathbf{k}\mathcal{M}}^\dagger$:

$$\begin{aligned} \text{quantum:} \quad \mathcal{P}_D^{\mathcal{M}}(k) &\equiv \frac{d\sigma_{D|\mathcal{M}}^2}{d\ln k} = \frac{k^3}{2\pi^2} |u_{\mathbf{k}\mathcal{M}}^{(D)}(t)|^2 (1 + 2\langle a_{\mathbf{k}\mathcal{M}}^\dagger a_{\mathbf{k}\mathcal{M}} \rangle), \\ \text{classical:} \quad \mathcal{P}_D^{\mathcal{M}}(k) &= \frac{k^3}{2\pi^2} |u_{\mathbf{k}\mathcal{M}}^{(D)}(t)|^2 \langle a_{\mathbf{k}\mathcal{M}}^* a_{\mathbf{k}\mathcal{M}} \rangle. \end{aligned} \quad (148)$$

In the quantum case, $\langle(-)\rangle$ denotes $\text{Trace}(\rho(-))$, where ρ is the density matrix operator; in the classical case, it denotes ensemble average with respect to

the probability distribution functional. If the modes are Gaussian-distributed, statistically homogeneous and isotropic, then this is all that is needed to specify the patterns in the field $D(\mathbf{x}, t)$. The local shape is characterized by the index

$$n_D(k) + 3 \equiv d \ln \mathcal{P}_D^{\mathcal{M}}(k) / d \ln k. \quad (149)$$

Thus $-n_D$ is a “fractal dimension”: zero is white noise, while three is scale invariance in D .

In the inflation picture, the wavenumbers in the observable regime are usually considered to be so high that any pre-inflation mode occupation, $\langle a_{\mathbf{k}\mathcal{M}}^\dagger a_{\mathbf{k}\mathcal{M}} \rangle$, is negligible, and only the unity zero point oscillation term appears. In that case, we connect to the random field description by making the real and imaginary parts of $a_{\mathbf{k}\mathcal{M}}$ Gaussian-distributed with variance $1/2$. Although quantization is at least self consistent in linear perturbation theory about a classical background, and gauge invariant, there are still obvious subtleties associated with the transition from a quantum to a classical random field description. A true inconsistency appears if we include the nonlinear backreaction of the fluctuations upon the background fields and upon themselves. For this, we would need a quantum gravity theory. The stochastic inflation theory is an attempt to bypass this using classical fields acted upon by quantum-derived noise (e.g., [180, 181]). In the inflation regime,

$$D \in \{\delta\phi_{inf}, \delta\phi_{is}, h_+, h_-, \delta \ln a, \delta \ln H, \delta q, \dots\}.$$

That is, D would refer to fluctuations in: (1) the inflaton field $\delta\phi_{inf}$ whose equation of state can give the negative pressure needed to drive the acceleration; (2) other scalar field degrees of freedom $\delta\phi_{is}$ which can, for example, induce scalar isocurvature perturbations. (If axions are the dark matter, ϕ_{is} would be the axion field.) The isocurvature baryon mode would need to have a ϕ_{is} (“isocons”) coupled some way to the baryon number, e.g., [246]; (3) gravitational wave modes h_+, h_- ; (4) the inhomogeneous scale factor $a(\mathbf{x}, t)$, the Hubble parameter $H(\mathbf{x}, t)$ and the deceleration parameter:

$$q(\mathbf{x}, t) \equiv -d \ln H a / d \ln a, \quad (150)$$

or other geometrical variables encoding scalar metric perturbations and their variations.¹ Inflation ends when q passes from negative to positive. Provided the fluctuations over the observable k -range remain Gaussian, the outcome of inflation is therefore a set of amplitudes for scalar metric (adiabatic) perturbations, gravity wave modes and various possible isocurvature modes, and primordial spectral tilts for each, in particular:

$$\text{scalar:} \quad \nu_s(k) \equiv n_s(k) - 1 \equiv \frac{d \ln \mathcal{P}_{\ln a|_{H*}}(k)}{d \ln k}$$

¹To be more precise, in terms of the variables of eq. (24), in the longitudinal gauge with $\Psi_\sigma = 0$, we have $\delta \ln a \equiv \varphi_L$, $(\delta \ln H) = (\bar{H}\bar{a})^{-1}\dot{\varphi}_L - \nu_L$, $\delta q = (\bar{H}\bar{a})^{-1}\dot{\nu}_L + (1 + \bar{q})\delta \ln H$, and the fluctuation used to characterize post-inflation amplitudes is $\delta \ln a|_{H*} = \delta \ln a - (d \ln \bar{a} / d \ln \bar{H})\delta \ln H = \varphi_{com}$.

$$\begin{aligned} & (a|_{H*} \equiv a(\mathbf{x}, t(\mathbf{x}, H^{-1}))), \\ \text{tensor:} \quad & \nu_t(k) \equiv n_t(k) + 3 \equiv \frac{d \ln \mathcal{P}_{GW}(k)}{d \ln k}, \end{aligned} \quad (151)$$

Measuring the power in scalar metric fluctuations on the time surfaces upon which the inhomogeneous Hubble parameter $H(\mathbf{x}, t)$ – the proper time derivative of $\ln a(\mathbf{x}, t)$ – is constant is useful [175, 192, 176, 181]: Once Ha exceeds k for a mode with wavenumber k , $(\delta \ln a|_{H*})$ becomes time-independent during an inflation epoch with a single dynamically-important scalar field, and it remains so through reheating and the passage from radiation into matter dominance until Ha falls below k (the wave “re-enters” the horizon). Although transforming calculations to a uniform Hubble hypersurface is instructive, it does not mean that solving the equations for fluctuations defined on that hypersurface is best. The perturbation quantities used in practice depend upon the gauge and choice of time surfaces, and are described in the next section.

In the post-inflation period,

$$D \in \{\delta \rho_{cdm}, \delta v_{cdm}, \delta \rho_B, \delta v_B, \delta f_\gamma, \delta f_{er\nu}, \delta f_{m\nu}, h_+, h_\times, \nu, \varphi, \Psi_\sigma, \dots\}.$$

That is, D would refer to fluctuations in the density and velocity of dark matter and baryons $(\delta \rho_{cdm}, \delta v_{cdm}, \delta \rho_B, \delta v_B)$, in the distribution functions for photons (δf_γ) and relativistic or semi-relativistic neutrinos $(\delta f_{er\nu}, \delta f_{m\nu})$, and in the metric (dispersing gravitational wave modes $h_{+,\times}$ and the scalar variables such as the “gravitational potential”, $\Phi_N = \nu_L$). The Gaussian nature of the statistics is not modified until mode-mode coupling occurs in the nonlinear regime.

6.2 Perturbed Einstein equations

6.2.1 Time-hypersurface and gauge freedom

In two relatively technical appendices, A and B, the Einstein–Boltzmann equations are viewed as defining a Cauchy problem: the spacetime metric plus matter variables step forward from a set of initial conditions through a sequence of spatial hypersurfaces, each labelled by a time coordinate. This “foliation” of spacetime into a $3 + 1$ split is described by the ADM formalism [167, 168, 169, 171, 178, 196]. Appendices A and B give the full nonlinear equations for transport and metric evolution, and only then are reduced to linear perturbation theory, because the nonlinear version illuminates the physical meaning of the perturbation terms. Because the ADM formalism restricts attention to foliations which are covered by a single time parameter, a change of foliation (timelike hypersurfaces) is conceptually intermingled with a change of coordinate system (gauge transformations). The gauge invariance aspect of this which looms so large in much of the cosmological literature is not as important as the choice of time surfaces upon which the perturbations are instantaneously measured. The time surfaces have a spatial 3-geometry, defined by a metric

$^{(3)}g_{ij}$, which are the geometrodynamical variables encoding the dynamics of the gravitational field. The theorist can decide how to push/pull his/her spatial hypersurfaces forward. This is encoded in the 4 remaining components of the spacetime metric, parameterized in terms of a lapse function N and a shift three-vector N^i , $i = 1, 2, 3$ ($g_{00} = -(N^2 - N_k N^k)$, $g_{0i} = N_i$). The contravariant 4-vector, e_n , with components $(e_n^\alpha) = N^{-1}(1, -N^1, -N^2, -N^3)$, is timelike and unity-normalized, $e_n^\alpha g_{\alpha\beta} e_n^\beta = -1$: it is the 4-velocity of observers who each have fixed positions on the spatial hypersurfaces (fiducial observers).

The covariant derivative of any 4-velocity U (of which e_n is a special case) can be decomposed into an acceleration 4-vector $\mathcal{A} \equiv \nabla_U U$ ($\mathcal{A}^a \equiv U^b U^a{}_{;b}$, where ‘;’ denotes covariant derivative), expansion rate θ , a vorticity ω_{ab} and an anisotropic shear σ_{ab} (where $a, b \in \{0, 1, 2, 3\}$):

$$U_{b;a} = -U_a \mathcal{A}_b + \frac{1}{3}\theta \perp_{ab} + \sigma_{ab} + \omega_{ab}, \quad \perp_{ab} \equiv g_{ab} + U_a U_b. \quad (152)$$

The tensor $\perp_b^a$ satisfies $\perp(U) = 0$, $\perp^2 = \perp$ (i.e., $\perp_c^a U^c = 0$ and $\perp_c^a \perp_b^c = \perp_b^a$), hence is a projection operator onto the three-dimensional subspace orthogonal to the flow U^a . The vorticity tensor is the antisymmetric (and trace-free) part of $\perp_a^c U_{b;c}$, θ is its trace, and σ_{ab} is the remaining symmetric trace-free part.

By construction, spacelike hypersurfaces exist which are orthogonal to the $U^\alpha = e_n^\alpha$ flow. This implies vanishing vorticity for the flow of time, $\omega_{ab} = 0$. (In general spacetimes, a global time parameter may not in fact exist.) The remaining spacelike part is the total shear, and its negative, K_{ab} , is called the extrinsic curvature:

$$\text{for } U = e_n, \quad K_{ab} \equiv -(\frac{1}{3}\theta \perp_{ab} + \sigma_{ab}). \quad (153)$$

It measures the relative deviation of the fiducial flow lines and defines how the spatial 3-geometry changes in time.

For a given flow U , in particular for the time flow e_n^α , the stress energy tensor T_{type}^{ab} for matter of type “type” can be decomposed into an energy density, ρ_{type} , a momentum current $J_{(e),\text{type}}^a$, an isotropic pressure, p_{type} and an anisotropic pressure tensor Π_{type}^{ab} :

$$\begin{aligned} T_{\text{type}}^{ab} &= \rho_{\text{type}} U^a U^b + J_{(e),\text{type}}^a U^b + U^a J_{(e),\text{type}}^b + p_{\text{type}} \perp^{ab} + \Pi_{\text{type}}^{ab}, \\ \perp^{ab} &\equiv g^{ab} + U^a U^b, \quad \rho_{\text{type}} \equiv U_a T_{\text{type}}^{ab} U_b, \quad J_{(e),\text{type}}^a \equiv \perp_c^a T_{\text{type}}^{cb} U_b, \\ p_{\text{type}} \perp^{ab} + \Pi_{\text{type}}^{ab} &\equiv \perp_c^a T_{\text{type}}^{cd} \perp_d^b, \quad (\Pi_{\text{type}})_a^a = 0. \end{aligned} \quad (154)$$

The total values $\rho_{\text{tot}}, J_{(e),\text{tot}}^a, p_{\text{tot}}, \Pi_{\text{tot}}^{ab}$ are just the sums of course.

In perturbation theory, we expand the spatial three metric, the lapse and the shift in terms of normal modes for the Einstein equations, expressed in terms of scalar metric variables $\varphi, \psi, \nu, \Psi_N$, the vector contributions, $h_{ij}^{(V)}, N_i^{(V)}$, and the (transverse traceless) tensor contributions, $h_{ij}^{(TT)}$:

$$^{(3)}g_{ij} = \left(^{(3)}\bar{g}_{ij} + ^{(3)}\bar{g}_{ij} 2\varphi - \bar{a}^2 ^{(3)}\bar{\nabla}_i ^{(3)}\bar{\nabla}_j 2\psi + \bar{a}^2 h_{ij}^{(V)} + \bar{a}^2 h_{ij}^{(TT)} \right),$$

$$N = \bar{N}(1 + \nu), \quad N_i = \bar{N} {}^{(3)}\nabla_i \Psi_n + N_i^{(V)}, \quad \Psi_\sigma \equiv \frac{\bar{a}^2}{\bar{N}} \dot{\psi} + \Psi_n. \quad (155)$$

The parameters of the background geometry are the average lapse $\bar{N}$, which is taken to be $\bar{a}$ if conformal time τ is chosen, and an unperturbed FRW background 3-metric ($k = 0, \pm 1$ gives the 3 FRW curvature possibilities and isotropic coordinates are used here):

$$\begin{aligned} {}^{(3)}\bar{g}_{ij} &= \bar{a}^2 f^2(r) \delta_{ij}, \quad f^{-1} = 1 + \frac{kr^2}{4d_{curv}^2}, \quad \chi = \tau_0 - \tau, \\ k = -1: \quad &\left\{ \frac{r}{2d_{curv}} = \tanh \frac{\chi}{2d_{curv}}, \quad \Sigma = fr = d_{curv} \sinh \frac{\chi}{d_{curv}} \right\}, \\ k = 1: \quad &\left\{ \frac{r}{2d_{curv}} = \tan \frac{\chi}{2d_{curv}}, \quad \Sigma = fr = d_{curv} \sin \frac{\chi}{d_{curv}} \right\}, \\ {}^{(3)}\bar{R} &= \frac{6k}{(d_{curv}\bar{a}(t))^2}, \quad \bar{K}_{ij} = -\frac{\dot{\bar{a}}}{N\bar{a}} f^2 \bar{a}^2 \delta_{ij}. \end{aligned} \quad (156)$$

Here τ_0 is the current conformal time, and χ is the comoving distance back to redshift z (and $\chi = c(\tau_0 - \tau)$ is the solution for radial photon geodesics). The covariant derivative with respect to the background 3-metric in the direction i is ${}^{(3)}\bar{\nabla}_i$, ${}^{(3)}\bar{\nabla}^i \equiv {}^{(3)}\bar{g}^{ij} {}^{(3)}\bar{\nabla}_j$, and the Laplacian is ${}^{(3)}\bar{\nabla}^2 = {}^{(3)}\bar{\nabla}^i {}^{(3)}\bar{\nabla}_i$. For flat ($k = 0$) models, ${}^{(3)}\bar{\nabla}_i = \partial_i$, ${}^{(3)}\bar{\nabla}^i \equiv \delta^{ij} \bar{a}^{-2} \partial_j$, and the Laplacian is ${}^{(3)}\bar{\nabla}^2 = \bar{a}^{-2} \sum \partial_j \partial_j$ – recall these $\bar{a}^{-1}$ factors in the following.

6.2.2 Scalar mode Einstein equations

The physical meaning of the scalar metric variables is determined by their relation to such physical quantities as the fluctuations in the 3-curvature scalar, $(\delta {}^{(3)}R)$, in the anisotropic 3-curvature tensor, $(\delta {}^{(3)}R')^i_j$, in the expansion rate, (δH) , in the anisotropic shear σ_j^i of e_n . For a conformal time choice, with $\bar{N} = \bar{a}$, from eqs. (248), (247), (244), (245), (248) we have:

$$(\delta {}^{(3)}R) = -4 {}^{(3)}\bar{\nabla}^2 \varphi - {}^{(3)}\bar{R} 2\varphi, \quad (157)$$

$$(\delta {}^{(3)}R')^i_j = -[{}^{(3)}\bar{\nabla}^i {}^{(3)}\bar{\nabla}_j - \frac{1}{3} \delta^i_j {}^{(3)}\bar{\nabla}^2] \varphi, \quad (158)$$

$$\bar{a}(\delta H) = \dot{\varphi} - \bar{N} \bar{H} \nu - \frac{1}{3} {}^{(3)}\bar{\nabla}^2 \bar{a} \Psi_\sigma, \quad (159)$$

$$\sigma_j^i = -({}^{(3)}\bar{\nabla}^i {}^{(3)}\bar{\nabla}_j - \frac{1}{3} \delta^i_j {}^{(3)}\bar{\nabla}^2) \Psi_\sigma. \quad (160)$$

In addition to the three metric scalars, for each type of matter present there will be a relative density perturbation, δ_{type} , a velocity perturbation which, because the flow is irrotational for scalar perturbations, can be written in terms of a velocity potential, $\Psi_{v,\text{type}}$: $\mathbf{v}_{\text{type}} = -\bar{a}^{-1} \nabla \Psi_{v,\text{type}}$,² Here type

²This is the usual definition for nonrelativistic matter, but is better defined in terms of the momentum current, $(\rho_{\text{type}} + p_{\text{type}}) \mathbf{v}_{\text{type}}$, especially for scalar fields ϕ for which $\Psi_{v,\phi} = \bar{a} \dot{\phi}^{-1} \delta \phi$ is ill-defined since $\dot{\phi}$ can vanish [178, 192].

runs over inflaton and isocon fields, massless and massive neutrinos, photons, baryons, CDM, etc. For some types of matter there may be an anisotropic pressure perturbation, $\pi_{t,\text{type}}$, and for photons and neutrinos, there will be higher moments, expressing all of the degrees of freedom in the perturbed distribution functions.

There are generally 10 Einstein equations, $G_b^a = 8\pi G_N T_b^a$. These split into 4 constraint equations, $G_n^n = 8\pi G_N T_n^n$ (where $G_{nn} \equiv e_n^\alpha G_{\alpha\beta} e_n^\beta$) and $G_n^I = 8\pi G_N T_n^I$, and 6 dynamical equations, $G_J^I = 8\pi G_N T_J^I$. (Here I, J refer to spatial components taken with respect to a triad e_I^α of spacelike 4-vectors perpendicular to e_n^α described in more detail in Appendix A.1.) Because spatial components of scalar variables can be expressed in terms of gradients, first integrals of the gradient equations can be done, reducing the total to 2 constraint and 2 dynamical equations.

The perturbed energy constraint (δG^{nn}) and momentum constraint (δG_n^I) equations are

$$2\bar{H}(\delta H) - \frac{2}{3} {}^{(3)}\bar{\nabla}^2 \varphi - \frac{1}{6} {}^{(3)}\bar{R} 2\varphi = \frac{8\pi G_N}{3} (\delta\rho)_{\text{tot}}, \quad (161)$$

$$\begin{aligned} \bar{a}(\delta H) + \frac{1}{3} {}^{(3)}\bar{\nabla}^2 \bar{a} \Psi_\sigma &= \dot{\varphi} - \bar{H} \bar{a} \nu \\ &= -4\pi G_N \sum_{\text{type}} (\bar{\rho} + \bar{p})_{\text{type}} \bar{a} \Psi_{v,\text{type}} - \frac{1}{6} {}^{(3)}\bar{R} \bar{a} \Psi_\sigma. \end{aligned} \quad (162)$$

A combination of the momentum and energy constraints gives a ‘‘Poisson–Newton’’ constraint equation:

$$\begin{aligned} - {}^{(3)}\bar{\nabla}^2 (\varphi + \bar{H} \Psi_\sigma) - \frac{1}{4} {}^{(3)}\bar{R} 2(\varphi + \bar{H} \Psi_\sigma) \\ = 4\pi G_N ((\delta\rho)_{\text{tot}} + 3\bar{H}(\bar{\rho} + \bar{p})_{\text{tot}} \Psi_{v,\text{tot}}). \end{aligned} \quad (163)$$

The dynamical Einstein equations are those for the isotropic pressure (δG_I^I) and the anisotropic stress ($\delta G_J^I - \frac{1}{3} \delta_J^I \delta G_K^K$); instead of (δG_I^I), it is more useful and usual to use the perturbed Raychaudhuri equation, $\delta R^n_n = -(\delta G_{nn} + \delta G_I^I)/2$:

$$\begin{aligned} \frac{\partial(\bar{a}\delta H)}{\partial\tau} + \frac{\dot{\bar{a}}}{\bar{a}}(\bar{a}\delta H) - (1 + \bar{q})(\bar{H}\bar{a})^2 \nu - \frac{1}{3} \bar{a}^2 {}^{(3)}\bar{\nabla}^2 \nu \\ = -\frac{4\pi G_N}{3} \bar{a}^2 ((\delta\rho) + 3(\delta p))_{\text{tot}}, \end{aligned} \quad (164)$$

$$\bar{a}^{-1} \dot{\Psi}_\sigma + \bar{H} \Psi_\sigma + (\varphi + \nu) = -8\pi G_N \sum_{\text{type}} \bar{p}_{\text{type}} \pi_{t,\text{type}}, \quad (165)$$

$$\Pi_{\text{type } ij}^{(S)} = ({}^{(3)}\bar{\nabla}_i {}^{(3)}\bar{\nabla}_j - \frac{1}{3} {}^{(3)}\bar{g}_{ij} {}^{(3)}\bar{\nabla}^2) \bar{p}_{\text{type}} \pi_{t,\text{type}}.$$

Note that only the combination Ψ_σ of the two scalars ψ, Ψ_n enters the linearized equations. Terms contributing to the total pressure are photons, $\bar{\rho}_\gamma \delta_\gamma/3$, massless neutrinos, $\bar{\rho}_{\text{erv}} \delta_{\text{erv}}/3$, hot and warm dark matter, and scalar field pressure if appropriate. The pressure of the baryons can be neglected – except on very small scales.

6.2.3 Useful gauge invariant combinations for scalar modes

Under change of the hypersurface from τ to $\tilde{\tau} = \tau + T(\mathbf{x}, \tau)$, the scalar metric variables change according to $\tilde{\varphi} = \varphi - \bar{H}\bar{a}T$, $\tilde{\Psi}_\sigma = \Psi_\sigma + \bar{a}T$, $\tilde{\nu} = \nu - \bar{a}^{-1}\partial_\tau(\bar{a}T)$ (see eq. (260) for the behavior of other variables). Three simple combinations can be made that are T -independent:

$$\begin{aligned}\nu_L &= \nu + \bar{a}^{-1}\dot{\Psi}_\sigma \equiv \Phi_A, & \varphi_L &= \varphi + \bar{H}\Psi_\sigma \equiv \Phi_H, \\ \varphi + \frac{(\bar{H}\bar{a})^{-1}\dot{\varphi} - \nu}{1 + \bar{q}} &= \varphi_{com} - \bar{H}^{-2} {}^{(3)}\bar{R}(\Psi_{v,\text{tot}} + \Psi_\sigma) \equiv (\delta \ln a|_{H*}), \\ \text{where } \varphi_{com} &\equiv \varphi - \bar{H}\Psi_{v,\text{tot}}.\end{aligned}\tag{166}$$

Here $\bar{q}$ is the mean deceleration parameter, expressible in terms of the mean densities and pressures of the matter present.

In the absence of mean curvature the metric combination $(\delta \ln a|_{H*})$ reduces to the scalar curvature potential on the hypersurfaces in which the net momentum current vanishes, φ_{com} . This quantity deserves some comment. In early universe calculations and to characterize the initial conditions for the photon transport through decoupling, the power in adiabatic scalar fluctuations on scales beyond the Hubble radius is best characterized in terms of quantities such as φ_{com} which become time-independent; φ_{com} has been used to simplify calculations of linear fluctuations generated by quantum noise since the early eighties by Mukhanov and others [176, 171, 185].

The variable $\ln a|_{H*}$ is the inhomogeneous scale factor as measured on time surfaces upon which the space creation rate, $H_* \equiv \partial \ln a / (N \partial \tau)$, is uniform. It gives a nice characterization of even nonlinear fluctuations that can arise in stochastic inflation [181]. However, H_* here is not exactly $H \equiv -K/3$, the usual Hubble parameter. For scalar perturbations, given a foliation, we can change spatial coordinates on the time surface to get $\psi = 0$, with all of Ψ_σ moved to the shift potential Ψ_n . In these coordinates,

$$\begin{aligned}H_* &= H + \frac{1}{3N} {}^{(3)}\nabla_j N^j \approx \bar{H} + \frac{1}{N}\dot{\varphi} - \bar{H}\nu, \\ (\delta H_*) &= (\delta H) + \frac{1}{3} {}^{(3)}\bar{\nabla}^2 \Psi_\sigma, & \ln a &= \ln \bar{a} + \varphi.\end{aligned}\tag{167}$$

Under the purely spatial transformation, φ remains unchanged. Both (δH_*) and φ are modified under time surface changes, but in such a way that the combination $(\delta \ln a|_{H*})$ is invariant.

For numerical or analytic calculations in inflation, it is impractical to work on a uniform Hubble foliation for complex calculations. One determines $(\delta \ln a|_{H*})$ by hypersurface shifting after the computations are done. In linear perturbation theory this is particularly simple. For example, suppose that the calculation has been performed in the longitudinal gauge, for which the variables of relevance are $\ln N$, $\ln a$ and $\ln H$. Keeping the same spatial coordinates, the time we transform to defines a scalar function $T(\mathbf{x}, \tau)$, and the inhomogeneous

scale factor and Hubble parameter become $a(\mathbf{x}, \tau + T(\mathbf{x}, \tau))$, $H_*(\mathbf{x}, \tau + T(\mathbf{x}, \tau))$. Choosing T to make H_* constant gives a nonlinear equation for $a|_{H_*}$. With linearization, $(\delta \ln a|_{H_*}) = \delta \ln a - \frac{d \ln a}{d \ln H} (\delta \ln H_*)$, i.e., eq. (166).

Bardeen [171] emphasized the virtues of φ_h , the value of φ on surfaces upon which $H \equiv -K/3$ is constant. The difference between (δH) and (δH_*) is a term of order $(k/(\bar{H}\bar{a}))^2$, hence φ_h differs from $(\delta \ln a|_{H_*})$ by terms of order $(k/(\bar{H}\bar{a}))^2$ as well; i.e., they are the same well outside the horizon. Another quantity which I have used extensively is Bardeen, Steinhardt and Turner's ζ [175, 232, 192, 178]:

$$\zeta = \varphi + \frac{(\delta \rho)_{\text{tot}}}{3(\bar{\rho} + \bar{p})_{\text{tot}}} = \varphi_{\text{com}} + \frac{{}^{(3)}\bar{\nabla}^2 \varphi_L}{3(\bar{H})^2(1 + \bar{q})}. \quad (168)$$

Again, far outside the “horizon”, $k/\bar{H}\bar{a} \ll 1$, $\zeta \approx \varphi_{\text{com}} \approx \delta \ln a|_{H_*}$.

The most commonly used gauge choice in nonlinear numerical relativity computations of black hole formation has been one on which K is not just uniform but is zero, because it turns out to be the time slicing which maximizes the 3-volume. This has the virtue of avoiding singularities, but in cosmology we usually care about following collapsing objects. The cosmological generalization of maximal slicing is one on which the Hubble parameter is uniform, i.e., basically the hypersurfaces whose scalar curvature parameter φ is used to characterize the initial conditions for adiabatic fluctuations.

6.2.4 Longitudinal and synchronous gauges

For scalar perturbations, there are no intrinsic dynamics of the gravitational field. This is similar to Newtonian theory in which, given the density, the Newtonian potential is found by solving the Poisson equation, but no ODEs in time. To specify the gauge, a single combination of the scalar degrees of freedom can be fixed. Two have been most widely used in cosmological calculations of radiative transport. In the longitudinal gauge (L), ψ and Ψ_n are both set to zero. Thus Ψ_σ is also set to zero, so the hypersurfaces have zero shear. The remaining two metric variables in Bardeen's notation [171] are $\Phi_A = \nu_L$, $\Phi_H = \varphi_L (= \delta \ln a)$. In the longitudinal gauge, one can use the anisotropic dynamical Einstein equation to algebraically relate $(\nu_L - \varphi_L)$ to the anisotropic stress, and the Poisson–Newton equation to get ν_L in terms of the total comoving energy density. All dynamical information is then carried by the matter present. Refs. [173, 138, 251] adopt this approach, but instead of solving for the longitudinal gauge $\delta_{\text{type}L}$, they solve for the comoving densities $\delta_{\text{typecom}} = \delta_{\text{type}} + 3(1 + \frac{\bar{p}}{\bar{\rho}})_{\text{type}} \bar{H} \Psi_{v, \text{type}}$. Ref. [260] solves for $\delta_{\text{type}L}$, but uses the momentum constraint equation in place of the anisotropic shear equation. The longitudinal gauge is considered to be the one closest to Newtonian in the nonrelativistic regime, for both metric variables are given by the perturbed Newtonian potential Φ_N : $\Phi_A \rightarrow \Phi_N$, $\Phi_H \rightarrow -\Phi_N$.

In the synchronous gauge (S), the lapse perturbation ν and Ψ_n are set to zero. One could solve the momentum constraint equation as a first order

ODE for φ_S in terms of the various velocity potentials, and then determine $\bar{a}(\delta H)$, $\propto$ the perturbation to the trace of the extrinsic curvature, through the energy constraint. We recommended this approach for inflation in [192]. The perturbed equations in the synchronous gauge for radiation and matter involve only $\bar{a}(\delta H)$ and $\dot{\varphi}_S$, so actually solving for φ_S is not really necessary, except to get $\bar{a}(\delta H)$. However, $\bar{a}(\delta H)$ can be directly determined from the Raychaudhuri equation, which is just an ODE at each point, and does not depend upon φ_S or even $\dot{\varphi}_S$. If one uses this equation to evolve $\bar{a}(\delta H)$, $\dot{\varphi}_S$ is set algebraically by the momentum constraint, and φ_S is not solved for. The transport equation for CDM in the synchronous gauge is simply $\dot{\delta}_{cdm} = -\bar{a}(\delta H)$. Transforming to normal $\bar{N} = 1$ time gives the usual density perturbation growth equation, $\dot{\delta}_{cdm} + 2\bar{H}\delta_{cdm} = 4\pi G_N \bar{\rho}_{cdm} \delta_{cdm} + 4\pi G_N \sum_{\text{type} \neq cdm} (\delta \rho_{\text{type}} + 3\delta p_{\text{type}})$.

The spatial hypersurfaces are those on which cold dark matter is comoving (the velocity potential for cdm particles in the longitudinal gauge is $\Psi_{v,cdm,L} = \Psi_{\sigma,S}$). In the nr limit, $\psi_S \rightarrow \Phi_N / (4\pi G_N \bar{\rho}_{nr} \bar{a}^2)$ becomes the potential for the displacement field that appears in the mapping from Lagrangian space to Eulerian space, $\mathbf{x}(\mathbf{r}, \tau) = \mathbf{r} - \mathbf{s}(\mathbf{r}, \tau)$, $\mathbf{s}(\mathbf{r}, \tau) = \nabla \psi_S$. The coordinates $x^I(\mathbf{r}, \tau)$ are Eulerian ones appropriate to the longitudinal gauge and the r^i coordinates are Lagrangian ones labelling the cold matter particles. The deformation tensor $e_{*i}^J = \partial x^J / \partial r^i$ defines a triad of orthogonal vectors e_{*i}^J for the space in which the cold dark matter is at rest (i.e., perpendicular to the flow e_n^α of the CDM). The way gravitational collapse manifests itself is through the shrinking of the comoving lengths $e_{*i}^J dr^i$, although dr^i remains fixed: i.e., collapse is viewed as a motionless distortion of the geometry. The synchronous gauge breaks down once e_{*i}^J becomes singular, i.e., once caustics form and shell crossing occurs. (Once the universe has become dominated by nonrelativistic matter, $\dot{\varphi}_S \rightarrow 0$, $\varphi_S \rightarrow -5\Phi_N/3$, $\bar{a}^{-1}\Psi_\sigma \rightarrow \tau\Phi_N/3$, $\psi_S = \partial_\tau(\bar{a}^{-1}\Psi_\sigma) \rightarrow \Phi_N/3$).

6.2.5 Tensor mode metric equations

The tensor modes satisfy the anisotropic $\delta G_I^J - \frac{1}{3}\delta_J^I \delta G_K^K$ Einstein equations. Expressed in conformal time this is

$$\begin{aligned} \ddot{h}_i^{(TT)j} + 2\bar{H}\bar{a}\dot{h}_i^{(TT)j} - \bar{a}^2 {}^{(3)}\bar{\nabla}^2 h_i^{(TT)j} + \frac{1}{3}\bar{a}^2 {}^{(3)}\bar{R}h_i^{(TT)j} \\ = 16\pi G_N \bar{a}^2 (\Pi_{\text{tot}}^{(T)})_i^j, \end{aligned} \quad (169)$$

where $(\Pi_{\text{tot}}^{(T)})_i^j$ is the tensor mode part of the anisotropic stress. The mode expansion (147) for gravity waves (in a flat FRW background) is

$$\begin{aligned} h_{ij}^{(TT)}(\mathbf{x}, \tau) \\ = w \sum_{\epsilon=\pm, \times} \sum_{\mathbf{k} \in \mathcal{M}} h_{(T\epsilon)} E_{ij}^{(T\epsilon)} e^{i\mathbf{k} \cdot \mathbf{x}} a_{\mathbf{k}(T\epsilon)} + h_{(T\epsilon)}^* E_{ij}^{(T\epsilon)*} e^{-i\mathbf{k} \cdot \mathbf{x}} a_{\mathbf{k}(T\epsilon)}^\dagger, \\ E^{(T+)} = (\mathbf{e}_1 \otimes \mathbf{e}_1 - \mathbf{e}_2 \otimes \mathbf{e}_2), \quad E^{(T\times)} = (\mathbf{e}_1 \otimes \mathbf{e}_2 + \mathbf{e}_2 \otimes \mathbf{e}_1), \\ \mathbf{k} \cdot \mathbf{e}_{\{1,2\}} = 0, \quad \mathbf{e}_1 \cdot \mathbf{e}_2 = 0. \end{aligned} \quad (170)$$

Here $\{\mathbf{e}_1, \mathbf{e}_2, \hat{k}\}$ form an orthonormal triad. With the wavevector $\mathbf{k}$ oriented in the z -direction, $\mathbf{e}_1$ in the x -direction and $\mathbf{e}_2$ in the y -direction, we have the usual

$$h_{(T+)}(\mathbf{k}, \tau) = \frac{E^{(T+)}{}^{ij} h_{ij}(\mathbf{k}, \tau)}{E^{(T+)}{}^{ij} E_{ij}^{(T+)}} = (h_{11} - h_{22})/2,$$

$$h_{(T\times)}(\mathbf{k}) = \frac{E^{(T\times)}{}^{ij} h_{ij}(\mathbf{k}, \tau)}{E^{(T\times)}{}^{ij} E_{ij}^{(T\times)}} = h_{12}.$$

The only degree of freedom of the stress energy tensor which has a nonzero amplitude in the tensor mode is the anisotropic stress, which has components

$$\bar{p}_{\text{tot}} \pi_{t, \text{tot}}^{(T\{+, \times\})} \equiv \frac{E^{(T\{+, \times\})}{}^{ij} \Pi_{\text{tot } ij}(\mathbf{k}, \tau)}{E^{(T\{+, \times\})}{}^{ij} E_{ij}^{(T\{+, \times\})}}.$$

Hence (for flat backgrounds) the Einstein equations reduce to

$$\ddot{h}_{(T\epsilon)} + 2\bar{H}\bar{a}\dot{h}_{(T\epsilon)} + k^2 h_{(T\epsilon)} = 16\pi G_N \bar{a}^2 \bar{p}_{\text{tot}} \pi_{\text{tot}}^{(T\epsilon)}, \quad \epsilon = +, \times. \quad (171)$$

For inflation-based models in which gravity waves are zero point fluctuations, the anisotropic stress driver can be ignored, even during evolution through the radiation-dominated epoch where the anisotropic stress may not be irrelevant. This contrasts with the scalar perturbation case for which the Einstein equations have source terms depending upon the density and the velocity potential, and one cannot solve for the metric variables without simultaneously solving for the radiation and matter. For tensor perturbations the predominant behavior is just free evolution from a given set of initial conditions for the waves.

The solution for $k \ll Ha$ is $h_{(T\{+, \times\})}$ constant for all relevant equations of state. Let us suppose that the gravitational waves are characterized by a power spectrum $\mathcal{P}_{h_{(T\{+, \times\})}}(k, \tau_e)$ at some initial time τ_e for which $k\tau_e \ll 1$. To see the character of the solutions for $k > Ha$, let us consider the case where we have only relativistic particles with density parameter now given by $\Omega_{er,0}$ (photons and massless or very light neutrinos) and nonrelativistic particles (baryons and cold dark matter say) with density parameter now $\Omega_{nr,0}$. When eq. (169) is expanded in $\mathbf{k}$ -modes, $h_{(T\{+, \times\})}$ obeys

$$\frac{d^2 h_{(T\epsilon)}}{dx^2} + 2 \frac{x/(2x_*) + a_{eq}^{1/2}}{x/(4x_*) + a_{eq}^{1/2}} \frac{1}{x} \frac{dh_{(T\epsilon)}}{dx} + h_{(T\epsilon)} = 0, \quad x \equiv k(\tau - \tau_e), \quad (172)$$

where $x_* = k[H_0\Omega_{nr,0}^{1/2}]^{-1}$ and

$$a - a_e = \frac{1}{4}[H_0\Omega_{nr,0}^{1/2}(\tau - \tau_e)]^2 + a_{eq}^{1/2}[H_0\Omega_{nr,0}^{1/2}(\tau - \tau_e)], \quad a_{eq} \equiv \frac{\Omega_{er,0}}{\Omega_{nr,0}}. \quad (173)$$

For waves which reach $Ha \sim k$ in the er -dominated regime, the solution is

$$\mathcal{P}_{h_{(T\{+, \times\})}}(k, \tau) = \mathcal{P}_{h_{(T\{+, \times\})}}(k, \tau_e) (j_0(k(\tau - \tau_e)))^2, \quad a - a_e \ll a_{eq}, \quad (174)$$

while for those which reach $Ha \sim k$ in the nr -dominated regime it is

$$\mathcal{P}_{h_{(T\{+, \times\})}}(k, \tau) = \mathcal{P}_{h_{(T\{+, \times\})}}(k, \tau_e) \left(\frac{3j_1(k(\tau - \tau_e))}{k(\tau - \tau_e)} \right)^2, \quad a - a_e \gg a_{eq}. \quad (175)$$

It is more complicated in the transition region or if there are other constituents in the equation of state, such as vacuum energy, decaying particles, light massive neutrinos, etc. These solutions explicitly show the constancy outside of the “horizon” and the loss of power due to the free-streaming of gravity wave perturbations inside the “horizon”, i.e., for $k(\tau - \tau_e) \gg 1$.

6.3 Connection with primordial post-inflation power spectra

The evolution equations for fluctuations $\delta\phi_j$ in scalar fields ϕ_j are

$$\begin{aligned} \frac{\partial^2 \delta\phi_j}{\partial \tau^2} + 2\bar{H}\bar{a} \frac{\partial \delta\phi_j}{\partial \tau} - \bar{a}^2 {}^{(3)}\bar{\nabla}^2 \delta\phi_j + \sum_i \bar{a}^2 \frac{\partial^2 V}{\partial \phi_j \partial \phi_i} \delta\phi_i \\ = -\dot{\bar{\phi}}_j (3\dot{\varphi} - \dot{\nu} - {}^{(3)}\bar{\nabla}^2 \bar{a} \Psi_\sigma) - 2 \frac{\partial V}{\partial \phi_j} \nu. \end{aligned} \quad (176)$$

The inflaton and isocons are coupled through a potential $V(\phi_{inf}, \phi_{is}, \dots)$. No explicit dissipative coupling term has been included (but is needed to turn oscillating scalar field into radiation and matter). The combinations

$$\delta\phi_j - \dot{\bar{\phi}}_j \bar{a}^{-1} \Psi_\sigma = \delta\phi_{jL} \quad \text{and} \quad \delta\phi_j + (\bar{H}\bar{a})^{-1} \dot{\bar{\phi}}_j \varphi = \delta\phi_j|_{\ln a} \quad (177)$$

are gauge invariant. Thus the perturbation in the longitudinal gauge is an invariant combination $\delta\phi_{jL}$. Scalar fields have no anisotropic stress in linear perturbation theory. For the longitudinal gauge this gives the simple relation $\nu_L = -\varphi_L$. The second gauge invariant combination, $\delta\phi_j|_{\ln a} = \delta\phi_{jL} + \frac{d\bar{\phi}_j}{d\ln a} \delta\ln a$, is the value of the scalar field on hypersurfaces of fixed $\ln a$; i.e., beginning in the longitudinal gauge, one forms $\phi_{\ln a} = \phi(\mathbf{x}, \tau + T(\mathbf{x}, \tau))$ with T defined by $\ln a(\mathbf{x}, \tau + T(\mathbf{x}, \tau))$ is constant. Mukhanov [177] showed in perturbation theory that the metric terms disappear in the scalar field evolution equation when the choice $\bar{a}\delta\phi_j|_{\ln a}$ is made, resulting in considerable simplification for the case of a single scalar field being important in inflation. In [181], we emphasized some of the virtues in the nonlinear case.

The equation for $h_{(T\{+, \times\})}$ is identical to that for scalar fields with no effective mass. (There is a (δH) term multiplying $h_{(T\{+, \times\})}$, but this is an ignorable quadratic nonlinearity.) A factor is required to make the actions the

same: $(m_{\mathcal{P}}/\sqrt{16\pi})h_{(T\{+, \times\})}$, where the Planck mass is related to Newton's gravitational constant by $m_{\mathcal{P}} \equiv G_N^{-1/2}$ in units with $\hbar = c = 1$.

We now describe the power spectra resulting from zero-point quantum oscillations found by solving these equations. During inflation Ha increases. The solution of the massless scalar equation shows rapid oscillation of the respective mode functions “inside the horizon”, almost freeze-out outside ($k < Ha$), with a power amplitude $\mathcal{P}_{\phi}^{1/2}(k, \tau_k) \approx H(\tau_k)/(2\pi)$ on the $k = Ha$ boundary. The Hawking temperature $H/(2\pi)$ result³ follows from a WKB solution to (169) evaluated at $k = Ha$ provided the effective masses of the scalars are small compared with H . The perturbation in the inflaton field ϕ_{inf} translates into scalar perturbations in the metric through $\delta \ln a = (H/\dot{\phi}_{inf})\delta\phi_{inf}$. If we denote the end of inflation by τ_e and horizon crossing by τ_k , the post-inflation spectra are

$$\begin{aligned} \mathcal{P}_{\ln a|_{H^*}}^{1/2}(k, \tau_e) &= \frac{1}{\sqrt{q+1}} \frac{\sqrt{4\pi}}{m_{\mathcal{P}}} \frac{H(\tau_k)}{2\pi} e^{u_s}, \\ \mathcal{P}_{GW}^{1/2} &= \sqrt{8} \frac{\sqrt{4\pi}}{m_{\mathcal{P}}} \frac{H(\tau_k)}{2\pi} e^{u_t}, \quad \mathcal{P}_{GW}(k) \equiv \mathcal{P}_{h_{(T+)}}(k) + \mathcal{P}_{h_{(T\times)}}(k). \end{aligned} \quad (178)$$

The correction factors u_t and u_s to “the $H/(2\pi)$ at $k = Ha$ WKB approximation” are in practice nearly zero. How near is now of considerable interest because the COBE results have created a desire for calculational precision [184, 6]. Complicated potential interactions between the inflaton and isocon degrees of freedom can also change these results.

In eq. (178), $H(\phi)$ and the deceleration parameter $q(\phi)$ are treated as functions of the inflaton field. These functions naturally follow from the Hamilton–Jacobi formulation [181, 182], in which $H(\phi)$ is related to the potential $V(\phi)$ by

$$H^2 = \frac{8\pi}{3m_{\mathcal{P}}^2} \left[\frac{1}{2} \left(\frac{m_{\mathcal{P}}^2}{4\pi} \frac{\partial H}{\partial \phi} \right)^2 + V(\phi) \right], \quad (1+q) = \frac{m_{\mathcal{P}}^2}{4\pi} \left[\frac{\partial \ln H}{\partial \phi} \right]^2. \quad (179)$$

The oft-used slow rollover approximation, valid in many inflation calculations but certainly not all, is the zeroth order solution in an expansion in $1+q$: $H^2 \approx 8\pi V/(3m_{\mathcal{P}}^2)$.

The adiabatic scalar and tensor tilts are logarithmic derivatives of eq. (178):

$$\begin{aligned} \frac{\nu_t}{2} &= \frac{n_t + 3}{2} = 1 + q^{-1} + C_t, \\ \frac{\nu_s}{2} &= \frac{n_s - 1}{2} = 1 + q^{-1} - q^{-1} \frac{m_{\mathcal{P}}^2}{4\pi} \frac{\partial^2 \ln H}{\partial \phi^2} + C_s. \end{aligned} \quad (180)$$

³In stochastic inflation, noise at the Hawking temperature radiates from short distances across the decreasing $(Ha)^{-1}$ boundary into an inhomogeneous background field built from longer wavelengths.

Here $C_{t,s}$ are small and essentially ignorable correction factors associated with derivatives of the $u_{t,s}$.

Eq. (180) shows that tilt mostly depends upon how far the acceleration is below the *critical* value of unity. For uniform acceleration, the scalar and tensor tilts are equal:

$$\nu_s = \nu_t = -2(p-1)^{-1}, \quad q+1 = p^{-1}. \quad (181)$$

It is realized by power law inflation, $a \propto t^p$, with p constant, and an exponential potential in ϕ . Over the small observable window we have in k -space, this is often a good approximation, e.g., for extended inflation, one of a class of theories with variable gravitational coupling. Of course, q must go negative for a viable model of inflation. Power law potentials of the form $V(\phi) = \lambda_e m_{\mathcal{P}}^4 (\phi/m_{\mathcal{P}})^{2n}/(2n)$ with n constant have the acceleration naturally dropping through zero: $q \approx -1 + (\phi/m_{\mathcal{P}})^{-2} n^2/(4\pi)$. In chaotic inflation examples [188], one often takes power law potentials with $n = 1$ or 2 . A characteristic of such potentials is that the range of values of ϕ which correspond to all of the large scale structure that we observe is actually remarkably small: e.g., for $n = 2$, the region of the potential curve responsible for the structure between the scale of galaxies and the scales up to our current Hubble length is just $4m_{\mathcal{P}} \lesssim \phi \lesssim 4.4m_{\mathcal{P}}$ [192]. Consequently, $H(\phi)$ does not evolve by a large factor over the large scale structure region and we therefore expect approximate uniformity of $\nu_s(k)$ and $\nu_t(k)$ over the narrow observable bands of k -space, and near-scale-invariance for both. Although this is usually quoted in the form of a logarithmic correction to the $\ln a|_{H^*}$ -spectrum, a power law approximation is quite accurate [190]:

$$\begin{aligned} \nu_s(k) &\approx -\frac{n+1}{N_I(k) - n/6}, & \nu_t &\approx -\frac{n}{N_I(k) - n/6}, \\ q+1 &\approx \frac{n/2}{N_I(k) + n/3}. \end{aligned}$$

$N_I(k)$ is the number of e-foldings from the point at which wavenumber k “crosses the horizon” (when $k = Ha$) and the end of inflation. For waves the size of our current Hubble length we have the familiar $N_I(k) \sim 60$, hence $\nu_s \approx -0.05, \nu_t \approx -0.03$ for $n = 2$ and $\nu_s \approx -0.03, \nu_t \approx -0.02$ for $n = 1$ (massive scalar field case). Further, the observable scales are sufficiently far from the reheating scale that N_I is relatively large over the observable range: e.g., over the range from our Hubble radius down to the galaxy scale, ν_s decreases by only about 0.01.

In natural inflation [189, 190], the inflaton for the region of k -space that we can observe is identified with a pseudo-Goldstone boson with a potential $V = 2\Lambda^4 \sin^2(\phi/(2f))$. This is similar to the axion, except that the symmetry breaking scale f is taken to be of order $m_{\mathcal{P}}$ and the energy scale for the potential is taken to be of order the grand unified scale, m_{GUT} , so that an effective weak coupling, $\lambda_e = \Lambda^4/(fm_{\mathcal{P}})^2 \sim (m_{GUT}/m_{\mathcal{P}})^4$ arises “naturally”, giving the required 10^{-13} for $m_{GUT} = 10^{16}\text{GeV}$. To obtain sufficient inflation and

a high enough post-inflation reheat temperature for baryogenesis, $f \gtrsim 0.3m_P$ is required. To have a tilted spectrum and also get enough inflation in our Hubble patch, ϕ/f must have started near the maximum at π , an inflection point where q is nearly -1 , hence tensor tilt and gravity wave power are both exponentially-suppressed; however the scalar $\nu_s \approx -m_P^2/(8\pi f^2)$ does not have to be small [190].

The index ν_s can have complex k -dependent structure when the acceleration changes considerably over the k -band in question. According to eqs. (180), the post-inflation gravitational wave spectrum will have power increasing with wavelength, whereas artfully using the $\partial^2 \ln H / \partial \phi^2$ term allows essentially any prescribed shape for the adiabatic scalar spectrum (e.g., [191, 192]). However, most broken scale invariance models which do give considerable variation in $\nu_s(k)$ over the relatively narrow band of k -space that we can observe are not very compelling, since rather dramatic features must be tuned to lie on the potential surface in just that stretch which corresponds to our observable band. A slowly varying $\nu_s(k)$ is certainly a better bet.

6.4 Relating scalar and tensor power measures to the dmr band-power

For early universe calculations and also to characterize the initial conditions for the photon transport through decoupling, the power in adiabatic scalar fluctuations on scales beyond the Hubble radius is best characterized in terms of quantities which become time-independent. We have seen that some examples are the spatial curvature of time surfaces on which there is no net flow of momentum, the expansion factor fluctuation on time surfaces with uniform space creation rate and ζ . An initially scale invariant adiabatic spectrum has k -independent power per $d \ln k$ in these variables (for $k/(\bar{H}\bar{a}) \ll 1$), while for models with spectral tilt ν_s , we have $\mathcal{P}_{\varphi_{com}}(k) = \mathcal{P}_{\varphi_{com}}(\tau_0^{-1})(k\tau_0)^{\nu_s}$, where we use the instantaneous comoving horizon size at the current epoch, τ_0 , as the normalization point. For CDM-like models (those with $\Omega = \Omega_{nr} = 1$ and $\tau_0 = 2H_0^{-1}$), these spectra are related to the portion of the dmr band power $\langle \mathcal{C}_\ell \rangle_{dmr}$ in the scalar adiabatic mode, $\langle \mathcal{C}_\ell^{(S)} \rangle_{dmr} = \langle \mathcal{C}_\ell \rangle_{dmr} / (1 + \tilde{r}_{ts})$, and to the quadrupole power, $\mathcal{C}_2^{(S)} = \mathcal{C}_2 / (1 + r_{ts})$, by

$$\begin{aligned} \mathcal{P}_{\varphi_{com}}(\tau_0^{-1}) &\approx 23.4 \langle \mathcal{C}_\ell^{(S)} \rangle_{dmr} e^{-1.99\nu_s(1+0.1\nu_s)} \approx 23.5 \mathcal{C}_2^{(S)} e^{-1.1\nu_s}, \quad (182) \\ \mathcal{P}_{\varphi_{com}}(k) &\approx \mathcal{P}_{\ln a|_{H*}}(k) \approx \mathcal{P}_\zeta(k) \quad \mathcal{P}_{\varphi_{com}}(k) = \mathcal{P}_{\varphi_{com}}(\tau_0^{-1})(k\tau_0)^{\nu_s}, \end{aligned}$$

i.e., roughly 3×10^{-9} . This relation, determined for the $\Omega_B = 0.05, h = 0.5$ CDM model, is quite insensitive to variations in h and Ω_B (e.g., 23.6 to 23.0 as Ω_B rises from 0.0125 to 0.20 for $\langle \mathcal{C}_\ell^{(S)} \rangle_{dmr}$ and almost no change for $\mathcal{C}_2^{(S)}$). For scales of order our present Hubble size, we also have $\mathcal{P}_\zeta \approx \frac{25}{9} \mathcal{P}_{\Phi_H} \approx \frac{25}{4} \mathcal{P}_{(\delta\rho)_{hor}}$, where $\Phi_H \equiv \varphi_L = -\nu_L = -\Phi_N$ is the perturbed Newtonian gravitational potential and $(\delta\rho)_{hor}$ is the density fluctuation (in the synchronous gauge) at “horizon crossing”, $k\tau = 1$.

Quantum noise in the transverse traceless modes of the perturbed metric tensor would also have arisen in the inflation epoch and for many models may have been quite significant, as is discussed below. The gravitational radiation power spectrum $\mathcal{P}_{GW} = \mathcal{P}_{h_+} + \mathcal{P}_{h_\times}$ is the sum of the two independent gravitational wave polarizations. It is related to the amplitude of the dmr band power $\langle \mathcal{C}_\ell^{(T)} \rangle_{dmr} = \langle \mathcal{C}_\ell \rangle_{dmr} \tilde{r}_{ts} / (1 + \tilde{r}_{ts})$ and to the quadrupole $\mathcal{C}_2^{(T)}$ by

$$\begin{aligned} \mathcal{P}_{GW}(\tau_0^{-1}) &\approx 17.6 \langle \mathcal{C}_\ell^{(T)} \rangle_{dmr} e^{-1.92\nu_t(1+0.1\nu_t)} \approx 13.7 \mathcal{C}_2^{(T)} e^{-1.25\nu_t}, \\ \mathcal{P}_{GW}(k) &= \mathcal{P}_{GW}(\tau_0^{-1})(k\tau_0)^{\nu_t}, \end{aligned} \quad (183)$$

with very little Ω_B dependence.

The inflation model determines the ratio of $\mathcal{P}_{GW}(\tau_0^{-1})$ to $\mathcal{P}_{\varphi_{com}}(\tau_0^{-1})$, through eq.(178), which is related to the tilt of the gravity wave spectrum, $-4\nu_t/(1 - \nu_t/2)$ in zeroth order, with small corrections associated with C_t and $u_t - u_s$ predominantly dependent upon $\nu_t - \nu_s$ and which can usually be ignored [184, 6]. The fits given above can then be used to relate the ratio of dmr band-powers (and quadrupoles) to the tilts (for $\Omega_{vac} = \Omega_{curv} = 0$):

$$\begin{aligned} \tilde{r}_{ts} &\equiv \frac{\langle \mathcal{C}_\ell^{(T)} \rangle_{dmr}}{\langle \mathcal{C}_\ell^{(S)} \rangle_{dmr}} \approx 1.33 \left[\frac{\mathcal{P}_{GW}(\tau_0^{-1})}{\mathcal{P}_{\varphi_{com}}(\tau_0^{-1})} \right] e^{-0.07\nu_t} e^{-1.99(\nu_s - \nu_t)}, \\ r_{ts} &\equiv \frac{\mathcal{C}_2^{(T)}}{\mathcal{C}_2^{(S)}} \approx 1.71 \left[\frac{\mathcal{P}_{GW}(\tau_0^{-1})}{\mathcal{P}_{\varphi_{com}}(\tau_0^{-1})} \right] e^{-0.15\nu_t} e^{-1.1(\nu_s - \nu_t)}, \\ \left[\frac{\mathcal{P}_{GW}(\tau_0^{-1})}{\mathcal{P}_{\varphi_{com}}(\tau_0^{-1})} \right] &= 8(1 + q)e^{2(u_t - u_s)} \approx \frac{-4\nu_t}{(1 - \nu_t/2)}. \end{aligned} \quad (184)$$

Recall from eq.(180) that the tensor tilt is simply related to the deceleration parameter $q = -a\ddot{a}/\dot{a}^2$ of the Universe in the inflationary epoch, $\nu_t/2 \approx 1 + q^{-1}$; although $1 + q^{-1}$ is the leading term for the scalar tilt, other terms can dominate when the deceleration is near the critical deSitter-space value of -1 . It is invariably negative. When assessing the effect of gravity waves on the normalization of the spectrum, as noted earlier it is useful to consider two limiting cases: $\nu_s \approx \nu_t$, which holds for the widest class of models, including power law and chaotic inflation, and $\nu_t \approx 0$, with ν_s arbitrary, which holds for some models such as “natural” inflation. To lowest order in ν_t , $\tilde{r}_{ts} \approx -5.3\nu_t$ and $r_{ts} \approx -6.8\nu_t$ (often rounded up to $-7\nu_t$, which is nearly the value one gets if only the naive Sachs-Wolfe formula is used to estimate $\mathcal{C}_2^{(S)}$.)

There are also corrections as one goes away from the $\Omega_{nr} = \Omega = 1$ models. For example, models with nonzero cosmological constant Ω_{vac} , but $\Omega_{nr} + \Omega_{vac} = 1$, have $\mathcal{P}_{GW}/\langle \mathcal{C}_\ell^{(T)} \rangle_{dmr}$ being only weakly dependent upon Ω_{vac} whereas $\mathcal{P}_\zeta/\langle \mathcal{C}_\ell^{(S)} \rangle_{dmr} \propto (1 - 0.6\Omega_{vac}^{3.5})$ is strongly dependent upon it (section 7.3).

6.5 The Boltzmann transport equation

In Appendices B and C, transport theory with polarization in the ADM framework is derived with full nonlinearities in the gravitational field included. The Boltzmann transport equation depends upon not only the spacetime foliation chosen, but also upon the momentum variables q^I chosen. A natural set to select are those referred to the orthonormal basis $\{e_n, e_I\}$, where the spatial triad e_I is normal to the hypersurface flow vector e_n^α ; however, this choice, p^I , is a physical momentum, not a comoving momentum, and we have seen that the transport equations are much simpler if we use comoving momentum, hence use $q^I = \Omega p^I$. The factor Ω must reduce to the average expansion factor, $\bar{a}$, in the unperturbed case, but can be inhomogeneous if we like. Choosing different $\Omega/\bar{a}$ corresponds to choosing different momentum-space gauges, and leads to different forms for the Boltzmann transport equation. This is only one example of the extra gauge transformation freedom that exists in dealing with transport phenomena. Momenta defined with respect to tetrads other than the $\{e_n, e_I\}$ also lead to modified equations.

Recall from section 3.1 that to treat polarized photons, four distribution functions are required, f_t, f_U, f_V, f_Q , corresponding to the four Stokes parameters. These are best understood as elements of a 2×2 polarization matrix $f_{ss'}$, where s denotes the photon polarization, $s = \pm 1$ for circular polarization, $s = 1, 2$ say for linear polarization, with associated polarization basis ε_s satisfying $\varepsilon_s \perp \hat{q}$. One can combine the tensor product basis $\varepsilon_s \otimes \varepsilon_{s'}$ with $f_{ss'}$ to make a (spatial) tensor of rank 2 which is conceptually extremely useful for understanding the Stokes parameters and how they behave under rotations:

$$\mathbf{f} = \sum_{ss'} f_{ss'} \varepsilon_s \otimes \varepsilon_{s'} = f_t \mathcal{E}_{(t)} + f_U \mathcal{E}_{(U)} + f_V \mathcal{E}_{(V)} + f_Q \mathcal{E}_{(Q)}. \quad (185)$$

The tensor basis $\mathcal{E}_{(\mu)}$ for the Stokes parameters are linear combinations of the $\varepsilon_s \otimes \varepsilon_{s'}$, defined by eq. (309). In Appendix C, the source function for Thomson scattering is derived using this language. A more conventional approach is to apply Chandrasekhar's classic development of the scattering source term for Rayleigh (and thus Thomson) scattering in a plane parallel atmosphere [200]. Of course, the transfer problem is not plane parallel; however, it effectively becomes so for each normal Fourier transform mode for flat universes [134]. The nonlinear Boltzmann transport equation for $\Delta_{\{t,Q,U,V\}}$ is given by eq. (284). The linearized version for photons takes the form (using $\bar{N}=\bar{a}$, i.e., conformal time)

$$\begin{aligned} & \frac{\partial}{\partial \tau} \Delta_{\{t,Q,U,V\}} + \hat{q}^J \frac{\partial}{\partial x^J} \Delta_{\{t,Q,U,V\}} \\ & = \mathcal{G}_{\{t,Q,U,V\} SW} + \mathcal{G}_{\{t,Q,U,V\} curv} + \mathcal{G}_{\{t,Q,U,V\} C}, \end{aligned} \quad (186)$$

in terms of a Sachs–Wolfe source from redshift effects, a source associated with mean curvature of the Universe, $\mathcal{G}_{\{t,Q,U,V\} curv}$, and the Thomson scattering source, $\mathcal{G}_{\{t,Q,U,V\} C}$. In nonlinear theory, there is a term associated with the

bending (lensing) of light, $\mathcal{G}_{\text{tbend}}$, but to linear order it manifests itself only if there is mean curvature. Mean curvature terms, grouped into $\mathcal{G}_{\{t,Q,U,V\} \text{ curv}}$, are described in section C.4. Although the solution method when there is mean curvature is quite similar to the flat case one, the discussion is complicated because the mode functions are not plane waves. For this section, we assume a flat background so the modes are characterized by a comoving wavevector k_I , with the understanding that we really mean the action of the operator $-i\bar{a}^{(3)}\nabla_I$.

6.5.1 Scalar mode transfer equations

For scalar perturbations (and flat universes – see section C.4 for mean curvature modifications),

$$\mathcal{G}_{tSW}^{(S)} = -i\hat{q} \cdot \hat{k} k \nu - \dot{\varphi} - (\hat{q} \cdot \hat{k})^2 k^2 \bar{a}^{-1} \Psi_\sigma, \quad (187)$$

$$\mathcal{G}_{QSW}^{(S)} = 0, \quad \mathcal{G}_{USW}^{(S)} = 0, \quad \mathcal{G}_{VSW}^{(S)} = 0,$$

$$\begin{aligned} \tau_C \mathcal{G}_{tC}^{(S)} &= -\Delta_t^{(S)} + \Delta_{t0}^{(S)} - i\hat{q} \cdot \hat{k} k \bar{a}^{-1} \Psi_{v,B} \\ &\quad - \frac{1}{2} P_2(\hat{q} \cdot \hat{k}) (\Delta_{t2}^{(S)} + \Delta_{Q2}^{(S)} + \Delta_{Q0}^{(S)}). \end{aligned} \quad (188)$$

$$\tau_C \mathcal{G}_{QC}^{(S)} = -\Delta_Q^{(S)} + \frac{1}{2} (1 - P_2(\hat{k} \cdot \hat{q})) (\Delta_{Q0}^{(S)} + \Delta_{t2}^{(S)} + \Delta_{Q2}^{(S)}), \quad (189)$$

$$\tau_C \mathcal{G}_{UC}^{(S)} = 0, \quad \tau_C \mathcal{G}_{VC}^{(S)} = -\Delta_V^{(S)} + \frac{3}{4} \hat{q} \cdot \hat{k} \Delta_{V1}^{(S)}. \quad (190)$$

The moments $\Delta_{\{t,Q,U,V\} \ell}^{(S)}$ are defined by expanding in Legendre polynomials:

$$\Delta_{\{t,Q,U,V\} \ell}^{(S)} = \sum_{\ell} (2\ell + 1) (-i)^\ell \Delta_{\{t,Q,U,V\} \ell}^{(S)} P_\ell(\hat{k} \cdot \hat{q}). \quad (191)$$

$\Delta_U^{(S)}$ and $\Delta_V^{(S)}$ are initially zero, and remain decoupled from other sources, so remain zero. Thus for each eigenmode only transport equations for $\Delta_t^{(S)}$ and $\Delta_Q^{(S)}$ need to be solved. However, when one reconstructs from these mode solutions the statistical distribution of the observed polarization pattern, $\Delta_{\{Q,U,V\}}(\mathbf{x}, \tau_0, \hat{q})$, the U and Q components mix, so both are nonzero.

These equations are valid for arbitrary gauge choices. The momenta have been chosen as the components defined with respect to an orthogonal basis so that the direction $\hat{q}$ is what would be measured. The change of momentum coordinates is itself part of a gauge transformation.⁴ It is often convenient in dealing with the transport equation to rewrite it with many of the explicit terms of form $\hat{q}^I {}^{(3)}\nabla_I$ brought into the transport operator. This is equivalent to a momentum component transformation. An explicit example is to use momenta

$$\tilde{q}^I = \exp[\nu + (\partial_\tau - \hat{q}^i \partial_i)(\bar{a}^{-1} \Psi_\sigma)] q^I, \quad (192)$$

⁴The oft-used approach of viewing the CMB photon transport equation as one for the radiation brightness, the integral of the distribution function over q but not $\hat{q}$, obscures this view. Also for most sources but Thomson scattering or for other particles such as massive neutrinos, the q -dependence is very relevant.

which transforms the distribution function and source terms to the gauge-invariant combination $\tilde{\Delta}_t^{(S)}$ introduced in section 5.2:

$$\begin{aligned}\tilde{\Delta}_t^{(S)} &= \Delta_t^{(S)} + \nu + \frac{\partial \bar{a}^{-1} \Psi_\sigma}{\partial \tau} - \hat{q}^i \partial_i \bar{a}^{-1} \Psi_\sigma, \\ \tilde{\mathcal{G}}_{tSW}^{(S)} &= \dot{\nu} - \dot{\varphi} + \frac{\partial^2 \bar{a}^{-1} \Psi_\sigma}{\partial \tau^2}, \\ \frac{1}{4} \tilde{\delta}_\gamma &\equiv \frac{1}{4} \delta_\gamma + \nu + \frac{\partial}{\partial \tau} \bar{a}^{-1} \Psi_\sigma, \\ \tilde{\Psi}_{v,\gamma} &\equiv \Psi_{v,\gamma} + \Psi_\sigma, \quad \tilde{\Delta}_{t\ell}^{(S)} \equiv \Delta_{t\ell}^{(S)} \quad (\ell \geq 2).\end{aligned}\tag{193}$$

If the Compton sources are small and the gravitational potential perturbations do not change in time (or equivalently if the time flow has $\bar{a}^{-1} \Psi_\sigma \propto \tau$), then the source term can be neglected, $\tilde{\Delta}_t^{(S)}$ propagates freely, and is particularly simple to integrate [88]. We used this quantity extensively in [134, 88]; it has the simple interpretation in the longitudinal gauge of basically saying it is the Tolman combination $e^{\nu_L} T_\gamma$ which free-streams.

These equations are coupled to the transport equations for massless neutrinos, hot, warm and cold dark matter, and baryons. Massless neutrinos obey a transfer equation identical to that for photons, except of course there is no Compton coupling, only the Sachs–Wolfe and curvature sources. (If the relativistic particles are decay products generated during evolution, there is also a $\mathcal{G}_{erv,decay}^{(S)}$ source term [252].) Just as for photons, these equations are solved by a moment expansion. For hot dark matter (light massive neutrinos), warm dark matter, etc. the transport equation has the form

$$\begin{aligned}\partial_\tau [\Delta_{hdm}] + i \frac{q}{q^n} \hat{q} \cdot \hat{k} k \Delta_{hdm} &= (\mathcal{G}_{hdm SW} + \mathcal{G}_{hdm curv}), \\ \mathcal{G}_{hdm SW}^{(S)} &= -i \frac{q^n}{q} \hat{q} \cdot \mathbf{k} \nu - \dot{\varphi} - (\hat{q} \cdot \mathbf{k})^2 \bar{a}^{-1} \Psi_\sigma, \quad q^n = \sqrt{q^2 + m^2 \bar{a}^2}.\end{aligned}\tag{194}$$

It is the semi-relativistic stage, when q/q^n is not simply unity or $q/(m\bar{a})$, that creates the difficulty. A straightforward method is to solve this by moment expansion, one for each neutrino momentum, q . To feedback into the perturbed Einstein equations one needs to appropriately sum over the momenta q , to get the energy density, velocity, pressure and anisotropic stress – with an adequate number of moments and energy groups, and proper treatment of boundary conditions in ℓ -space, one can get away with only a few hundred extra coupled ODEs to be added to the already formidable number of moment equations used for the photons [260, 261]. In earlier work [195] we described another method, expressing the metric equations as integro-differential equations – which had the penalty of integrating over past time to get the current perturbed energy–momentum tensor of the neutrinos. To make the method practical and indeed relatively rapid for CMB anisotropy calculations, optimal sampling of the past history [261] and a switch into (essentially) cold dark matter equations once the

particles were strongly in the nonrelativistic regime and the wavenumber was much below the redshift-dependent Jeans wavenumber was helpful ([134, 2, 233] and eqs. (350), (352)).

The mass and momentum conservation equations for nr dark matter and for baryons are (eqs. (280), (282))

$$\text{CDM: } \frac{1}{3}\dot{\delta}_{cdm} + \dot{\varphi} + \frac{1}{3}k^2\bar{a}^{-1}(\Psi_{v,cdm} + \Psi_\sigma) = 0, \quad (195)$$

$$\bar{a}^{-1}\dot{\Psi}_{v,cdm} = \nu, \quad (196)$$

$$\text{baryons: } \frac{1}{3}\dot{\delta}_B + \dot{\varphi} + \frac{1}{3}k^2\bar{a}^{-1}(\Psi_{v,B} + \Psi_\sigma) = 0, \quad (197)$$

$$\bar{a}^{-1}\dot{\Psi}_{v,B} = \nu + n_e\sigma_T\frac{4}{3}\frac{\rho_\gamma}{\rho_B}\bar{a}^{-1}\Psi_{v,\gamma B}, \quad (198)$$

$$\text{relative velocity potential: } \Psi_{v,\gamma B} \equiv \Psi_{v,\gamma} - \Psi_{v,B}. \quad (199)$$

In eq. (198), the baryon pressure is neglected, a valid approximation for primary anisotropy calculations: it manifests itself through the post-recombination baryon Jeans length, $k^{-1} \sim 1h^{-1}\text{kpc}$, very small compared to the $\sim 5h^{-1}\text{Mpc}$ damping scale for primary anisotropies [134]. For Thomson scattering with Δ_t independent of q , the first few moments of the photon distribution function are related to the density and pressure perturbations, velocity potential and anisotropic stress by

$$\Delta_{t0}^{(S)} = \frac{\delta_\gamma}{4} = \frac{(\delta p)_\gamma}{4\bar{p}_\gamma}, \quad \Delta_{t1}^{(S)} = \frac{k\Psi_{v,\gamma}}{3\bar{a}}, \quad \Delta_{t2}^{(S)} = k^2\frac{\pi_{t,\gamma}}{12},$$

and the first two moments of the $\Delta_t^{(S)}$ Boltzmann transport equation are just the energy and momentum conservation equations for photons:

$$\text{photons: } \frac{1}{4}\dot{\delta}_\gamma + \dot{\varphi} + \frac{1}{3}k^2\bar{a}^{-1}(\Psi_{v,\gamma} + \Psi_\sigma) = 0, \quad (200)$$

$$\bar{a}^{-1}\dot{\Psi}_{v,\gamma} - \bar{H}\Psi_{v,\gamma} - (\frac{1}{4}\delta_\gamma + \nu) + \frac{1}{6}k^2\pi_{t,\gamma} = -n_e\sigma_T\Psi_{v,\gamma B}, \quad (201)$$

The *right-hand side* of eq. (198) gives the body force (i.e., per unit volume) from Compton drag felt by the baryons, and the *right-hand side* of eq. (201) gives the equal and opposite body force felt by the photons:

$$\begin{aligned} \{\text{Force}_B\}_{C\text{-drag}} &= -\frac{1}{\bar{a}^4}2\int\frac{d^3\mathbf{q}}{(2\pi)^3}\mathcal{S}_t\mathbf{q} = -(\rho_\gamma + p_\gamma)n_e\sigma_T(\mathbf{v}_B - \mathbf{v}_\gamma), \\ \{\text{Force}_\gamma\}_{C\text{-drag}} &= -\{\text{Force}_B\}_{C\text{-drag}}. \end{aligned} \quad (202)$$

Compton drag effectively damps the gas motion down to $z \sim 300$ for CDM-type models if the universe remains ionized, but lets up at z below ~ 1000 for normal recombination.

There are two kinds of limiting behavior in the baryon plus photon transport equations that simplify calculations. The first is at the beginning of computations, when the photons plus baryons are tightly coupled, acting like a single fluid, albeit with a shear viscosity and a thermal conductivity (section C.3.1).

In this limit, the infinite hierarchy of moment equations is not needed, but is truncated by assuming the $\ell = 3$ term vanishes. The anisotropic stress $\pi_{t,\gamma}$ is then related to $\Psi_{v,\gamma B}$, which also has a thermal diffusion contribution to it. To get these transport coefficients accurately, $\Psi_{v,\gamma B}$ must be expanded to second order in τ_C . These are used until it is unsafe to do so (using a conservative safety tolerance).

The full transport equations are then solved, with an algorithm for opening up the number of multipoles being calculated: the main effect of transport is to propagate a pulse in ℓ -space localized around $\ell = k\tau$ from low ℓ to high ℓ as τ increases to τ_0 . In fig. 14, the nature of the pulse (smoothed in ℓ , as described below) at its final location at τ_0 is shown for representative ℓ 's.⁵

A second regime, used long after photon decoupling and only if the curvature term and $\tilde{G}_{tSW}^{(S)}$ are negligible, is free-streaming in the $\tilde{\Delta}_t$ variable. It can be used to propagate from some stopping point $\tau_s(k)$ to the present in one step. This translates the ℓ -space pulse from a location centered on $\ell \sim k\tau_s$ to one centered on $\ell \sim k\tau_0$.

The hypersurface choice should be the one it is easiest to make the computations in. What often happens is that the equations suggest best variables for numerical reasons which pick out the gauge choice. The perturbed source function was derived in the comoving gauge of the baryons, then transformed to other gauges. Even with evolution in the synchronous gauge at the beginning of a CDM-dominated evolution, after photon decoupling when the radiation free-streams, a combination of the photon distribution function and metric variables is suggested which turns out to be the Tolman combination $e^{\nu_L}T_\gamma$ in the longitudinal gauge. Quantities which are manifestly invariant under infinitesimal coordinate transformations are also usually numerically preferred, such as the photon entropy per baryon and differences in velocities (the most obvious of which is velocity relative to CDM). A crucial one for numerical accuracy is the relative velocity of the photons and baryons.

In figs. 19 and 20, a few scalar perturbation mode functions are shown: fig. 19 shows relative density perturbations as computed in the synchronous gauge, while fig. 20 shows some gauge invariant velocity potentials $\Psi_{v,\gamma B}$, $\Psi_{v,B\text{ cdm}}$ and the metric variables $\dot{\varphi}$ and $\Psi_{\sigma,S}$. The behavior of the relative density perturbations in the longitudinal gauge outside the horizon is dramatically differ-

⁵The continuum limit of the moment hierarchy in ℓ – Appendix C.2.1, eqs. (333), (334) – gives a wave equation for $D(k\tau, Q) = \Delta_{t\ell}^{(S)}$ in the variables $k\tau$ and $Q = \ell + 1/2$ with the wave solution $Q^{-1/2}\text{fn}(Q - k\tau)$ which conserves the power $\sum (2\ell + 1)|\Delta_{t\ell}^{(S)}|^2, \sim \int \text{fn}^2(q) dq$. Here fn is a function describing the pulse whose form is determined by the power that is injected at the base of the hierarchy, through the sources acting on $\ell = 0, 1, 2$. The $Q^{-1/2}$ prefactor is a little more complicated with curvature, but the pulse description is still valid, $D \propto \text{fn}(kd_{\text{curv}} \arcsin(Q/kd_{\text{curv}}) - k\tau)$ for closed models, $D \propto \text{fn}(kd_{\text{curv}} \text{arcsinh}(Q/kd_{\text{curv}}) - k\tau)$ for open models. Thus it is the combination $Q/(k\mathcal{R}(\chi))$ which is relevant for propagation of the ℓ -space pulse, where $\mathcal{R}(\chi)$ is defined by eq. (130).

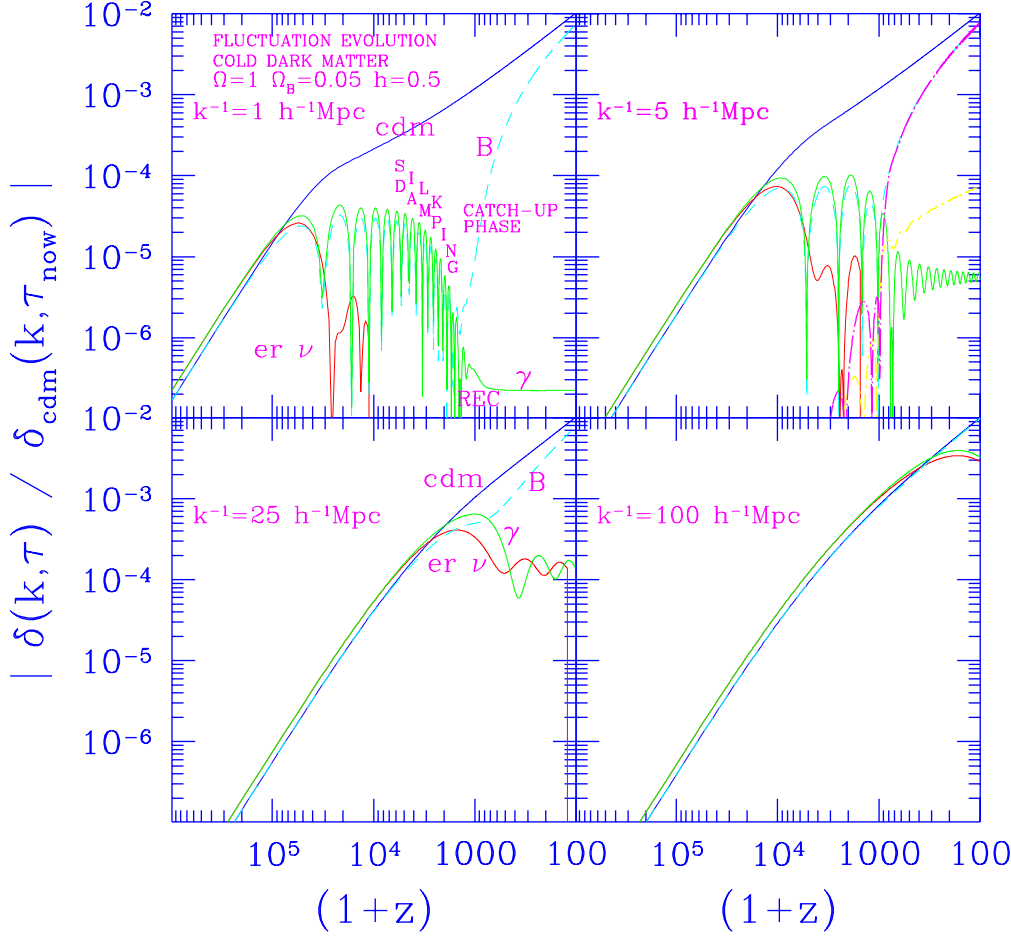


Figure 19: The synchronous gauge evolution of scalar perturbations with the 4 wavenumbers shown for the standard CDM model with normal recombination illustrates such basic physical phenomena as Hubble drag on the CDM perturbation growth in the *er*-dominated regime after the wave “enters the horizon”, Silk damping of the baryon and photon perturbations, the catch-up of the baryons to the CDM after photon decoupling. $k = 1 \text{ h}^{-1} \text{ Mpc}$ is about the highest k one needs to go to get an accurate computation of $\mathcal{C}_\ell$ for this model. Even so, by $z = 100$ one needs to follow multipoles up to $\ell = 460$, and the number of photon ODEs is twice this because of the polarization. After free-streaming to $z = 0$, one needs to go to ℓ about 6000. Although this is easiest to do with the one step free-streaming method, it is also quite feasible to do the full Boltzmann equation integration numerically. For relativistic neutrinos, modes only up to $\ell = 40$ were included, but once they exert a negligible effect on the metric variables they are shut off.

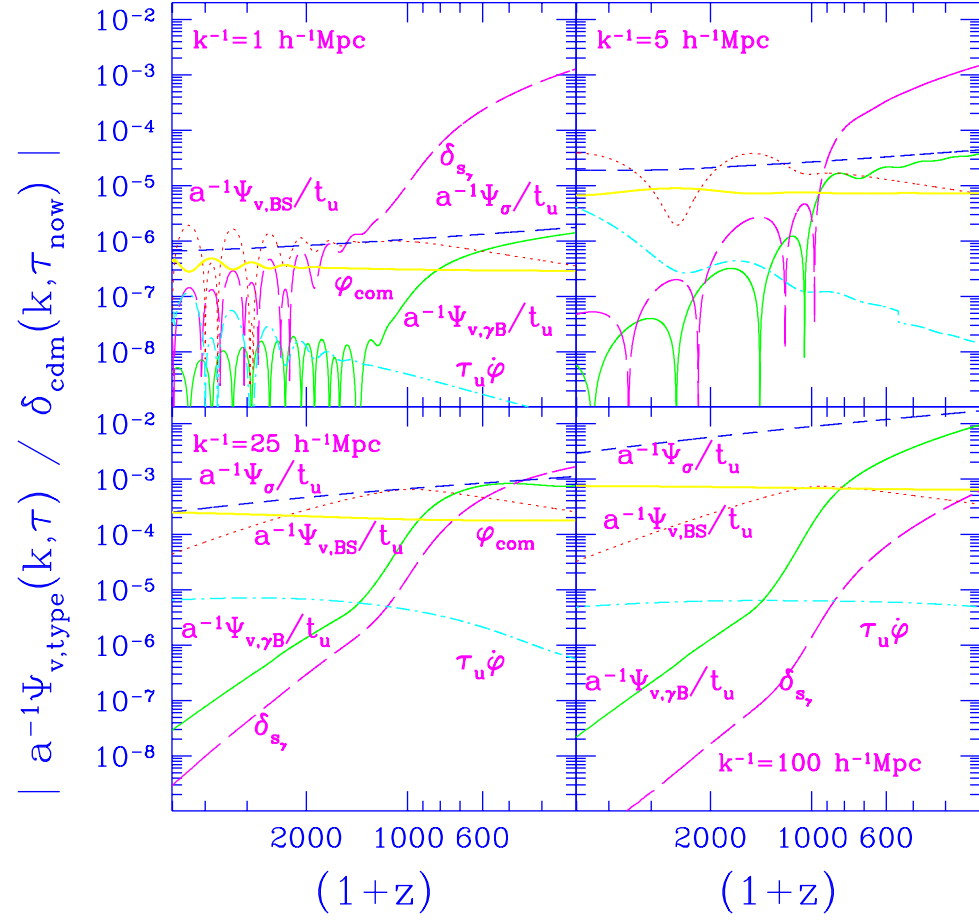


Figure 20: The gauge invariant relative velocity potentials $\bar{a}^{-1}\Psi_{v,\gamma B}$, $\bar{a}^{-1}\Psi_{v,B\,cdm} \equiv \Psi_{v,BS}$ and the photon entropy per baryon perturbations are shown for the standard CDM model with normal recombination for the 4 wavenumbers of the last figure. They are all normalized to the amplitude of the CDM density fluctuation at the current time if linear growth prevailed. The synchronous gauge metric variables $\bar{a}^{-1}\Psi_{\sigma S} \equiv \bar{a}^{-1}\Psi_{v,cdm,L}$ and $\dot{\varphi}_S$ and the comoving curvature parameter φ_{com} are also shown: $\dot{\varphi}_S$ becomes negligible and $\bar{a}^{-1}\Psi_{\sigma S} \propto \tau$ at late times. The velocity potentials and $\dot{\varphi}$ are in units used for the Boltzmann integration code, so the relative magnitudes are meaningful. (The physical and conformal time units of the code are $ct_u = 22.28$ Mpc, $\tau_u = 1280$ yr, $t_u/\tau_u = (a_0)_u = 56776$.)

ent (see eq. (261)):

$$\frac{\delta_{\text{type},L}}{3(1+\bar{p}/\bar{\rho})_{\text{type}}} = \frac{\delta_{\text{type},S}}{3(1+\bar{p}/\bar{\rho})_{\text{type}}} - \bar{H}\Psi_{\sigma,S}. \quad (203)$$

For $\tau \ll \tau_{eq}$ and $k\tau \ll 1$, $\bar{H}\Psi_{\sigma,S}$ is approximately constant, hence so are the relative perturbations. At late times with $k\tau \gg 1$, the $\bar{H}\Psi_{\sigma,S}$ is dominated by $\delta_{\text{type},S}$, so $\delta_{\text{type},L} \approx \delta_{\text{type},S}$. That is why one can compute transfer functions for density perturbations in either gauge without hypersurface shifting. The quantities $\delta_{s_\gamma} = (1+\bar{p}_\gamma/\bar{\rho}_\gamma)^{-1}\delta_\gamma - \delta_B$ (fig. 20) and $(1+\bar{p}/\bar{\rho})_{\text{type}}^{-1}\delta_{\text{type}} - \delta_{cdm}$ are gauge invariant of course. The latter are useful for accurately following scalar isocurvature CDM models (for small k).

A catalogue of mode functions with varying k are generated. Depending upon the accuracy one wishes anywhere from many hundreds to many thousands are typical for a CDM calculation. The output of the Einstein–Boltzmann calculations is therefore $\Delta_{\{t,Q\},\ell}^{(S)}(k, \tau_0)$ – even in open or closed FRW models, where $k^2/\bar{a}^2$ is the eigenvalue of $-(^{(3)}\nabla)^2$. This allows one to form the k -space spectra for given ℓ , $d\mathcal{C}_{\{t,Q\},\ell}^{(S)}/d\ln k$, which, when integrated over $\ln k$, yield the spectra $\mathcal{C}_{\{t,Q\},\ell}^{(S)}$. Figure 14 showed the standard CDM example for $\ell = 4, 10, 59, 121$. The $\ell = 59, 121$ cases have been averaged over nearby ℓ 's, from $\ell - \delta\ell$ to $\ell + \delta\ell$, to smooth out the dominant rapid oscillation associated with the typical $j_\ell^2(k\tau_0)$ behavior. If one has sparse k coverage, just a few hundred logarithmically spaced from $(10^{-7} \text{ h}^{-1} \text{ Mpc})^{-1}$ to $(1 \text{ h}^{-1} \text{ Mpc})^{-1}$, then $\delta\ell$ should not be too small. With many thousand, little smoothing is needed. Another approach to smoothing is to wait until the $\ln k$ integration has been done. Too much smoothing lowers the heights of the Doppler peaks, too little leaves high frequency oscillations in $\mathcal{C}_\ell$.

Figure 21(a) shows the differential spectrum $d\mathcal{C}_{t,\ell}^{(S)}/d\ln k$ for the quadrupole and a window, corresponding to typical half-degree-beam anisotropy experiments, that probes the same k -band as many large scale structure observations. The no-recombination CDM model has very little power at $\ell = 214$ as expected. The nonzero Λ and standard CDM models look similar except for a shift to smaller k for the nonzero Λ model associated with τ_0 being larger. Notice that the quadrupole probes k 's whose wavelength exceeds the size of our Hubble patch, although unless the power spectrum is rising rapidly to small k , waves with $k\tau_0 < 1$ contribute very little to the observed quadrupole, and even less to the octopole and higher multipoles. Still it is this behavior which allows one to set useful constraints on “fluctuations bigger than the horizon”.

Figure 22 shows where the polarization power, $\mathcal{C}_{Q,\ell}^{(S)}$ lies in ℓ -space for scalar modes when there is standard recombination and early reionization. Figure 21(b) shows $d\mathcal{C}_{Q,\ell}^{(S)}/d\ln k$, i.e., where the polarization power lies in k -space, for the ℓ choices of fig. 14. The polarization is a 10% effect in $\Delta T/T$ [134]. The polarization power spectrum can be used, for example, to make theoretical polarization maps, [88] and Appendix C.1. [166] has shown that

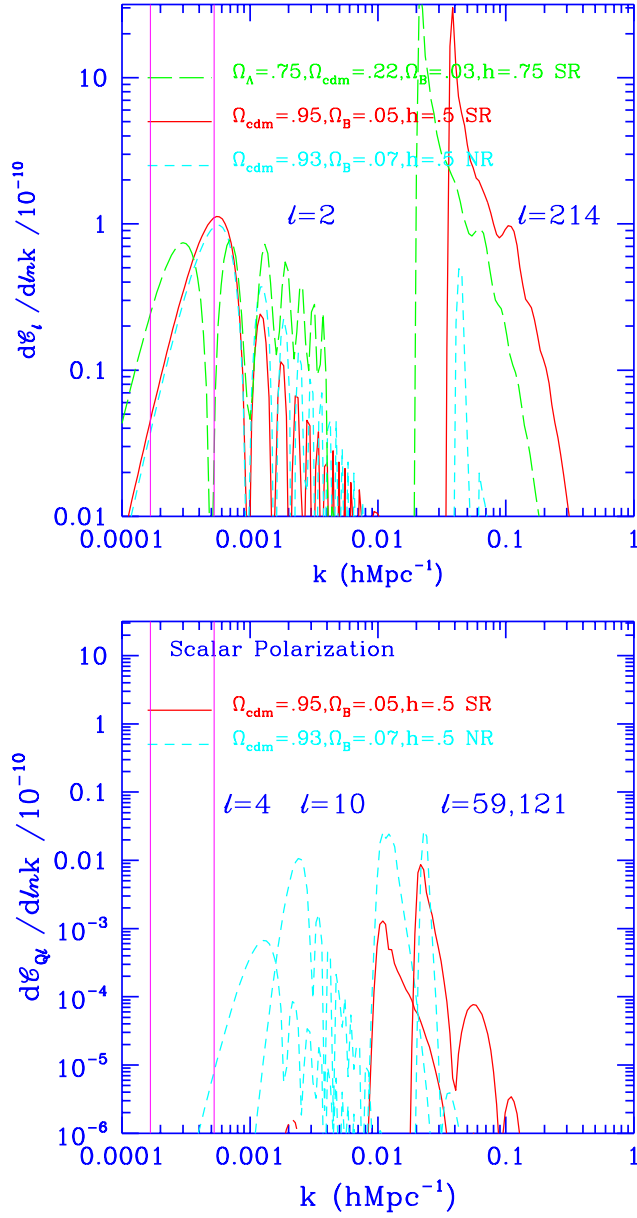


Figure 21: (a) $d\mathcal{C}_\ell/d\ln k$ for the scale invariant models listed, for the quadrupole and a $\ell \sim 200$ multipole that lies within the MAX and MSAM windows, which probes k extending into the large scale structure region. The vertical lines are defined by $k^{-1} = 2cH_0^{-1}$ and $\pi k^{-1} = 2cH_0^{-1}$, when half a wavelength equals the horizon size. The extension beyond this line is what one means when one says that CMB data can constrain fluctuations bigger than our horizon: a huge increase would be ruled out by the quadrupole observations. (b) How the polarization power is concentrated in k for selected multipoles. The contribution at low ℓ for the SR model is negligible.

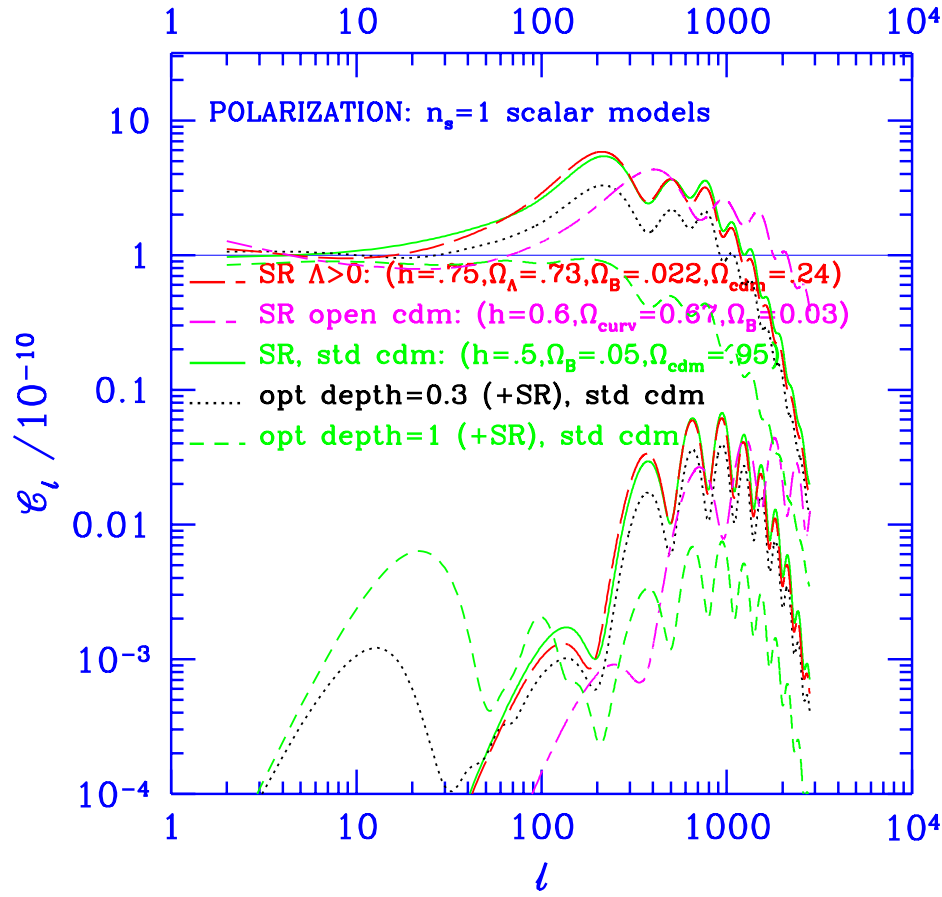


Figure 22: Polarization power spectra for the models shown demonstrate that over a limited multipole band the polarization power has signals about 10% of the primary signal. As experimental noise decreases, it can provide a signature for early reionization models.

there is a small but interesting cross-correlation between polarization and total anisotropy maps which may be useful in differentiating among polarization components. Given the strides in decreasing receiver noise, it seems quite plausible that the 10% effect (on selected angular scales) can be used to differentiate among models, in particular it could provide a nice signature for early reionization models since the polarization power is concentrated at relatively low ℓ 's, around $\ell \sim 10 - 50$, whereas it is a small angle signature with normal recombination. Of course, the presence or absence of a Doppler peak in $\mathcal{C}_{t,\ell}^{(S)}$ (upper curves) is a more direct signature, but the more signals we have to select on primary anisotropies the better.

As we have seen, after Compton scattering has become negligible as a source, at say $\tau_s(k)$, the solution to the radiative transfer problem for the Tolman combination $\tilde{\Delta}_t^{(S)}$ is

$$\begin{aligned} \tilde{\Delta}_t^{(S)}(\hat{q}, \mathbf{k}, \tau_0) &= e^{-\zeta_C(\tau_0|\tau_s(k))} e^{-i\hat{k}\cdot\hat{q}k(\tau_0-\tau_s(k))} \tilde{\Delta}_t^{(S)}(\hat{q}, \mathbf{k}, \tau_s(k)) \\ &+ \int_{\tau_s(k)}^{\tau_0} d\tau e^{-\zeta_C(\tau_0|\tau)} e^{-i\hat{k}\cdot\hat{q}k(\tau_0-\tau)} \left(\dot{\nu} - \dot{\varphi} + \frac{\partial^2 \bar{a}^{-1} \Psi_\sigma}{\partial \tau^2} \right). \end{aligned} \quad (204)$$

The first term represents the free-streaming of the temperature pattern at τ_s to τ_0 . The second term involving a line-of-sight integral of $(\dot{\nu} - \dot{\varphi} + (\partial^2 \bar{a}^{-1} \Psi_\sigma / \partial \tau^2))$ is the integrated Sachs–Wolfe effect. This term vanishes for standard $\Omega_{nr} = \Omega = 1$ universes, provided we take $\tau_s \gg \tau_{dec}, \tau_{eq}$. The classical $\Phi_N/3$ Sachs–Wolfe factor (where Φ_N is the Newtonian gravitational potential perturbation) is easiest to see in the synchronous gauge: $\nu_S = 0$ defines the gauge, $\dot{\varphi}_S \rightarrow 0$ and $\bar{a}^{-1} \Psi_\sigma \rightarrow \frac{1}{3} \Phi_N \tau$ for $\tau_s \gg \tau_{eq}$, hence the $\partial_\tau [\bar{a}^{-1} \Psi_\sigma]$ term in $\tilde{\Delta}_t^{(S)}$ gives $\Phi_N/3$. Of course, photon bunching and Doppler effects also have small influences even at low ℓ . Figure 23 contrasts the shape of the spectrum that would result if only $\Phi_N/3$ contributed with the exact result. The effective tilt in $\mathcal{C}_\ell^{(S)}$ for this $n_s = 1$ case is $\nu_{\Delta T} = 0.15$, not 0 as the $\Phi_N/3$ approximation would give.

The integrated Sachs–Wolfe effect is important if we try to take $\tau_s(k)$ too close to τ_{eq} or if the gravitational potential changes as a result of a change in the equation of state of the Universe, for example the period between τ_Λ , when vacuum energy becomes important, and τ_0 . For an initially scale invariant spectrum it causes an upturn in $\mathcal{C}_\ell^{(S)}$ at low ℓ , as shown in fig. 23.

Not only do open models have a nontrivial integrated Sachs–Wolfe effect, there is also a direct effect of the curvature on the mode function evolution, as is described in section C.4. Of course whatever mechanism generated the ultra-large-scale mean curvature may well have had associated with it strong fluctuations on observable scales, so much so that this is an argument against large mean curvature because of the absence of such effects in the CMB. Even if the background curvature is determined by an entirely different mechanism, it should influence the fluctuation generation mechanism. An open issue in open models has always been what is a natural shape for the spectrum for k near d_{curv}^{-1} . Power laws in kd_{curv} , $\sqrt{(kd_{curv})^2 - 1}$, etc. have often been adopted.

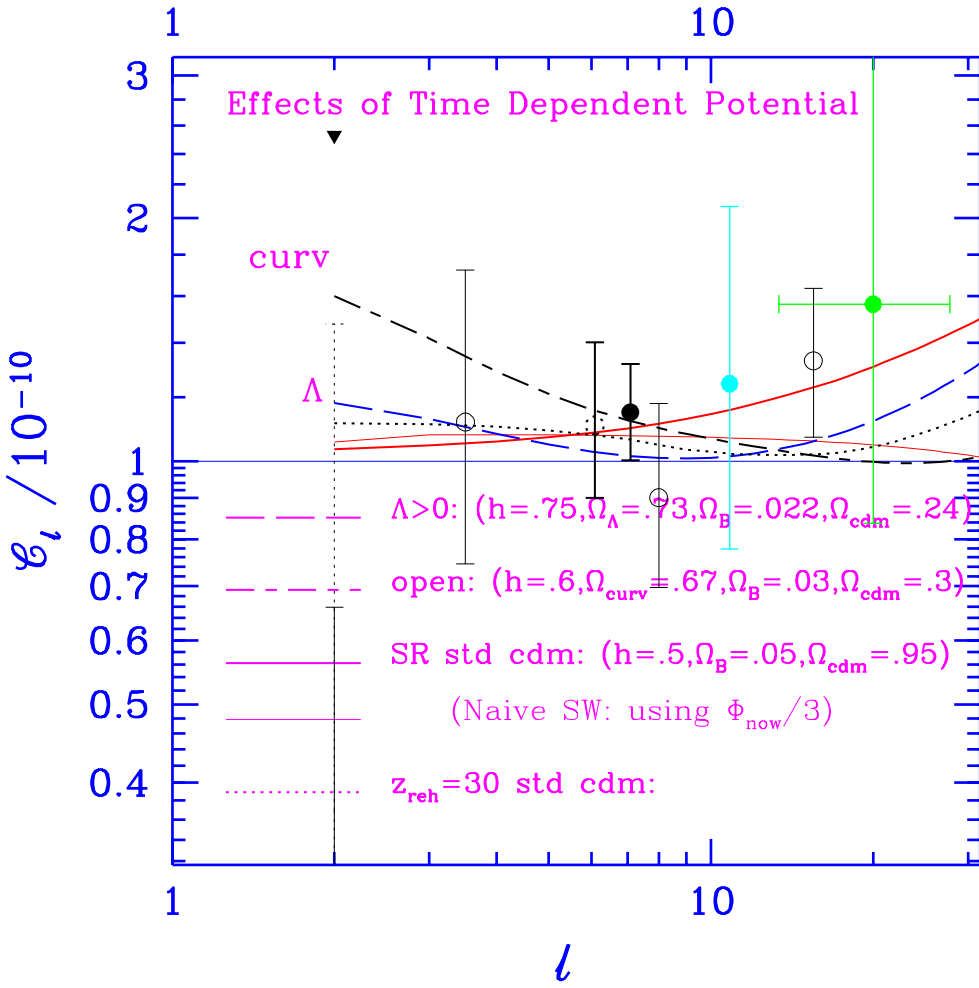


Figure 23: This illustrates the role of the integrated Sachs–Wolfe effect for scalar perturbations when there is vacuum energy or when there is negative curvature. The vacuum effect was first considered by Kofman and Starobinski (1985). The $h = 0.6$ open model also has an integrated Sachs–Wolfe effect, but uses a scale-invariant initial condition from inflation which is naturally truncated at $kd_{\text{curv}} = 1$; by contrast, a $h = 0.55, \Omega = 0.6$ model with the same 13 Gyr age turns down at low ℓ . Also shown is a comparison between the $\mathcal{C}_\ell$ from the full Einstein–Boltzmann transport and the $\mathcal{C}_\ell$ found using just 1/3 of the current gravitational potential.

The case shown in fig. 23 has equal power per decade in the initial gravitational potential power spectrum (or more correctly in $\mathcal{P}_{\varphi_{com}}$), that would arise if the fluctuations were generated by quantum oscillations during an inflation epoch subsequent to the mean curvature generation, for tilt $\nu_s \approx 0$ [245, 244, 305].

Spurred on by the promise of high precision space experiments [152, 154], a considerable fraction of the CMB theoretical community with Boltzmann transport codes compared their approaches and validated the results to ensure subpercent accuracy [301]. An important byproduct of this was an emphasis on speed, since one hopes to constrain a large multidimensional parameter space with the anisotropy data. The important issues for methods based on solving the moment equations are discussed in various places in these notes; although the techniques were in place prior to the COBE discovery, to get the high accuracy with speed has been somewhat of a challenge: e.g., if the number of wavenumbers run is too small then smoothing is required, and this smooths the $\mathcal{C}_\ell$ curves, but to run the number needed to avoid smoothing is slow.

Although solving the hierarchy of moment equations became the standard approach for evaluating the transport of photons and neutrinos, there are alternatives. One can cast the entire problem of photon transport in terms of integral equations in which the multipoles with $\ell > 2$ are expressed as history-integrals of metric variables, photon-bunching ($\Delta_{t,0}^{(S)}$), Doppler and polarization (e.g., $\Delta_{t,2}^{(S)}$) sources; and the problem of neutrino transport, massive and massless, can be cast into history-integrals of metric variables only. This approach was used by [195, 134, 233, 261] for hot and warm dark matter to evaluate moments that fed into the metric equations (eq. 350). It was used by Kaiser [266] to evaluate photon polarization. If applied to just the integrated Sachs-Wolfe term it can augment the one-step free-streaming result and allow one to begin the free-streaming transport to the present shortly after recombination without any loss of accuracy. It has now been used by Seljak and Zaldarriaga [306] to develop an accurate and fast code for $\mathcal{C}_\ell$ evaluation. One aspect of the speedup is that since $\mathcal{C}_\ell$'s do not change that rapidly with ℓ , one does not need to evaluate the history-integrals ℓ by ℓ , whereas with the moment hierarchy the Δ_ℓ 's are all coupled to each other.

6.5.2 Tensor mode transfer equations

For tensor perturbations, and for flat universes, we have seen that for wavenumber $\mathbf{k}$ there are two independent tensor modes defined in terms of two transverse traceless matrices, $E_{ij}^{(T\{+, \times\})}$ (satisfying $k^j E_{ij}^{(T\{+, \times\})} = 0, E_{ij}^{(T\{+, \times\})} \delta^{ij} = 0$). The expansion of $h_{ij}^{(TT)}$ and the anisotropic stresses $\Pi_{type\,ij}$ in this basis gives the $h_{(T\{+, \times\})}$ and $\pi_{type}^{(T\epsilon)}$ mode functions, and the reduction of the $(G')_j^i$ dynamical Einstein equation to eq. (171), $\ddot{h}_{(T\epsilon)} + 2\bar{H}\bar{a}\dot{h}_{(T\epsilon)} + k^2 h_{(T\epsilon)} = 16\pi G_N \bar{a}^2 \bar{p}_{tot} \pi_{tot}^{(T\epsilon)}$ (for the flat case).

The radiation field can also be expanded in these modes. The natural mode

variables are $\tilde{\Delta}_{\{t,U,V,Q\}}^{(T\epsilon)}$ in the expansion

$$\Delta_{ij}^{(T)} = w \sum_{(\mu)=t,Q,U,V} \sum_{\epsilon=+,\times} \sum_{\mathbf{k}} \tilde{\Delta}_{(\mu)}^{(T\epsilon)} \frac{E^{(T\epsilon)} \cdot \mathcal{E}_{(\mu)}}{\mathcal{E}_{(\mu)} \cdot \mathcal{E}_{(\mu)}} E_{ij}^{(T\epsilon)} e^{i\mathbf{k} \cdot \mathbf{x}} a_{\mathbf{k}(T\epsilon)} + \text{cc.} \quad (205)$$

The $\mathcal{E}_{(\mu)}$ are the tensor product combinations of the polarization basis ε_s , eq. (185) and Appendix C, eq. (309). In the frame in which $\hat{k}$ is the pole and $\hat{q}$ has polar coordinates (θ, ϕ) , the $E^{(T\epsilon)} \cdot \mathcal{E}_{(\mu)}$ terms are proportional to either $\cos(2\phi)$ or $\sin(2\phi)$ and functions of $\mu = \cos(\theta)$ that are at most quadratic, and are given by eq. (372) of Appendix C.6. Apart from an overall sign, these are the combinations first suggested by Polnarev [230] and which we used in [140]:

$$\begin{aligned} \Delta_t^{(T)} &= -\tilde{\Delta}_t^{(T+)}(1 - \mu^2) \cos 2\phi - \tilde{\Delta}_t^{(T\times)}(1 - \mu^2) \sin 2\phi \\ \Delta_Q^{(T)} &= \tilde{\Delta}_Q^{(T+)}(1 + \mu^2) \cos 2\phi + \tilde{\Delta}_Q^{(T\times)}(1 + \mu^2) \sin 2\phi \\ \Delta_U^{(T)} &= \tilde{\Delta}_U^{(T+)}2\mu \sin 2\phi - \tilde{\Delta}_U^{(T\times)}2\mu \cos 2\phi. \end{aligned} \quad (206)$$

The source functions in these modes, $\tilde{\mathcal{G}}_{\{t,U,V,Q\}SW}^{(T\{+,\times\})}$, are also derived in Appendix C.6:

$$\tilde{\mathcal{G}}_{tSW}^{(T\epsilon)} = \frac{1}{2}\dot{h}_{(T\epsilon)}, \quad \tilde{\mathcal{G}}_{\{U,V,Q\}SW}^{(T\epsilon)} = 0, \quad (207)$$

$$\tilde{\mathcal{G}}_{tC}^{(T\epsilon)} = -\tau_C^{-1}(\tilde{\Delta}_t^{(T\epsilon)} - \Upsilon_{(T\epsilon)}), \quad \tilde{\mathcal{G}}_{VC}^{(T\epsilon)} = -\tau_C^{-1}\tilde{\Delta}_V^{(T\epsilon)}, \quad (208)$$

$$\tilde{\mathcal{G}}_{QC}^{(T\epsilon)} = -\tau_C^{-1}(\tilde{\Delta}_Q^{(T\epsilon)} + \Upsilon_{(T\epsilon)}), \quad \tilde{\mathcal{G}}_{UC}^{(T\epsilon)} = -\tau_C^{-1}(\tilde{\Delta}_U^{(T\epsilon)} - \Upsilon_{(T\epsilon)}),$$

$$\Upsilon_{(T\epsilon)} \equiv$$

$$\begin{aligned} &\frac{3}{8} \int \frac{1}{2} d\mu' [\frac{1}{2}(1 - (\mu')^2)^2 \tilde{\Delta}_t^{(T\epsilon)} - \frac{1}{2}(1 + (\mu')^2)^2 \tilde{\Delta}_Q^{(T\epsilon)} + \frac{1}{2}(2\mu')^2 \tilde{\Delta}_U^{(T\epsilon)}] \\ &= \frac{1}{10} \tilde{\Delta}_{t0}^{(T\epsilon)} + \frac{1}{7} \tilde{\Delta}_{t2}^{(T\epsilon)} + \frac{3}{70} \tilde{\Delta}_{t4}^{(T\epsilon)} + \frac{3}{5} \tilde{\Delta}_{Q0}^{(T\epsilon)} - \frac{6}{7} \tilde{\Delta}_{Q2}^{(T\epsilon)} + \frac{3}{70} \tilde{\Delta}_{Q4}^{(T\epsilon)}. \end{aligned}$$

Recall from section 6.5.1 that in the scalar case only $\Delta_t^{(S)}$ and $\Delta_Q^{(S)}$ can be excited, so two transfer equations are required [134]. In the tensor case, $\tilde{\Delta}_t^{(T\epsilon)}$, $\tilde{\Delta}_Q^{(T\epsilon)}$ and $\tilde{\Delta}_U^{(T\epsilon)}$ can be excited, but the source for $\tilde{\Delta}_Q^{(T\epsilon)} + \tilde{\Delta}_U^{(T\epsilon)}$ has only a pure damping term, so the combination will be unexcited in the early universe and will remain so – as will $\tilde{\Delta}_V^{(T\epsilon)}$. Thus, the four Stokes radiative transfer equations again reduce to two.

The back action on the gravity wave collisionless damping is from the anisotropic stress for the photons,

$$\pi_\gamma^{(T\epsilon)} = 12(\frac{2}{15} \tilde{\Delta}_{t0}^{(T\epsilon)} + \frac{4}{21} \tilde{\Delta}_{t2}^{(T\epsilon)} + \frac{2}{35} \tilde{\Delta}_{t4}^{(T\epsilon)}), \quad (209)$$

with a similar contribution from extremely relativistic neutrinos. There will also be a contribution from hot or warm dark matter in the *er* and semi-relativistic phase.

The main features of the solution can be readily understood by writing the transport equations in terms of a combination that only has $\dot{h}_{(T\epsilon)}$ as a source:

$$\begin{aligned}(\partial_\tau + ik\mu + \tau_C^{-1})[\tilde{\Delta}_t^{(T\epsilon)} - \tilde{\Delta}_Q^{(T\epsilon)}] &= \frac{1}{2}\dot{h}_{(T\epsilon)}, \\(\partial_\tau + ik\mu + \tau_C^{-1})[\tilde{\Delta}_Q^{(T\epsilon)}] &= \tau_C^{-1}\Upsilon_{T\epsilon},\end{aligned}\quad (210)$$

As in the scalar case, these equations are solved by expanding in Legendre polynomials [140]. The polarization induced by the tensor mode is quite small ($\ll 1\%$) [141]. To the extent that polarization and the small back action of the anisotropic stress of the neutrinos and photons upon the gravity waves can be neglected, the solution (for the flat background case) is simply

$$\tilde{\Delta}_{t\ell}^{(T\epsilon)} \approx \int_0^{\tau_0} e^{-\zeta_C(\tau)} d\tau j_\ell(k\chi) \frac{1}{2}\dot{h}_{(T\epsilon)}(\tau), \quad \chi \equiv \tau_0 - \tau. \quad (211)$$

The $e^{-\zeta_C(\tau)}$ implies that waves that entered the horizon before decoupling will not be able to develop anisotropy in $\tilde{\Delta}_t^{(T\epsilon)}$ until after recombination, when the gravity waves will have already decayed as a result of collisionless dispersion, as embodied in the spherical Bessel function behavior of $h_{(T\epsilon)}$.

To go from the mode variable $\tilde{\Delta}_{(\mu)}^{(T\epsilon)}$ to angular power spectra, one must take into account the angular dependence of $E^{(T\epsilon)} \cdot \mathcal{E}_{(\mu)}$. For $\tilde{\Delta}_t^{(T\epsilon)}$, the angular power spectrum found by summing over $\mathbf{k}$ and polarizations is

$$\begin{aligned}\frac{d\mathcal{C}_{t\ell}^{(T)}}{d\ln k} &= \ell(\ell+1) \left(1 - \frac{1}{\ell^2}\right) \left(1 + \frac{2}{\ell}\right) \frac{k^3}{2\pi^2} \frac{1}{2} \sum_{\epsilon=+, \times} \left\langle \left| \frac{\tilde{\Delta}_{t,\ell-2}^{(T\epsilon)}}{(1 - \frac{1}{2\ell})(1 + \frac{1}{2\ell})} \right. \right. \\ &\quad \left. \left. + 2 \frac{\tilde{\Delta}_{t,\ell}^{(T\epsilon)}}{(1 - \frac{1}{2\ell})(1 + \frac{3}{2\ell})} + \frac{\tilde{\Delta}_{t,\ell+2}^{(T\epsilon)}}{(1 + \frac{1}{2\ell})(1 + \frac{3}{2\ell})} \right|^2 \right\rangle.\end{aligned}\quad (212)$$

If we assume recombination is sudden at $\tau = \tau_{dec}$ and use the eq. (211) approximation, this reduces to the classical Abbott and Wise [225] formula for $\mathcal{C}_{t\ell}^{(T)}$. With the full $e^{-\zeta_C(\tau)}$ included, the approximation is a good one compared with the results of the full integration. Typical solutions are shown in figs. 4.6 and 7. The decline at ℓ above ~ 50 is due to the loss of gravity wave power by decoupling. Note also the rise at $\ell = 2$. Even though eq. (211) is simple in form, the directional decay of $h_{(T\epsilon)}$ implies eq. (212) even for $\mathcal{C}_2^{(T)}$, hence the ratio $r_{ts} = \mathcal{C}_2^{(T)}/\mathcal{C}_2^{(S)}$, needs numerical evaluation. The feedback of the anisotropic stress in relativistic neutrinos and photons upon the gravitational wave evolution does have a significant ($\sim 20\%$) effect at $\ell \gtrsim 100$, which is somewhat larger for smaller Ω_{nr}/Ω_{er} , but by then the power has fallen off sufficiently that it is unmeasurable. In fact, in fig. 7 the $\mathcal{C}_\ell^{(T)}$ actually have both curves, with and without anisotropic stress feedback, drawn, and on this scale one cannot see a difference. For open universes, the influence of the gravitational radiation on the spectrum extends to higher ℓ because of the angle-distance relation [305].

7 Connection with other cosmic probes of k-space

7.1 Density power spectra and characteristic scales

A byproduct of the linear perturbation calculations used to compute $\Delta T/T$ is the transfer function for density fluctuations, which maps the initial density fluctuation spectrum in the very early universe into the final post-recombination one. From this, fluctuation spectra appropriate to the linear regime for the density, velocity and gravitational potential can be constructed. Various (co-moving) wavenumber scales determined by the transport of the many species of particles present in the universe characterize these spectra. The most important of these for dark matter dominated universes is the scale of the horizon at redshift Ω_{nr}/Ω_{er} when the density in nonrelativistic matter, $\Omega_{nr}\bar{a}^{-3}$, equals that in relativistic matter, $\Omega_{er}\bar{a}^{-4}$,

$$k_{Heq}^{-1} = 5 \Gamma_{eq}^{-1} h^{-1} \text{ Mpc}, \quad \Gamma_{eq} = \Omega_{nr} h [\Omega_{er}/(1.68\Omega_\gamma)]^{-1/2}. \quad (213)$$

(It is defined by $k\tau_{eq} = \pi$, where τ_{eq} is the conformal time at er/nr equality.) In [231], we adopted the functional form:

$$\begin{aligned} \mathcal{P}_\rho(k) &\propto k^{3+n_s(k)} \{1 + [ak + (bk)^{3/2} + (ck)^2]^p\}^{-2/p}, \\ (a, b, c) &= (6.4, 3.0, 1.7) \Gamma^{-1} h^{-1} \text{ Mpc}, \quad p = 1.13, \\ \Gamma &\approx \Gamma_{eq} e^{-(\Omega_B(1+\Omega_{nr}^{-1}(2h)^{1/2})-0.06)}, \end{aligned} \quad (214)$$

where Γ is an effective index. For $\Gamma = 0.5$, this accurately fits the linear power-spectrum of the standard adiabatic CDM model with $\Omega_{nr} = 1$, $h = 0.5$ and $\Omega_B = 0.03$ [134]. Although one does not expect that this fit will be highly accurate if we change Ω_B , and indeed the best-fit parameters a, b, c, ν do vary with Ω_B [134, 88], it is usually sufficiently accurate for large scale structure work to use a simple $\exp[-2(\Omega_B - 0.03)]$ correction factor for modest Ω_B variations, even if h varies [250]; a further improvement [251] occurs for low Ω_{nr} if the 2 is replaced by $(1 + \Omega_{nr}^{-1}(2h)^{1/2})$, as indicated. (The oft-used $\Omega_B \rightarrow 0$ form given in BBKS [232], Appendix G,

$$\begin{aligned} \mathcal{P}_\rho(k) &\propto k^{3+n_s(k)} \{1 + ak + (bk)^2 + (ck)^3 + (dk)^4\}^{-1/2} \frac{[\ln(1 + ek)]^2}{(ek)^2}, \\ (a, b, c, d, e) &= (3.97, 16.4, 5.57, 6.85, 2.39) \tilde{\Gamma}^{-1} h^{-1} \text{ Mpc}, \\ \tilde{\Gamma} &= \Gamma_{eq} e^{-\Omega_B(1+\Omega_{nr}^{-1}(2h)^{1/2})}, \end{aligned}$$

is best fit by $\Gamma = 0.53$, and with the $\tilde{\Gamma}$ form the fits are at least as good as the eq. (214) form [250]. The coefficients have been increased by $(2.728/2.70)^2$ over the BBKS values. For the standard CDM model, both transfer function formulae fit to better than 3% to $k^{-1} = 1 h^{-1} \text{ Mpc}$, with eq.(214) better over the crucial large scale structure region.)

To fit the APM angular correlation function using a power spectrum for galaxies described by eq. (214) requires $0.15 \lesssim \Gamma \lesssim 0.3$ [231] for $n_s = 1$ and $0.2 \lesssim n_s \lesssim 0.6$ for $\Gamma = 0.53$ [190, 249]. More generally, $dn_{p,eff}(k)/d\Gamma \approx 2$ over the APM waveband, hence it is $\Gamma + \nu_s/2$ that should lie in the 0.15-0.3 range [6]. A recent estimate of Γ using power spectra from redshift surveys as well as from the APM data suggests $\Gamma \approx 0.23$ fits best [250]. Figure 24 compares the COBE-normalized $n_s = 1, \Gamma = 0.5$ linear density power spectrum with an $n_s = 1, \Gamma = 0.25$ and an $n_s = 0.6, \Gamma = 0.5$ spectrum.

To lower Γ into the 0.15 to 0.3 range one can [233]: lower h ; lower Ω_{nr} ; or raise Ω_{er} ($= 1.68\Omega_\gamma$ with the canonical three massless neutrino species present). Raising Ω_B also helps. Low density CDM models in a spatially flat universe (i.e. with $\Lambda > 0$) lower Ω_{nr} to $1 - \Omega_\Lambda$. CDM models with decaying neutrinos raise Ω_{er} [233, 252]: $\Gamma \approx 1.08\Omega_{nr}h(1 + 0.96(m_\nu\tau_d/\text{keV yr})^{2/3})^{-1/2}$, where m_ν is the neutrino mass and τ_d is its lifetime. Decaying neutrino models have the added feature of a bump in the power at subgalactic scales to ensure early galaxy formation, a consequence of the large effective Ω_{nr} of the neutrinos before they decayed. As we saw in section 6.3, we expect a tilt in inflation models, so we can probably relax the amount by which Γ needs to be lowered. One could do it entirely by tilt by invoking one of the inflation models of section 6.3 utilizing a deceleration parameter $q \approx -(n_s + 1)/2$ or, for natural inflation, the curvature in $\ln H$ away from the peak of the potential, $\frac{m_\nu^2}{4\pi} \partial^2 \ln H / \partial \phi^2 \approx (n_s - 1)/2$.

Generally, more scales are needed to characterize the spectrum than just k_{Heq} :

$$k_{\nu d\text{amp}}^{-1} \approx 6 (\Omega_{nr}\Omega_\nu(2h)^2)^{-1/2} (g_{m\nu}/2)^{1/2} h^{-1} \text{ Mpc}, \quad (215)$$

$$k_{H\text{rec}}^{-1} \approx 41 (\Omega_{nr})^{-1/2} h^{-1} \text{ Mpc}, \quad (216)$$

$$k_{S\text{ilk}}^{-1} \approx 3.8 (\Omega_{nr})^{-1/2} h^{-1} \text{ Mpc}, \quad (217)$$

$$k_{JB\text{rec}}^{-1} \approx 0.0016 (\Omega_{nr})^{-1/2} h^{-1} \text{ Mpc}, \quad (218)$$

$$k_{curv}^{-1} \equiv d_{curv} \approx 3000 |1 - \Omega_{tot}|^{-1/2} h^{-1} \text{ Mpc}. \quad (219)$$

These are: (215) the collisionless damping scale for hot dark matter (massive neutrinos), with $g_{m\nu}$ the number of massive species (counting particle and antiparticle); (216) the horizon scale at recombination; (217) the Silk damping scale; (218) the baryon Jeans length at recombination (below which the baryon power spectrum in CDM models is effectively filtered, multiplying the power by approximately $(1 + (k/k_{JB\text{rec}})^2/2)^{-2}$); and (219) the curvature scale for open universes (in which case k is not exactly wavenumber).

One could try to mimic some of these effects on the power spectrum by modifying Γ . In hot/cold hybrid models, there is a stable light neutrino of mass m_ν contributing a density $\Omega_\nu = 0.3(m_\nu/7.2 \text{ eV})(2h)^{-2}$, combining with the CDM and baryon densities to make a total $\Omega_{nr} = 1$. A Γ -shape is not a very accurate representation of the entire spectrum, dropping from about 0.5 for small k to $\Gamma \sim 0.22(\Omega_\nu/0.3)^{-1/2}$ over the band $0.04\text{--}2 (h^{-1} \text{ Mpc})^{-1}$ of relevance to large scale structure calculations [233, 253, 231]. For pure hot

dark matter models, BBKS showed that a good fit is provided by a Γ -law – with $\Omega_{er} = 1.46\Omega_\gamma$ for one species of massive neutrino (hence $\Gamma = 1.07\Gamma_{cdm}$) – but with an exponential filtering multiplying the Γ -form of $\mathcal{P}_\rho^{1/2}$: $D_\nu = \exp[-0.32(kR_{f\nu}) - (kR_{f\nu})^2]$, where $R_{f\nu} = 2.6(\Omega_\nu h)^{-1} \text{ h}^{-1} \text{ Mpc}$. The damping is dominated by the Gaussian part of the filter. If we define a characteristic Gaussian filtering length by the radius at which the filtering function drops to $1/e^2$, then this radius defines $k_{\nu damp}^{-1} = 1.1R_{f\nu}$. For the mixed hot/cold models, the $\mathcal{P}_\rho^{1/2}$ modification factor

$$D_\nu = \left[\frac{(1 + (Ak)^2 + (1 - \Omega_\nu)^{\beta-1} \Omega_{er}(Bk)^4)}{1 + (Bk)^2 - (Bk)^3 + (Bk)^4} \right]^\beta,$$

$$\beta = \frac{1}{4}(5 - \sqrt{25 - 24\Omega_\nu}), \quad B = 10.73 \frac{1.14}{(\Omega_\nu + 0.14)},$$

$$A = 69.06 \frac{(1 + 10.91\Omega_\nu)\sqrt{\Omega_\nu(1 - .9465\Omega_\nu)}}{(1 + (9.26\Omega_\nu)^2)}$$

is quite accurate [255], even for finite Ω_B [261].

For warm dark matter, Γ is the same as for the CDM model and a rough fit to the influence of free-streaming is provided by the exponential damping factor form $D_\nu = \exp[-kR_{fw} - (kR_{fw})^2]$, where $R_{fw} = 0.2(\frac{g_{w,dec}}{100})^{-4/3}(\Omega_{warm}h)^{-1} \text{ h}^{-1} \text{ Mpc}$, where $g_{w,dec}$ is the effective number of particle degrees of freedom when the warm-particles decoupled, typically about 60–300 for minimal grand unified theories over the range of decoupling temperature $T \sim 1\text{--}10^{18} \text{ GeV}$, and $\Omega_{warm}h^2 = 1.0(g_{w,dec}/100)^{-1}(m_{warm}/\text{keV})$.

Scales characterizing the CMB anisotropy power spectrum include k_{Silk}^{-1} , k_{curv}^{-1} , and k_{Hrec}^{-1} (above which causal processes cannot occur at the recombination epoch). In addition, we have seen that $k_{LS}^{-1} \approx (5\text{--}10)\Omega_{nr}^{-1/2} \text{ h}^{-1} \text{ Mpc}$, the fuzziness of the last scattering surface below which destructive interference damps CMB anisotropies, is very important. Associated with these physical scales are angular scales $\theta_{LS} \approx (3'\text{--}6')\Omega_{nr}^{1/2}$ and $\theta_{Hrec} \approx 2^\circ\Omega_{nr}^{1/2}$, evaluated using the angle-distance relation $\theta(d) = 0.95^\circ \Omega_{nr} d/100 \text{ h}^{-1} \text{ Mpc}$ appropriate for an $\Omega = \Omega_{nr} = 1$ universe and for an $\Omega = \Omega_{nr} \ll 1$ universe.

7.2 The observable range in k-space

Figure 24 contrasts k -space filters $F_W(k)$ for representative CMB anisotropy experiments (characterized by ℓ -space filters $\overline{W}_\ell$) with the bands in k -space probed by large scale structure (LSS) observations and the bands associated with the formation of collapsed structures such as clusters. The LSS probes shown are: the angular correlation of galaxies $w_{gg}(\theta)$; the power spectrum and redshift space correlation function of galaxies $\xi_{gg}(r)$ as probed by redshift surveys; large scale streaming velocities LSSV; and the correlation function of clusters ξ_{cc} . The abundances of clusters (“cls”) and groups (“gps”) provide information on slightly smaller scales. Abundances of galaxies (and quasars and

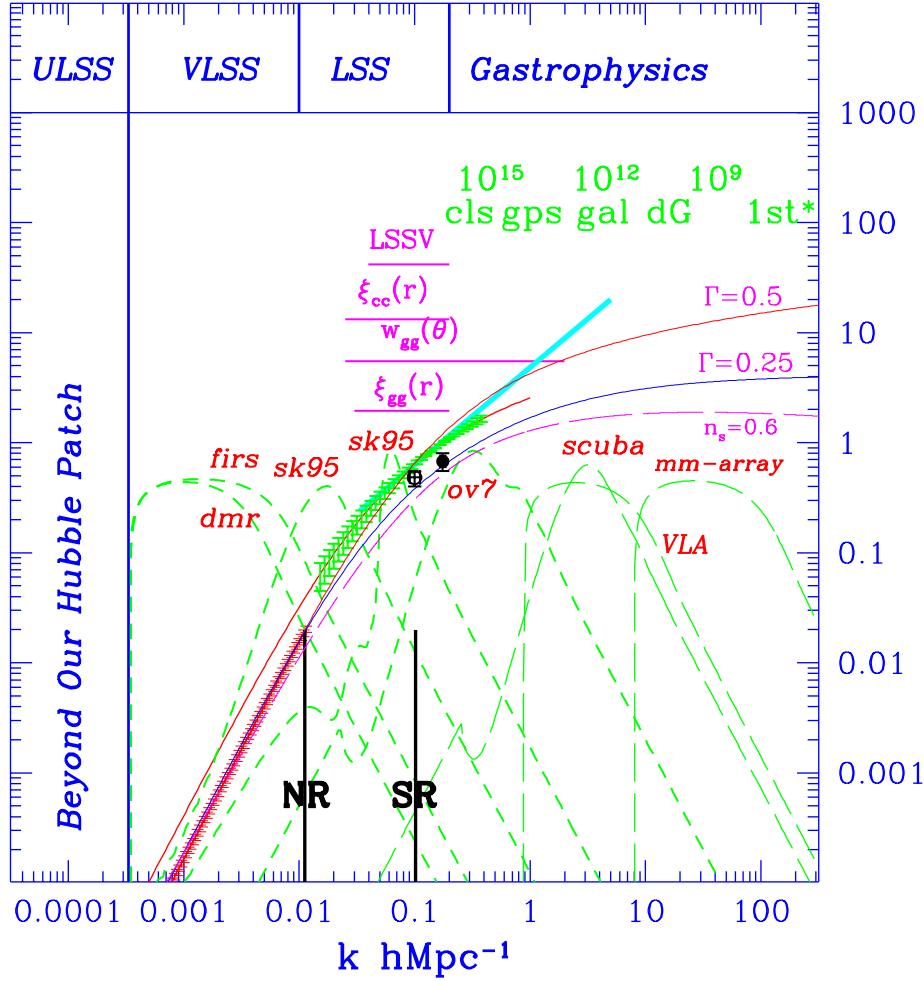


Figure 24: Cosmic waveband probes. The bands of cosmic fluctuation spectra probed by LSS observations are contrasted with the bands that current CMB experiments can probe. The (linear) density power spectrum for the standard $n_s = 1$ CDM model, labelled $\Gamma = 0.5$, is contrasted with (COBE-normalized) power spectra that fit the galaxy clustering data, one tilted ($n_s = 0.6$, $\Gamma = 0.5$) and the other scale invariant with a modified shape parameter ($n_s = 1$, $\Gamma = 0.25$). Biasing must raise the spectra up (uniformly?) to fit into the hatched w_{gg} range and nonlinearities must raise it at $k \gtrsim 0.2 h \text{ Mpc}^{-1}$ to (roughly) match the heavy solid ($\gamma = 1.8$) line. The solid data point in the cluster-band denotes the constraint on the power spectrum from the abundance of clusters, and the open data point at $10 h^{-1} \text{ Mpc}$ denotes an estimate from streaming velocities (for $\Omega_{nr} = \Omega = 1$ models). 138

damped Lyman alpha systems) at high redshift provide valuable information on the power in higher k -bands, but these probes are sensitive to gas dynamical processing which may obscure the hierarchical relationship between object and primordial fluctuation waveband; indeed damping processes or tilted initial spectra may require some of the shorter distance structure to arise from fragmentation and other nongravitational effects.

We can define a k -space filter $F_W(k)$ as one acting upon a k -space “power spectrum for $\Delta T/T$ fluctuations” $\mathcal{P}_{\Delta T}(k)$:

$$F_W(k) = \mathcal{P}_{\Delta T}^{-1} \mathcal{I} \left[\overline{W}_\ell \frac{d\mathcal{C}_\ell}{d \ln k} \right], \quad \mathcal{P}_{\Delta T}(k, \tau) = \mathcal{I} \left[\frac{d\mathcal{C}_\ell(k, \tau)}{d \ln k} \right], \quad (220)$$

with $d\mathcal{C}_\ell/d \ln k$ evaluated at some post-decoupling time τ . $\mathcal{P}_{\Delta T}$ defined in this way is basically conserved through free streaming [88] and $\int \mathcal{P}_{\Delta T} d \ln k$ gives the total anisotropy power. (The choice of $\mathcal{P}_{\Delta T}$ is really up to the theorist; e.g., a filter acting on the primordial gravitational potential power spectrum can be constructed by choosing $\mathcal{P}_{\Delta T} = \mathcal{P}_{\Phi_N}$.) In [88], we showed that a rather good approximation to $\mathcal{C}_\ell$ is obtained by putting the $\mathcal{P}_{\Delta T}(k)$ of eq. (220) in place of $\mathcal{P}_{\mathcal{G}_0 \mathcal{G}_0}(k)$ in eq. (125) and $\delta(\tau - \tau_s)$ in place of $\mathcal{V}$, where τ_s is the time at which $\mathcal{P}_{\Delta T}(k, \tau_s)$ is evaluated (which should be well after decoupling). The k -filters actually plotted in the figure use the high ℓ approximation for the Bessel function product:

$$\mathcal{F}_W(k) = \sum_{\ell < k\chi_s - \frac{1}{2}} \frac{(\ell + \frac{1}{2})}{k\chi_s \sqrt{(k\chi_s)^2 - (\ell + \frac{1}{2})^2}} \overline{W}_\ell. \quad (221)$$

The filters shown in fig. 24 are for the large angle *dmr* and *firs* experiments, with beams $\sim 7^\circ$ and $\sim 3.9^\circ$, two filters for the *sk95* experiment, showing the k -space that it covers, and the Caltech OVRO *ov7* (1.8' beam) experiment. *max* and *msam* cover ranges between the *sk95* points. CAT [151], WhiteDish [102] and a new OVRO (7' beam) experiment (*ov22*) cover the region between *sk95* and *ov7*. These experiments are all sensitive to primary anisotropies. The line labelled SR shows the length scale below which the primary power is basically erased if hydrogen recombination is standard; NR denotes the scale if there is an early injection of energy which ionizes the medium. These depend upon Ω_{nr} , Ω_B etc. The light long-dashed filters at high k show the bands probed by very small angle microwave background experiments, the VLA, the SCUBA array on the sub-mm telescope JCMT, and the OVRO mm-array. Although their beams are too small to see primary CMB anisotropies, they will provide invaluable probes of secondary anisotropies generated by nonlinear effects, including redshifted dust emission from galaxies and Thomson scattering from nonlinear structures in the pregalactic medium.

The (linear) density fluctuation power spectra (actually their square roots, $\mathcal{P}_\rho^{1/2}(k)$) shown in fig. 24 are for three ($\Omega_{nr} = 1$) models normalized to the COBE *dmr* data (i.e., within the small- k hatched region which includes the

8% *dmr* error on overall amplitude): a standard $\Gamma = 0.5$ CDM model with $n_s = 1$, one with the spectrum tilted to $n_s = 0.6$, and an $n_s = 1$ CDM model whose shape is characterized by $\Gamma = 0.25$. To fit the galaxy clustering data requires $0.15 \lesssim \Gamma + \nu_s/2 \lesssim 0.3$. A biasing factor b_g is relied upon to move the curves up into the allowed w_{gg} band (i.e., into the higher- k hatched region) and nonlinearities to bend the shape upward to match the (approximate) 1.8 law for $k^{-1} < 5 \text{ h}^{-1} \text{ Mpc}$ (heavy line extending the hatched w_{gg} region). Power spectra derived from the QDOT [235], IRAS 1.2 Jansky [236] and CfA2 [237] redshift surveys are compatible with the range inferred from w_{gg} when account is taken of redshift space distortions and biasing offsets between IRAS and optically identified galaxies. Cluster-cluster correlations and galaxy-cluster cross correlations [240] also seem to be compatible with this inferred spectrum. Power spectrum estimates derived from the abundance of clusters as a function of temperature [120] and from the Mark III peculiar velocity catalogue [298] are also shown. There is a lesson to draw from an overview figure like this while we concentrate on the LSS power issue that has led to intense research on variations in the scale-invariant minimal-CDM theme for many years: the great success inherent in the extrapolation over so many decades from COBE normalization to large and small scale structure formation suggests that scale invariance cannot be wildly broken and nonminimality cannot be too extreme, even if the generation mechanism has nothing to do with inflation.

7.3 Relating the cluster-amplitude σ_8 and the *dmr* band-power

Apart from the shape parameters for $\mathcal{P}_\rho(k)$, there is also an overall amplitude parameter, which we now take to be $\langle \mathcal{C}_\ell^{(S)} \rangle_{dmr} = \langle \mathcal{C}_\ell \rangle_{dmr} / (1 + \tilde{r}_{ts})$, where $\tilde{r}_{ts} = \langle \mathcal{C}_\ell^{(T)} \rangle_{dmr} / \langle \mathcal{C}_\ell^{(S)} \rangle_{dmr}$. The band-powers obtained from the 4-year *dmr* data as a function of the phenomenological slope $\nu_{\Delta T}$ for each frequency channel and for the 53+90+31 *A+B* GHz map were given in section 4.5. The effective slope of the standard $n_s = 1$ $\Omega_B = 0.05$ CDM model of figs. 7, 23 is $\nu_{\Delta T} \approx 0.15$ over the *dmr* band; variation in Ω_B and H_0 does not change this very much as fig. 8(b,c) shows; nor does a change in the recombination history (fig. 7). Vacuum-dominated models do raise the slope to low ℓ because of the time-dependence of the gravitational potential [110]: fig. 23 shows it is not well represented by a single power law, but if we were forced to choose an effective index it would be $\nu_{\Delta T} \approx 0$.

Before the COBE detection, normalization of the density spectrum was done using σ_8 , the rms (linear) mass density fluctuations on the scale of $8 \text{ h}^{-1} \text{ Mpc}$, or to a biasing factor b_g for galaxies, which was usually assumed to obey $b_g \sigma_8 \approx 1$ e.g., [134, 243]. The COBE-normalized value of σ_8 is thus extremely important for deciding on viability of any specific model of cosmic structure formation. Bayesian determinations of σ_8 from the *dmr* data for a number of selected models can be used to calibrate a more general relation

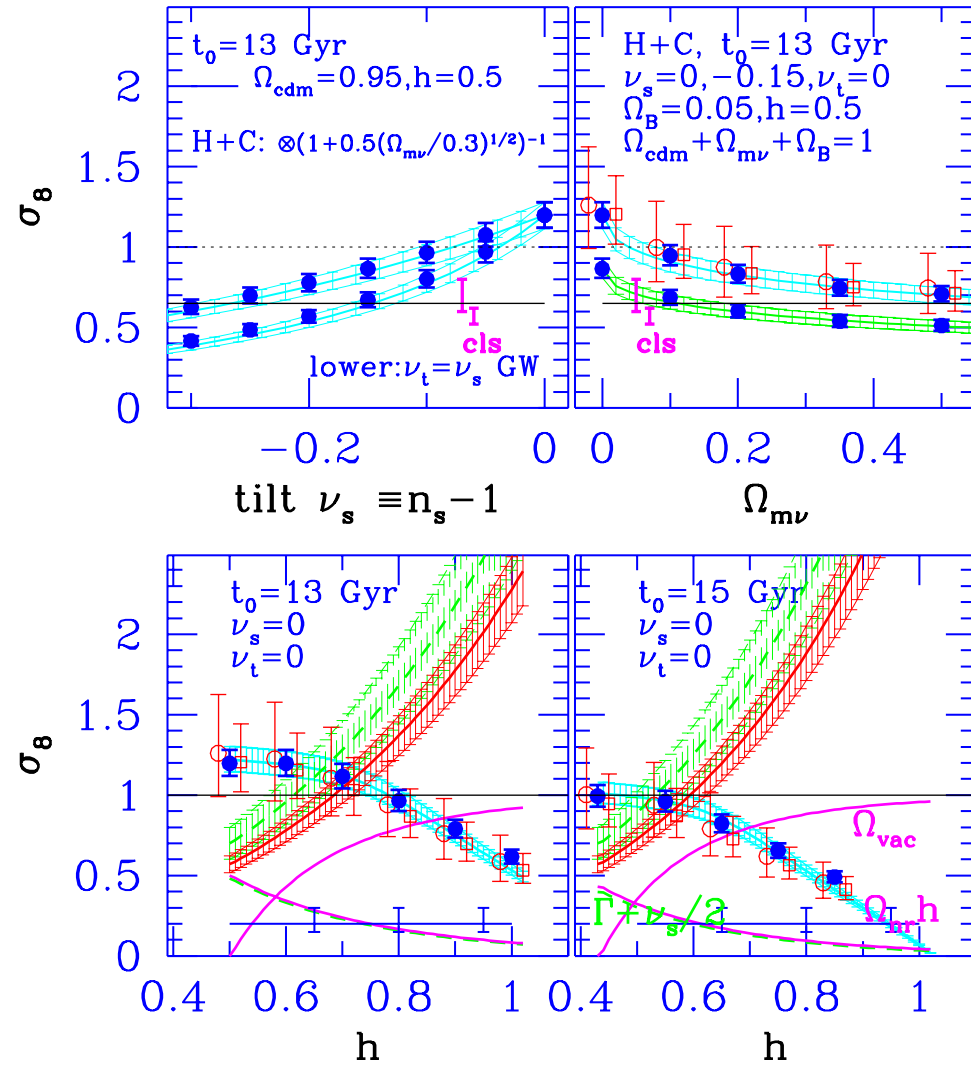


Figure 25: This illustrates the accuracy and utility of the fitting formula for σ_8 . Top left shows the average and $\pm 1\sigma$ variation of σ_8 against tilt for $\Omega_{nr} = 1$ CDM models, with no gravity waves (upper) and with them. A reduction factor for hot/cold hybrid models is also given. The heavy closed data points are σ_8 's derived using the exact $\mathcal{C}_\ell$. The two vertical lines denote two estimates of σ_8 from clusters. Upper right shows $\sigma_8(\Omega_{m\nu})$ for $n_s = 1$ (upper) and 0.85 hot/cold models. Open circles shifted left of the *dmr* points are σ_8 's for the *sp94* data, open squares for the *sk93+94* data. The lower panels show $\sigma_8(h)$ for a sequence with fixed age, no mean curvature, and $\Omega_{vac}(h) = 1 - \Omega_{nr}$, the rising curve. The solid dropping curve is $\Omega_{nr}h$ and the almost indistinguishable dashed one is Γ , the error bars defining its likely range. The rising hatched regions are the two cluster σ_8 estimates.

between σ_8 and the dmr band-power, following [231, 190, 6]. Although we have seen in fig. 23 that the naive Sachs–Wolfe formula with $d\mathcal{C}_\ell/d\ln k = 2\ell(\ell+1)\mathcal{P}_\Phi(k)j_\ell^2(k\chi_{dec})/9$ is not particularly good for the standard CDM model, we can use the scalings predicted by $\mathcal{P}_\Phi$ to parameterize the $\langle\mathcal{C}_\ell\rangle_{dmr}^{1/2}/\sigma_8$ relation:

$$\begin{aligned} \Gamma\text{-law:} \quad \sigma_8 &\approx \frac{1.25}{f_{SW}} \frac{10^5 \langle\mathcal{C}_\ell\rangle_{dmr}^{1/2}}{(1+\tilde{r}_{ts})^{1/2}} \frac{\Omega_{nr}^{-0.77}(2(\Gamma-0.03))}{(1+0.55(\Omega_\nu/0.3)^{1/2})} e^{2.63\nu_s}, \quad (222) \\ 4\text{yr}(53\&90\&31)\text{a\&b:} \quad 10^5 \langle\mathcal{C}_\ell\rangle_{dmr}^{1/2} &\approx [0.82 + 0.26(1 - \frac{\nu_{\Delta T}}{2})^{2.8}] \times 1_{-0.06}^{+0.07}, \\ f_{SW} &\approx (1 + 0.12\Omega_{vac})(1 + \Omega_{vac}^{10}), \quad \nu_{\Delta T} \approx 0.15(1 - \Omega_{vac}) + \nu_s, \\ \tilde{r}_{ts} &\approx 5.4 \frac{(-\nu_t)}{(1-\nu_t/2)} e^{-0.07\nu_t} e^{-1.99(\nu_s-\nu_t)} (1 - 0.6\Omega_{vac}^{3.5}). \end{aligned}$$

Here $\nu_s \equiv n_s - 1$. The fit was originally made in n_s with Γ fixed at 0.5, and in Γ with n_s fixed at 1, but it works well even when both vary significantly from these standard values. The $\Omega_{vac} = 0$ formulae are the same as I gave in [6, 146]. How well the fitting formula does is shown in fig. 25. As expected, this calibration works well for the H_0 variation of fig. 8(c) and, using $\Gamma = \Gamma_{eq} e^{-[\Omega_B(1+\Omega_{nr}^{-1}(2h)^{1/2})-0.06]}$, works well for the cases of fig. 8(b).

Although $f_{SW} = 1$ takes into account some of the enhancements over the naive Sachs–Wolfe formula by normalizing to the calculated $\sigma_8\text{--}\langle\mathcal{C}_\ell\rangle_{dmr}^{1/2}$ relation for standard CDM, it does not take into account the enhancement of $\langle\mathcal{C}_\ell\rangle_{dmr}^{1/2}$ associated with the time dependence of the gravitational potential when Λ dominates, hence for that case we expect f_{SW} to exceed unity. Using the Ω_{vac} dependences of f_{SW} , $\nu_{\Delta T}$ and $\tilde{r}_{ts}$ (section 6.4) allows good σ_8 fits, as the lower panels of fig. 25 show. All models shown have $\Omega_B h^2 = 0.0125$, with the rest of the nr -matter in cold dark matter ($\Omega_{nr} = \Omega_{cdm} + \Omega_B$). For these sequences of models with a uniform age t_0 , the variation of Ω_{vac} with Hubble parameter (the rising curve in fig. 25) is

$$\begin{aligned} h &= h_1 \Omega_{nr}^{-1/2} \frac{\ln[\sqrt{\Omega_{vac}/\Omega_{nr}} + \sqrt{\Omega_{vac}/\Omega_{nr} + 1}]}{\sqrt{\Omega_{vac}/\Omega_{nr}}}, \quad h_1 \equiv 0.5(13\text{Gyr}/t_0), \\ \Omega_{vac}(h) &\sim 0.9(0.3(h/h_1 - 1)^{0.3} + 0.7(h/h_1 - 1)^{0.4}). \end{aligned} \quad (223)$$

The latter is a rough inversion. The ages shown in fig. 25 bracket a recent estimate for globular cluster ages, $14.6_{-1.6}^{+1.7}$ Gyr [111]. The $\Omega_{vac} = 0$ model with 13 Gyr age is therefore the $H_0 = 50$ standard CDM model, and $H_0 = 43$ for the 15 Gyr age.

For open CDM models, the COBE-determined σ_8 goes down with decreasing Ω (and increasing h). These models are not so attractive because Ω drops so precipitously with increasing h for fixed age. Equation (222) has not been modified to treat open models (see e.g., [292]).

Section 4.6 showed that different combinations of cosmological parameters can lead to sufficiently similar spectra that it will be quite an experimental

challenge to differentiate among them [144]. In the near term, we must rely on such important ratios as $\sigma_8/\langle\mathcal{C}_\ell\rangle_{dmr}^{1/2}$ and the shape of the galaxy correlations to further constrain cosmological parameter space. We can also hope to constrain parameters through observations of galaxies at high redshift and by large scale streaming velocities. As is evident from fig. 24, to have a COBE-normalized power spectrum pass through the error bars associated with the power spectrum from cluster abundances on the scale of $\sim 0.2\text{ h}^{-1}\text{ Mpc}$ and the LSSV estimate at $\sim 0.1\text{ h}^{-1}\text{ Mpc}$ [297, 298], and to satisfy the shape restriction, albeit with a free galaxy biasing factor b_g , is like threading the eye of a needle, and clearly severely restricts the range of models. Much discussion in the post-COBE era has been about which COBE-normalized models pass these tests. We now consider a few examples of the use of eq. (222) in conjunction with the current large scale structure data. As an illustration here I will consider the shape and cluster constraints and, to a lesser extent, the LSSV constraint, on models of fixed age with variable tilt and Ω_{vac} . In fig. 25, the dashed curve shows the power spectrum shape parameter Γ , almost indistinguishable from the $\Omega_{nr}h$ curve, for the 13 and 15 Gyr model sequence. The rising curves with error bars denote estimates of σ_8 from clusters. The upper rising regions also roughly denote the σ_8 behavior, as derived from optical galaxy samples, in units of $b_g\sigma_8$.

The mass enclosed within $8\text{ h}^{-1}\text{ Mpc}$ is that of a typical rich cluster, $1.2 \times 10^{15}\Omega_{nr}(2h)^{-1}M_\odot$. Because rich clusters are rare events in the medium, their number density is extremely sensitive to the value of σ_8 . The abundance as a function of cluster mass, velocity dispersion or X-ray temperature also depends upon the shape of $\mathcal{P}_\rho(k)$ in the cluster band of fig. 24, i.e., on n_s and Γ . Cluster X-ray data implies $0.6 \lesssim \sigma_8 \lesssim 0.8$ for CDM-like $\Omega_{nr} = 1$ theories, with the best value depending upon Γ , n_s , some issues of theoretical calibration of models, and especially which region of the dn_{cl}/dT_X data one wishes to fit, since the data prefer a local spectral index $d\ln\mathcal{P}_\rho/d\ln k$ substantially flatter over the cluster region than the standard CDM model gives [120]. I believe a good target number is 0.7 and values below about 0.5 are unacceptable, but because CDM spectra do not fit the data well, this normalization depends upon whether one focusses on the high or low temperature end. Other authors who concentrated on the low to median region found lower values for $n_s = 1$ models, 0.57 ± 0.05 [293] and 0.50 ± 0.04 [294], but do not fit the high T_X end well. (A small upward correction should be applied to these low estimates to account for the nonzero redshift of the calibrating samples.) For $\Omega_\Lambda \neq 0$, a higher value is better [231, 254]; [293] adopt $\Omega_{nr}^{-0.56}$ as the correction, [294] give a more moderate dependence, $\Omega_{nr}^{-0.53+0.13\Omega_{nr}}$ for nonzero vacuum models, $\Omega_{nr}^{-0.47+0.10\Omega_{nr}}$ for open models. The rising curves with error bars in Fig. 25 show the higher and lower σ_8 estimates from clusters. Allowed models would have to lie in the overlap region between the cluster σ_8 and the dmr σ_8 .

There are many estimates of the combination $\sigma_8\Omega_{nr}^{0.56}$ that are obtained by relating the galaxy flow field to the galaxy density field inferred from redshift surveys, which all take the form $[b_g\sigma_8]\beta_g$, where β_g is a numerical factor whose

value depends upon data set and analysis procedure: in [295], the rather varied estimations are reviewed, and raw averages are given, 0.78 ± 0.33 for IRAS-selected galaxy surveys, 0.71 ± 0.25 for optically-selected galaxy surveys. Even more recent estimates give $\beta_g \sim 0.4\text{--}0.6$. (In this case, the $\Omega_{nr}^{0.56}$ is the factor by which the linear growth rate $\dot{D}/D$ differs from the Hubble expansion rate $\dot{a}/a$.) For this to be a σ_8 estimator requires the simplifying assumption of linear amplification bias, and a choice for $b_g\sigma_8$. It is usual to take $b_g \approx \sigma_8^{-1}$ for galaxies, but b_g can depend upon the galaxy types being probed, upon scale, and could be bigger or smaller than σ_8^{-1} , and certainly cannot be determined by theory alone. Recent estimates of parameters derived from the LSSV data, in this case the Mark III velocity catalogue [296], are $\Gamma = 0.5 \pm 0.15$, b_g nearly unity and $\sigma_8\Omega_{nr}^{0.56} \sim 0.85 \pm 0.1$, with sampling errors adding another ~ 0.1 uncertainty [297, 298]. The emphasis in [298] is on $\mathcal{P}_\rho(k)$ estimation from the LSSV data since it allows a direct comparison with models in figures like fig. 24. For example, they give $\mathcal{P}_\rho^{1/2}\Omega_{nr}^{0.56} = 0.48_{-0.08}^{+0.07}$ at $k^{-1} \approx 10 \text{ h}^{-1} \text{ Mpc}$, which compares with $\mathcal{P}_\rho^{1/2} \sim 0.54\sigma_8(1 - 0.65(\Gamma + \nu_s/2 - 0.5))$ for tilted Γ models. However, the 17% should be augmented by a theoretical “cosmic variance” sample error, which may be quite large. In [299], parametric models give similar results, $\mathcal{P}_\rho^{1/2}\Omega_{nr}^{0.56} = 0.49_{-0.08}^{+0.07}$, $\sigma_8\Omega_{nr}^{0.56} \sim 0.88 \pm 0.15$. (Earlier work on LSSV concentrated on estimates of large scale *rms* bulk flows, e.g., over 40 and 60 $\text{h}^{-1} \text{ Mpc}$ regions: $\sigma_v(40 \text{ h}^{-1} \text{ Mpc})$ had the same 17% data errors, but there the cosmic variance fluctuations contributed 50% uncertainty; even so the $\Gamma = 0.5$ model needed $n_s > 0.83$ with the typical gravity wave contribution and > 0.55 without [190].) Since the peculiar velocity data relies on having spatially-independent and accurate distance indicators (e.g., the empirical Tully–Fisher relation between luminosity and rotation velocity in spiral galaxies), how seriously we take the LSSV constraints depends upon how reliable we think the indicators are – a subject of much debate.

Fig. 25 shows σ_8 is a sensitive function of n_s : for CDM models with $\Omega_{nr} = 1$, it is far too high at 1.2 for $n_s=1$, but too low by $n_s \approx 0.76$ with the “standard” gravity wave contribution ($\nu_t = \nu_s$) or by $n_s \approx 0.60$ if there is no tensor mode contribution. However, the shape constraint wants lower n_s . In [162], we marginalize likelihood functions determined with the COBE data (and smaller angle data) using a prior probability requiring that $\Gamma + \nu_s/2$ be 0.22 ± 0.08 and $\sigma_8\Omega_{nr}^{0.56}$ be $0.65_{-0.08}^{+0.15}$ in order to condense the tendencies evident in fig. 25 into single numbers with error bars. Threading the “eye of the needle” this way is so exacting that the error bars are too small to take too seriously. Sample numbers using only the 4-year *dmr* data and these priors are $n_s = 0.76_{-0.03;0.06}^{+0.03;0.06}$ for $h = 0.5$ with gravity waves, $n_s = 0.61_{-0.04;0.08}^{+0.04;0.09}$ without. For $h = 0.7$ and $\Omega_{vac} = 0.66$, we get $n_s = 0.99_{-0.02;0.04}^{+0.03;0.06}$, and when Hubble parameters in the range from 0.5 to 1 are marginalized over, the preferred index is $n_s = 0.99_{-0.04;0.08}^{+0.06;0.18}$ with gravity waves, $n_s = 0.95_{-0.10;0.17}^{+0.09;0.19}$ without. These are of course significantly better than can be determined from *dmr* alone (section 4.5).

For the decaying neutrino model with $n_s = 1$ to have $\sigma_8 > 0.5$ we need $\Gamma > 0.22$, i.e., $m_\nu \tau_d < 14$ keV yr. The hot/cold hybrid model formula in eq. (222) is for one massive neutrino species. As fig. 25 shows, an $n_s=1$ hot/cold hybrid model with $\Omega_\nu < 0.3$ would have $\sigma_8 > 0.8$; however, even with a modest tilt to $n_s = 0.95$ this can drop to 0.7 for $\Omega_\nu = 0.25$. (See also ref. [255].) That is, little tilt is required, in contrast to the CDM case.

It is also evident from fig. 25 that the cluster data in combination with the dmr data stops h from becoming too high for a fixed age, but also would prefer a nonzero Λ value, with $H_0 \sim 60 - 70$ for 13 Gyr, and $H_0 \sim 50 - 60$ for 15 Gyr. When the tilt is allowed to vary as well, the preferred values lower to very near 50 and 43, respectively, i.e., with little Ω_{vac} : $h < 0.70$ at 2σ with gravity waves, $h < 0.56$ with no gravity waves for 13 Gyr; $h < 0.56$ at 2σ with gravity waves for 15 Gyr. For the hot/cold models, the values near 50 and 43 are preferred even more, even with very little tilt.

The redshift of galaxy formation cannot be too low or we would get too few $z \sim 4$ quasars and too little neutral gas compared with that inferred using the damped Lyman alpha systems seen in the spectra of quasars. A fairly conservative estimate of the redshift of galaxy formation is [190] $z_{gf} \approx 1.3\sigma_{0.5}\Omega_{nr}^{-0.23} - 1$, where $\sigma_{0.5} \equiv \sigma_\rho(0.5 h^{-1} \text{ Mpc})$ is the analogue of σ_8 but at a galactic mass scale rather than a cluster mass scale and $D/a \approx \Omega_{nr}^{-0.23}$ for the linear growth rate $D(t)$ at high redshift has been used. This suggests $2 \lesssim \sigma_{0.5}\Omega_{nr}^{-0.23} \lesssim 5$ or so. For the Γ models with tilt we have roughly $\sigma_{0.5} \sim 6.4\sigma_8 e^{\nu_s} (\Gamma/0.5)^{0.44}$. (If we characterize galactic scales by the baryonic mass then we should use $\sigma_{1 \text{ Mpc}} \equiv \sigma_\rho(1 \text{ Mpc})$ rather than $\sigma_{0.5}$ in the z_{gf} estimation if $\Omega_B h^2$ is treated as fixed by primordial nucleosynthesis. For the Γ models with tilt, $\sigma_{1 \text{ Mpc}} \sim \sigma_{0.5}/(2h)^{0.3}$.) The z_{gf} requirement leads to serious constraints on n_s in standard CDM models: $n_s > 0.76$ with gravity waves, $n_s > 0.63$ without. With $\Gamma < 0.5$, the restrictions on the primordial spectral index from galaxy and cluster formation are even more severe (for $\Omega_{nr} \sim 1$), but the $\Omega_{nr}^{-0.23}$ factor ameliorates the situation for $\Lambda \neq 0$ models. The z_{gf} constraint is also the Achilles' heel of hot/cold hybrid models with $\Omega_\nu \gtrsim 0.3$ [233, 255]. Observations of the CMB on small scales could in principle help to normalize the power spectrum there; e.g., using sub-mm sky observations as in fig. 15 (if one could get redshifts by other means).

7.4 The future

A consistent story that accommodates all of the current data on the CMB, large scale structure, the Hubble parameter, the ages of stars, the deceleration parameter, clusters, lensing, etc. does not yet leap out at us. With the large Sloan and 2df redshift surveys, we will have a wealth of LSS data to compare with the evolving CMB spectrum, and many of the current puzzles will be definitively answered. As we have seen, if just the shape of the density power spectrum over the LSS band and the amplitude of the power spectrum on cluster scales are considered to be known, then the range of inflation and dark matter models

is restricted considerably when combined with the COBE anisotropy level (and indeed the anisotropy levels of intermediate angle experiments). Whether the solution will be a simple variant on the CDM+inflation theme [233], involving slight tilt (or more radical broken scale invariance), stable $e\nu$ -mass neutrinos, decaying ($>keV$)-neutrinos, vacuum energy, low H_0 , high baryon fraction, negative mean curvature or some combination, is still open, but can be decided as the observations tighten, and, in particular, as the noise in the $\mathcal{C}_\ell$ figure subsides, revealing the details of the Doppler peaks, a very happy future for those of us who wish to peer into the mechanism by which structure was generated in the Universe.

Although there are undoubtedly many surprises in store for us as the anisotropy data improves, we should be very encouraged by how far we have come since the COBE discovery. We are now beginning to map the sky's primary and secondary anisotropy signals. It is fitting to end by pointing back to fig. 11 that shows the anisotropy at low resolution as revealed by COBE, and forward to the interferometric arrays (VSA, CBI, VCA), long duration balloon experiments (ACE, Boomerang, Maxima, Top Hat,...) and especially the all-sky satellite experiments (MAP, COBRAS/SAMBA), that will tell us the parameters defining how cosmic structure formed in detail.

Acknowledgements

First a nod to my fellow CMB enthusiasts who kept digging for three decades until gold was revealed. Now we're all rich. And the data-bank deposits just keep accumulating. Fifteen years of collaboration with George Efstathiou on CMB topics and strong interactions with Bernard Carr and Craig Hogan on the spectrum, Steve Myers, Paul Steinhardt, Rick Davis, Rob Crittenden, Andrew Jaffe, Lloyd Knox, Yoram Lithwick, Dmitry Pogosyan and Tarun Souradeep on the anisotropy is especially noted. Support from a Canadian Institute for Advanced Research Fellowship, NSERC and from the Institut d'Astrophysique de Paris, where some of these lecture notes were written, is gratefully acknowledged.

A The ADM formalism and perturbation theory

The ADM treatment of the Cauchy problem in relativity [167] is well covered in MTW [196]). The ADM formalism is the natural language for numerical relativity, so there has been intense post-MTW development; in particular, Jimmy York's highly influential 1979 Batelle and 1982 Les Houches lecture notes [168, 169]. The approach to perturbation theory which I ascribe to [195, 2, 192] is based upon this 3+1 split. I usually use either the synchronous gauge or the longitudinal gauge, but with liberal use of transformation to other

variables and hypersurfaces if it simplifies analytic or numerical calculations or helps in understanding. This approach underlied Bardeen's influential 1980 paper and many of the main papers in the subject. However, there was also excessive zeal for the "gauge invariant approach" that made sacrosanct the perturbation to the lapse and the inhomogeneous scale factor in the longitudinal gauge. These variables refer to just one choice of time slicing, which is sometimes a rather bad choice from the point of view of hypersurface warping. By contrast, the much-maligned synchronous gauge – for which the hypersurfaces are those on which cold dark matter is at rest – is often excellent and a great workhouse in General Relativity, e.g., Landau and Lifshitz [179]. Bardeen's China lectures [178] redress the balance, giving a clear compact enunciation of the issues starting from the ADM formalism in a paper which deserves to be better known in cosmology.

The main equations for perturbation theory are given in sections A.2, B.4, C.2, C.3.1, C.3.2, C.4 for scalar modes and in sections A.3, C.6 for tensor modes. The other sections develop these equations from first principles.

A.1 The ADM equations

A foliation is a set of spacelike 3-surfaces $\{{}^{(3)}\mathcal{G}\}$ that fills spacetime, for which a closed 1-form Ω exists which is normal to the surfaces. It is therefore locally exact, i.e., can be written as $\Omega = d\tau$, where τ is a time coordinate labelling the hypersurfaces. The metric can be decomposed into the ADM form in terms of the lapse function N , the shift (three) vector N^i , and a spatial metric ${}^{(3)}g_{ij}$:

$$\begin{aligned} ds^2 &= -N^2 d\tau^2 + {}^{(3)}g_{ij}(dx^i + N^i d\tau)(dx^j + N^j d\tau), \\ g_{00} &= -N^2 + N_k N^k, \quad g_{0i} = N_i \equiv {}^{(3)}g_{ij} N^j, \quad g_{ij} = {}^{(3)}g_{ij}, \\ g^{00} &= -\frac{1}{N^2}, \quad g^{0i} = \frac{N^i}{N^2}, \quad g^{ij} = {}^{(3)}g^{ij} - \frac{N^i N^j}{N^2}, \quad {}^{(3)}g^{ij} {}^{(3)}g_{jk} \equiv \delta_k^i. \end{aligned} \quad (224)$$

Here, x^i are local coordinates on the $\tau = \text{constant}$ surfaces and ${}^{(3)}g^{ij}$ is the contravariant spatial 3-metric.

One can refer tensors to the coordinate basis, dx^α and its dual basis $\partial_\alpha \equiv \partial/\partial x^\alpha$ or to a more general contravariant basis (tetrad), e_a , and its covariant dual basis, e^a , where $a = 0, 1, 2, 3$; for the spatial components with respect to the basis I shall use $I, J, K, \dots$. It is natural to choose the 4-velocity $e_n = N^{-1}\partial_0 - N^{-1}N^i\partial_i$ as the timelike basis vector (and $e^n = N d\tau$): it describes observers comoving with the flow of time (section 6.2.1). The spatial triad $\{e_1, e_2, e_3\}$ is chosen to be perpendicular to e^n ($\langle e^n, e_I \rangle = \langle e_n, e^I \rangle = 0$). Thus, $\{e_I\}$ is invariant under the action of the projector $\perp^{\alpha\beta} = g^{\alpha\beta} + e_n^\alpha e_n^\beta$. Tetrads are not usually expressible as coordinate bases (i.e., are nonholonomic), but components of tensors with respect to tetrads often have more direct physical meaning than components referred to coordinates. With the e_I chosen to be perpendicular to e_n , to go from spatial coordinate components of a tensor $\mathcal{T}_{kl\dots}^{ij\dots}$ to triad components $\mathcal{T}_{KL\dots}^{IJ\dots}$, one just forms $\mathcal{T}_{kl\dots}^{ij\dots} e_i^I e_j^J e_k^K e_l^L \dots$; 3-space

spatial covariant derivatives with respect to the 3-metric $^{(3)}g_{ij}$ are denoted by $\mathcal{T}_{kl\dots|m}^{ij\dots}$ or by $^{(3)}\nabla_m \mathcal{T}_{kl\dots}^{ij\dots}$, with $\mathcal{T}_{KL\dots|M}^{IJ\dots}$ or $^{(3)}\nabla_M \mathcal{T}_{KL\dots}^{IJ\dots}$ denoting the action of the covariant derivative $^{(3)}\nabla_{e_M}$ on the tensor. If $\mathcal{T}$ is invariant under projection, then $^{(3)}\nabla \mathcal{T} = [\perp \ ^{(4)}\nabla] \mathcal{T}$, where $^{(4)}\nabla$ is the covariant derivative with respect to the 4-metric $^{(4)}g_{ij}$. The 3-space metric coefficients in the e_I basis is $^{(3)}g_{IJ} \equiv e_I \cdot e_J$, e.g., δ_{IJ} for an orthonormal choice. Considered as matrices, $(e_I^i) = [(e_j^J)^{tr}]^{-1}$. The matrix e_j^J is sometimes called the deformation tensor since $e_j^J dr^j$ gives the proper length of an element of coordinate length dr^j .

In the following equations, we shall refer the time components to the basis e_n , using the subscript n . For the spatial components, because e_I and e^I are just linear combinations at each point of the ∂_i and ∂^i , independent of e_n and e^n , the transformation to basis components involves changing i, j to I, J (with some care for the treatment of the shift; also note that although e_I^0 vanishes, $e^I_0 = \delta^{IJ} N_j e_j^J$ does not for nonzero shift – however, e_n^I does vanish.) It is useful to introduce a modified basis e_{*I} which “takes out the expansion of the Universe” from e_I : $e_{*I}^I = A^{-1} e^I$, $e_{*I} = A e_I$. Here $A(\mathbf{x}, \tau)$ is a “conformal factor” that should reduce to $\bar{a}(\tau)$ for homogeneous backgrounds, but also could be spatially dependent for fluctuations if it results in simplified equations.

The only nonvanishing components of the extrinsic curvature (with respect to the basis) are

$$\begin{aligned} K_{IJ} &= e_I^i e_J^j \frac{1}{2N} \left(-\frac{\partial^{(3)}g_{ij}}{\partial \tau} + N_{j|i} + N_{i|j} \right) \\ &= -\frac{\dot{A}}{NA} \delta_{IJ} - \frac{1}{2N} (\delta_{KJ} e_{*I}^i + \delta_{IK} e_{*J}^i) \dot{e}_{*i}^K + \frac{1}{2N} e_I^i e_J^j (N_{j|i} + N_{i|j}). \end{aligned} \quad (225)$$

The last form assumes the basis is orthonormal, and is explicitly given to show (with zero shift) that this is just the familiar matrix relation for the shear tensor in terms of the deformation tensor when one maps from Lagrangian to Eulerian space in Newtonian dynamics in the expanding Universe.

We define an inhomogeneous Hubble parameter in terms of the trace of the extrinsic curvature $K = K_I^I$ and a hypersurface anisotropic shear in terms of the anisotropic part of the extrinsic curvature, $(K')_{IJ}$. Letting

$$a(\mathbf{x}, \tau) \equiv [\det(^{(3)}g)]^{1/6} = \exp \left[\frac{1}{6} \text{Trace} \ln(^{(3)}g_{ij}) \right], \quad (226)$$

we have

$$\begin{aligned} H \equiv \theta &\equiv -\frac{K}{3} = \frac{1}{N} \frac{\partial \ln a}{\partial \tau} - \frac{1}{3} \frac{1}{N} ^{(3)}\nabla_j N^j = e_n[\ln a] - \frac{1}{3} \frac{1}{N} \partial_j N^j \\ &= \frac{\dot{A}}{NA} + \frac{1}{3} \frac{1}{N} e_{*K}^i \dot{e}_{*i}^K - \frac{1}{3} \frac{1}{N} N_{|j}^j, \\ \sigma_{IJ} &\equiv -(K')_{IJ} \equiv -(K_{IJ} - \frac{1}{3} ^{(3)}g_{IJ} K) \end{aligned} \quad (227)$$

$$\begin{aligned}
&= \left(\frac{1}{2}(\delta_{KJ}e_{*I}^i + \delta_{IK}e_{*J}^i) - \frac{1}{3}\delta_{IJ}e_{*K}^i \right) \frac{1}{N}\dot{e}_{*i}^K \\
&\quad - e_I^i e_J^j \frac{1}{N} \left(\frac{1}{2}(N_{i|j} + N_{j|i}) - \frac{1}{3}{}^{(3)}g_{ij}N_{|k}^k \right), \quad (228) \\
\sigma_{ij} &= \frac{1}{2} \frac{1}{N} a^2 \frac{\partial^{(3)}g_{ij}/a^2}{\partial\tau} - \frac{1}{N} \left(\frac{1}{2}(N_{i|j} + N_{j|i}) - \frac{1}{3}{}^{(3)}g_{ij}N_{|k}^k \right).
\end{aligned}$$

Just as the stress–energy tensor was decomposed in eq. (154), so the Einstein tensor G^{ab} can be decomposed into $\{G_{nn}, G_n^I, \frac{1}{3}G_I^I, G_{IJ} - \frac{1}{3}{}^{(3)}g_{IJ}G_K^K\}$, and the 10 Einstein equations written in this form. The energy constraint equation can be re-expressed as an inhomogeneous Friedmann equation, which is also the general relativistic version of the Poisson–Newton equation:

$${}^{(4)}G_{nn} = \frac{1}{2}({}^{(3)}R + \frac{2}{3}K^2 - (K')_{IJ}(K')^{IJ}) = 8\pi G_N \rho_{\text{tot}}, \quad (229)$$

$$\text{i.e., } H^2 = \frac{8}{3}\pi G_N \rho_{\text{tot}} + \frac{1}{3}\sigma^2 - \frac{1}{6}{}^{(3)}R, \quad \sigma^2 \equiv \frac{1}{2}\sigma_{IJ}\sigma^{IJ}. \quad (230)$$

The momentum constraint equation is

$${}^{(4)}G^{ni} = {}^{(3)}\nabla_j((K')^{ij} - \frac{2}{3}K{}^{(3)}g^{ij}) = 8\pi G_N J_{(e),\text{tot}}^i. \quad (231)$$

The isotropic dynamical equation ($G_I^I/3$) is

$$\frac{2}{3}e_n[K] + \frac{2}{3}\frac{1}{N}{}^{(3)}\nabla^2 N - \frac{1}{3}K^2 - \frac{1}{6}{}^{(3)}R - \frac{1}{2}(K')_{IJ}(K')^{IJ} = 8\pi G_N p_{\text{tot}} \quad (232)$$

The curvature term in eq. (229) can be eliminated by forming the combination $R_n^n = -(G_{nn} + G_I^I)/2$ equation, which is the Raychaudhuri equation for this zero vorticity hypersurface flow ($\omega^2 \equiv \frac{1}{2}\omega_{IJ}\omega^{IJ} = 0$):

$$3e_n[H] + 3H^2 + 2(\sigma^2 - \omega^2) - \frac{1}{N}{}^{(3)}\nabla^2 N + 4\pi G_N(\rho + 3p)_{\text{tot}} = 0. \quad (233)$$

The anisotropic dynamical Einstein equations ($(G')^{ij}$) are

$$\begin{aligned}
&-e_n[(K')_j^i] + K(K')_j^i + {}^{(3)}(R')_j^i - \frac{1}{N} \left({}^{(3)}\nabla^i {}^{(3)}\nabla_j - \frac{1}{3}\delta_j^i {}^{(3)}\nabla^2 \right) N \\
&\quad - \frac{1}{N}(K')_j^k {}^{(3)}\nabla_k N^i + \frac{1}{N}(K')_k^i {}^{(3)}\nabla_j N^k = 8\pi G_N (\Pi_{\text{tot}})_j^i. \quad (234)
\end{aligned}$$

I now give a few examples of the stress–energy tensors which we shall have occasion to use. The stress energy of a classical fluid can be decomposed into a comoving density $\rho_{com} = U_a T^{ab} U_b$, momentum current $J_{(e)com}^a$, pressure p_{com} , and anisotropic stress Π_{com}^{ab} , defined by eq. (154) but with U the 4-velocity of

the fluid in question. The fluid may be imperfect, with shear and bulk viscosity, η, ζ , and a thermal conductivity κ , obeying the constitutive relations [264, 196]:

$$p_{com} = p(\rho_{com}, T) - \zeta \theta_{(U)}, \quad \Pi_{com}^{ab} = -2\eta \sigma_{(U)}^{ab}, \quad (235)$$

$$J_{(e)(U)}^a = -\kappa T \perp_{(U)}^{ab} ({}^{(4)}\nabla_b [\ln T] + \mathcal{A}_{(U)b}), \quad (236)$$

where T is the fluid temperature and $p(\rho, T)$ is the equation of state. The fluid's acceleration is $\mathcal{A}_{(U)b} \equiv {}^{(4)}\nabla_U U_b$, where the subscript (U) indicates projection with respect to U , e.g., $\theta_{(U)} \equiv \perp_{(U)b}^a T^{bc} \perp_{(U)ca}$. The stress energy derived from a distribution function is given by eq. (276) below. A last example is a scalar field, ϕ , interacting through a potential $V(\phi, \dots)$; projecting onto $U = e_n$, we have

$$\rho_\phi = \frac{1}{2}(e_n[\phi])^2 + {}^{(3)}g^{IJ}e_I[\phi]e_J[\phi] + V, \quad (237)$$

$$J_{(e),\phi}^I = -T_{(\phi)n}^I = -{}^{(3)}g^{IJ}e_n[\phi]e_I[\phi],$$

$$p_\phi = \frac{1}{2}(e_n[\phi])^2 - \frac{1}{6}{}^{(3)}g^{IJ}e_I[\phi]e_J[\phi] - V,$$

$$\Pi_{(\phi)IJ} = e_I[\phi]e_J[\phi] - \frac{1}{3}{}^{(3)}g_{IJ}{}^{(3)}g^{KM}e_K[\phi]e_M[\phi].$$

The scalar field evolution equation, ${}^{(4)}\nabla^2\phi = \partial V/\partial\phi$, is

$$\text{scalar field momentum: } \Pi^{(\phi)} = e_n[\phi], \quad (238)$$

$$e_n[\Pi^{(\phi)}] - K\Pi^{(\phi)} - {}^{(3)}g^{IJ}e_J[\ln N]e_I[\phi] - {}^{(3)}\nabla^2\phi + \frac{\partial V}{\partial\phi} = 0.$$

One can split the 20 independent components of the spacetime curvature tensor ${}^{(4)}R_{abcd}$ into 14 that just depend upon the properties of the 3-geometry, as embodied in the space curvature tensor ${}^{(3)}R_{ijkm}$, and upon the extrinsic curvature and its spatial derivatives,

$${}^{(4)}R_{ijkm} = {}^{(3)}R_{ijkm} + (K_{ik}K_{jm} - K_{im}K_{jk}), \quad (239)$$

$${}^{(4)}R_{ijkn} = {}^{(3)}\nabla_j K_{ik} - {}^{(3)}\nabla_i K_{jk}, \quad (240)$$

$$\begin{aligned} {}^{(3)}R_{ijkm} = & {}^{(3)}g_{ik}{}^{(3)}R_{jm} - {}^{(3)}g_{im}{}^{(3)}R_{jk} + {}^{(3)}g_{jm}{}^{(3)}R_{ik} \\ & - {}^{(3)}g_{jk}{}^{(3)}R_{im} + \frac{1}{2}{}^{(3)}R \left({}^{(3)}g_{im}{}^{(3)}g_{jk} - {}^{(3)}g_{ik}{}^{(3)}g_{jm} \right), \end{aligned}$$

and into 6 dynamical components that depend upon how the extrinsic curvature changes in time, i.e., dependent upon ${}^{(3)}\mathcal{G}$ -evolution:

$$\begin{aligned} {}^{(4)}R_{ijn} = & K_{ik}K_j^k + N^{-1}{}^{(3)}\nabla_i{}^{(3)}\nabla_j N \\ & + \left\{ \frac{1}{N}\dot{K}_{ij} - \frac{1}{N}[N^k{}^{(3)}\nabla_k K_{ij} + K_{ik}{}^{(3)}\nabla_j N^k + K_{kj}{}^{(3)}\nabla_i N^k] \right\}. \end{aligned} \quad (241)$$

Eqs. (240), (239) are called the Gauss–Codazzi equations in the differential geometry of surfaces.¹

¹Many of these quantities are most naturally expressed in terms of Lie derivatives: e.g., K_{IJ} is the Lie derivative of ${}^{(3)}g_{ab}$ with respect to e^n , the term in curly brackets in eq. (241) is the Lie derivative of K_{ij} with respect to e^n , $\mathcal{L}_{e^n}K_{ij}$, and the term in square brackets is the Lie derivative of K_{ij} along the shift vector $N^K e_K$, which vanishes for zero shift.

Normal coordinates have $N^i = 0$. In perturbation theory this defines time-orthogonal gauges. Because the equations simplify, this has also often been adopted in numerical relativity. Gaussian normal coordinates have $N = 1$ (or $\bar{a}$) as well, defining the synchronous gauge. There is a gauge which maximizes the 3-space volume, one with $K = 0$, which was used to retard horizon formation in black hole calculations, but is of little interest for cosmology. Constant K hypersurfaces are used to characterize the outcome of inflation calculations, and have been generally advocated for inhomogeneous numerical cosmology because they are singularity-avoiding, e.g., [172]. However, this positive feature is a negative one if we are interested in following the collapse of cosmic structures such as clusters. Other choices that have been used in black hole calculations share this singularity-avoiding characteristic. There is also a large class of comoving hypersurfaces, one for each “type” of matter present, and one on which the total energy current $J_{(e),\text{tot}}^a$ vanishes. These are very useful for deriving source functions, etc. and are sometimes useful for calculations.

Perturbation theory beyond first order in General Relativity depends upon exactly what spacetime we expand about. It is often useful to take out some aspect of the dynamics via a conformal transformation on 4-space ($g_{\alpha\beta} = \Omega^2 \tilde{g}_{\alpha\beta}$) or on 3-space (${}^{(3)}g_{ij} = A^2 {}^{(3)}g_{*ij}$). In the usual cosmological perturbation theory, it is $A^2 = \bar{a}^2$ or $\Omega^2 = \bar{a}^2$ which is removed, but inhomogeneous parts could also be transformed. The spatial metric ${}^{(3)}g_{ij}$ can even in the nonlinear case be decomposed into terms that we can identify with scalar, vector and tensor (transverse traceless) components, but the nature of these depend upon exactly what we pull out in A or Ω and the Einstein equations couple them – unless the metric coefficients and the conformal factors are all treated fully linearly. Nonlinear choices of some interest are $\Omega = N(\mathbf{x}, \tau)$ and $A = a(\mathbf{x}, \tau) \equiv (\det({}^{(3)}g))^{1/6}$.

A.2 Scalar perturbations

In the following, unperturbed variables and covariant derivative operators have bars over them. For scalar perturbations, we have

$${}^{(3)}g_{ij} = {}^{(3)}\bar{g}_{ij}(1 + 2\varphi) - \bar{a}^2({}^{(3)}\bar{\nabla}_j {}^{(3)}\bar{\nabla}_i + {}^{(3)}\bar{\nabla}_i {}^{(3)}\bar{\nabla}_j)\psi, \quad (242)$$

$$g_{00} = -\bar{N}^2(1 + 2\nu), \quad g_{0i} = N_i = \bar{N}({}^{(3)}\bar{\nabla}_i \Psi_n), \quad (243)$$

$$e_0^n = -\bar{N}(1 + \nu), \quad e_n^0 = \bar{N}^{-1}(1 - \nu), \quad e_n^i = -({}^{(3)}\bar{\nabla}^i \Psi_n),$$

$$\Psi_\sigma \equiv \Psi_n + \frac{\bar{a}^2}{\bar{N}}\dot{\psi}, \quad (244)$$

$$(\delta H) \equiv -\frac{1}{3}(\delta K) = \frac{1}{\bar{N}}\dot{\varphi} - \bar{H}\nu - \frac{1}{3}({}^{(3)}\bar{\nabla}^2 \Psi_\sigma), \quad (245)$$

$$\sigma_j^i \equiv -(K')_j^i = -({}^{(3)}\bar{\nabla}^i {}^{(3)}\bar{\nabla}_j - \frac{1}{3}\delta_j^i {}^{(3)}\bar{\nabla}^2)\Psi_\sigma, \quad (246)$$

$$(\delta {}^{(3)}R')_j^i = -[({}^{(3)}\bar{\nabla}^i {}^{(3)}\bar{\nabla}_j - \frac{1}{3}\delta_j^i {}^{(3)}\bar{\nabla}^2)]\varphi, \quad (247)$$

$$(\delta^{(3)}R) = -4^{(3)}\bar{\nabla}^2\varphi - {}^{(3)}\bar{R}2\varphi, \quad {}^{(3)}\bar{R} = 6\frac{k_c}{d_{curv}^2\bar{a}^2}, \quad (248)$$

$$\rho_{\text{type}} = \bar{\rho}_{\text{type}} + (\delta\rho)_{\text{type}} = \bar{\rho}_{\text{type}}(1 + \delta_{\text{type}}), \quad (249)$$

$$J_{(e),\text{type}I} = T_{\text{type}nI} = -(\bar{\rho} + \bar{p})_{\text{type}} {}^{(3)}\bar{\nabla}_I \Psi_{v,\text{type}}, \quad (250)$$

$$U_{\text{type}I} = -{}^{(3)}\bar{\nabla}_I \Psi_{v,\text{type}}, \quad p_{\text{type}} = \bar{p}_{\text{type}} + (\delta p)_{\text{type}},$$

$$\text{fluid acceleration: } \mathcal{A}_{(U)I} = -{}^{(3)}\bar{\nabla}_I \Psi_{\mathcal{A},\text{type}},$$

$$\Psi_{\mathcal{A},\text{type}} = \frac{1}{N} \dot{\Psi}_{v,\text{type}} - \nu,$$

$$\text{hypersurface acceleration: } \mathcal{A}_{(e_n)I} = {}^{(3)}\bar{\nabla}_I \nu, \quad (251)$$

$$\Pi_{\text{type}ij} = ({}^{(3)}\bar{\nabla}_i {}^{(3)}\bar{\nabla}_j - \frac{1}{3} {}^{(3)}\bar{g}_{ij} {}^{(3)}\bar{\nabla}^2) \bar{p}_{\text{type}} \pi_{t,\text{type}},$$

$$\text{scalar field: } \phi = \bar{\phi} + \delta\phi, \quad \Psi_{v,\phi} = (\bar{e}_n[\phi])^{-1} \delta\phi. \quad (252)$$

$$(\delta\rho)_{\text{tot}} \equiv \bar{\rho}_{\text{tot}} \delta_{\text{tot}} \equiv \sum_{\text{type}} \bar{\rho}_{\text{type}} \delta_{\text{type}},$$

$$\Psi_{v,\text{tot}} \equiv \frac{\sum_{\text{type}} (\bar{\rho} + \bar{p})_{\text{type}} \Psi_{v,\text{type}}}{(\bar{\rho} + \bar{p})_{\text{tot}}}, \quad (253)$$

$$J_{(e),\text{tot}I} = T_{\text{tot}nI} = -(\bar{\rho} + \bar{p})_{\text{tot}} {}^{(3)}\bar{\nabla}_I \Psi_{v,\text{tot}}. \quad (254)$$

Thus φ fully parameterizes the Ricci 3-space tensor ($k_c = 0, \pm 1$ gives the 3 FRW curvature possibilities). The velocity potentials for “type”-matter are $\Psi_{v,\text{type}}$. It is also convenient to define a total velocity perturbation through eqs. (253), (254). Ψ_n is like a velocity potential for the shift, and, as we shall see, only the combination Ψ_σ , which is a potential for the anisotropic shear of the hypersurfaces σ_j^i , enters into the equations of motion. The anisotropic stress for type-matter can also be expressed in terms of a scalar potential, $\pi_{t,\text{type}}$. It vanishes for scalar fields, as eq. (237) shows, and also for perfect fluids, including CDM and the baryons. It does not vanish for photons and relativistic neutrinos. The acceleration of a fluid moving with velocity U is $\mathcal{A}_{(U)I} = \bar{a}^{-1} U^n \bar{e}_n[\bar{a}U_I] + e_I[\ln N]$ to first order, and is $e_I[\ln N]$ to all orders for the time surfaces (as is shown in Appendix B), yielding eq. (251) expressed in terms of an acceleration potential. The acceleration, $\nabla_U U$, is from nongravitational forces only, hence the $+{}^{(3)}\bar{\nabla}_I \nu$ term is there to take out the gravitational acceleration derived from geodesic motion.

The expansion of ${}^{(3)}g_{ij}$ is based on the removal of the 3-space conformal factor $\bar{a}$ rather than some inhomogeneous function. We define $\bar{\nabla}^j$ in terms of the ${}^{(3)}\bar{g}_{ij}$ without the $\bar{a}^2$ taken out, so that ${}^{(3)}\bar{\nabla}^i = \bar{a}^{-2} {}^{(3)}g_*^{ij} \partial_j$ (with ${}^{(3)}g_*^{ij} = \delta^{ij}$ for a flat Universe) has extra $\bar{a}$ terms designed to confuse the reader. So does the Laplacian ${}^{(3)}\bar{\nabla}^2$. One of the advantages in working in an orthonormal basis is that the correct $\bar{a}$ multipliers enter into the expressions (e.g., ${}^{(3)}\bar{\nabla}_I = \bar{a}^{-1} \partial/\partial x^I$ for a flat Universe).

The energy constraint and (the first integral of) the momentum constraint

are

$$2\bar{H}(\delta H) - \frac{2}{3} {}^{(3)}\bar{\nabla}^2 \varphi - \frac{1}{6} {}^{(3)}\bar{R} 2\varphi = \frac{8\pi G_N}{3} (\delta\rho)_{\text{tot}}, \quad (255)$$

$$\begin{aligned} (\delta H) + \frac{1}{3} {}^{(3)}\bar{\nabla}^2 \Psi_\sigma &= \frac{1}{N} \dot{\varphi} - \bar{H}\nu \\ &= -4\pi G_N \sum_{\text{type}} (\bar{\rho} + \bar{p})_{\text{type}} \Psi_{v,\text{type}} - \frac{1}{6} {}^{(3)}\bar{R} \Psi_\sigma. \end{aligned} \quad (256)$$

It is sometimes better to work with a modified form of the energy constraint equation, found by inserting the relation for (δH) from the momentum constraint equation into eq. (255):

$$\begin{aligned} - {}^{(3)}\bar{\nabla}^2 (\varphi + \bar{H}\Psi_\sigma) - \frac{1}{4} {}^{(3)}\bar{R} 2(\varphi + \bar{H}\Psi_\sigma) &= 4\pi G_N (\delta\rho)_{\text{com,tot}}, \quad (257) \\ (\delta\rho)_{\text{com,tot}} &\equiv ((\delta\rho)_{\text{tot}} + 3\bar{H}(\bar{\rho} + \bar{p})_{\text{tot}} \Psi_{v,\text{tot}}), \quad \Phi_H \equiv \varphi + \bar{H}\Psi_\sigma, \end{aligned}$$

involving the energy density in the frame in which the total energy current $J_{(e),\text{tot}}$ vanishes and Bardeen's gauge invariant Φ_H , which is also φ in the longitudinal gauge: $\varphi_L = \Phi_H$.

The Raychaudhuri equation, slightly reworked, is

$$\begin{aligned} \frac{1}{\bar{N}^2} \left[\frac{\partial \bar{N}(\delta H)}{\partial \tau} + \frac{\partial \ln(\bar{N}^{-1} \bar{a}^2)}{\partial \tau} \bar{N}(\delta H) - \frac{\partial \bar{H}}{\partial \tau} \bar{N}\nu - \frac{1}{3} \bar{N}^2 {}^{(3)}\bar{\nabla}^2 \nu \right] \\ = -\frac{4\pi G_N}{3} ((\delta\rho) + 3(\delta p))_{\text{tot}}. \end{aligned} \quad (258)$$

Note that $\bar{N}(\delta H)$ is negative: a growing density perturbation slows the expansion rate. The $(G')_J^I$ simplifies considerably when expressed in terms of the potentials:

$$\frac{1}{\bar{N}} \dot{\Psi}_\sigma + \bar{H}\Psi_\sigma + (\varphi + \nu) = -8\pi G_N \sum_{\text{type}} \bar{p}_{\text{type}} \pi_{t,\text{type}}. \quad (259)$$

Although for scalar perturbations, the constraint equations together with the matter conservation equations form a complete system from which the dynamical Einstein equations follow by taking appropriate time derivatives and linear combinations, sometimes it is worth it to solve the Raychaudhuri equation, extra time derivative and all, or the anisotropic G_{IJ} equation in the place of one of the constraint equations. In a gauge with $\nu=0$, the Raychaudhuri equation becomes a simple ODE for $\bar{N}(\delta H)$ at each point in the space. The momentum constraint equation is an ODE for φ , but it turns out that only $\dot{\varphi}$ enters the matter evolution equations and its expression in terms of the velocity potentials can be substituted. This is the usual approach taken for solving scalar perturbations in the synchronous gauge, and is the one adopted in the Bond and Szalay and Bond and Efstathiou papers [195, 134, 88].

Although it is fine to solve eq. (258) for the evolution of matter and radiation through photon decoupling and free-streaming to the present, intractable numerical problems arise in inflation calculations with scalar fields [192]: a robust solution strategy for solving synchronous gauge fluctuations does exist: the momentum constraint is treated as an ODE for φ , and (δH) is then fixed through the energy constraint equation.

For the synchronous gauge, the anisotropic G_{IJ} equation follows from a combination of the matter evolution equations and the other Einstein equations and is not usually separately solved for. It is an algebraic relation for zero shear hypersurfaces ($\Psi_\sigma = 0$), e.g., for the longitudinal gauge (with $\psi = \Psi_n = 0$). For example, if there is no anisotropic stress (e.g., universes with only perfect fluids and/or scalar fields), then $\nu_L = -\varphi_L$. In [192], we also solved for scalar field fluctuations in the longitudinal gauge, using eq. (259) and a sum of the Raychaudhuri and energy constraint equations, a dynamical equation of second order in φ_L . For the CMB problem, the standard approach [138] has been to also use a constraint equation, the Poisson equation, eq. (257), relating the total comoving energy density to $^{(3)}\bar{\nabla}^2 \nu_L$. [260] use the momentum constraint equation instead of the anisotropic shear equation.

Under scalar mode gauge transformations [171, 178], $\tau_{\text{new}} = \tau_{\text{old}} + T$, $x_{\text{new}}^i = x_{\text{old}}^i + \bar{a}^2 {}^{(3)}\bar{\nabla}^i L$, where T and L are scalar functions, we have

$$\begin{aligned}
\nu_{\text{new}} &= \nu_{\text{old}} - \frac{1}{\bar{N}} \frac{\partial \bar{N} T}{\partial \tau}, & \varphi_{\text{new}} &= \varphi_{\text{old}} - \bar{H} \bar{N} T, \\
\psi_{\text{new}} &= \psi_{\text{old}} + L, & \Psi_{n\text{new}} &= \Psi_{n\text{old}} - \frac{\bar{a}^2}{\bar{N}} \dot{L} + \bar{N} T \\
\Psi_{\sigma,\text{new}} &= \Psi_{\sigma,\text{old}} + \bar{N} T, & \Psi_{v,\text{type,new}} &= \Psi_{v,\text{type,old}} - \bar{N} T, \\
(\delta H)_{\text{new}} &= (\delta H)_{\text{old}} - \dot{\bar{H}} T - \frac{1}{3} {}^{(3)}\bar{\nabla}^2 \bar{N} T, \\
(\delta H)_{*\text{new}} &= (\delta H)_{*\text{old}} + (1 + \bar{q}) \bar{H}^2 \bar{N} T, \\
(\delta \rho)_{\text{tot,new}} &= (\delta \rho)_{\text{tot,old}} + 3 \bar{H} (\bar{\rho} + \bar{p})_{\text{tot}} \bar{N} T, \\
(\delta \rho)_{\text{type,new}} &= (\delta \rho)_{\text{type,old}} - \dot{\bar{\rho}}_{\text{type}} T, \\
(\delta p)_{\text{type,new}} &= (\delta p)_{\text{type,old}} - \dot{\bar{p}}_{\text{type}} T, & \pi_{t,\text{new}} &= \pi_{t,\text{old}}, \\
\Psi_{\mathcal{A},\text{type,new}} &= \Psi_{\mathcal{A},\text{type,old}}, \\
\text{scalar field: } (\delta \phi)_{\text{new}} &= (\delta \phi)_{\text{old}} + \bar{e}_n[\phi] \bar{N} T.
\end{aligned} \tag{260}$$

The modified inhomogeneous Hubble parameter H_* is defined by eq. (167). Notice that the “acceleration potential” of a fluid is gauge invariant. The unperturbed momentum of the scalar field is $\bar{e}_n[\phi]$.

To transform from the synchronous to the longitudinal gauge:

$$\begin{aligned}
\bar{N} T &= -\Psi_{\sigma,S}, & \Phi_A \equiv \nu_L &= \frac{1}{\bar{N}} \dot{\Psi}_{\sigma,S}, & \Phi_H \equiv \varphi_L &= \varphi_S + \bar{H} \Psi_{\sigma,S}, \\
\Psi_{v,\text{type},L} &= \Psi_{v,\text{type},S} + \Psi_{\sigma,S}, & \Psi_{v,\text{cdm},L} &= \Psi_{\sigma,S}, \\
\delta_{\text{type},L} &= \delta_{\text{type},S} + \frac{d \ln \bar{\rho}_{\text{type}}}{d \ln a} \bar{H} \Psi_{\sigma,S},
\end{aligned} \tag{261}$$

$$\text{scalar field: } (\delta\phi)_L = (\delta\phi)_S - \bar{e}_n[\bar{\phi}]\Psi_{\sigma,S}. \quad (262)$$

Φ_A and Φ_H are gauge invariant. Some other gauge invariant quantities that are often used are:

$$\zeta = \varphi + \frac{(\delta\rho)_{\text{tot}}}{3(\bar{\rho} + \bar{p})_{\text{tot}}}, \quad (263)$$

$$\varphi_{\text{com}} = \varphi - \bar{H}\Psi_{v,\text{tot}}, \quad (264)$$

$$(\delta\rho)_{\text{com,type}} = (\delta\rho)_{\text{type}} - \frac{1}{\bar{N}}\dot{\bar{\rho}}_{\text{type}}\Psi_{v,\text{type}}, \quad (265)$$

$$(\delta p)_{\text{type}} - \frac{\dot{\bar{p}}_{\text{type}}}{\dot{\bar{\rho}}_{\text{type}}}(\delta\rho)_{\text{type}}. \quad (266)$$

Also gauge invariant are any differences between quantities which may themselves not be gauge invariant, such as velocity and appropriately normalized density differences:

$$\Psi_{v,\text{type}_1,\text{type}_2} \equiv \Psi_{v,\text{type}_1} - \Psi_{v,\text{type}_2}, \quad (267)$$

$$\frac{(\delta\rho)_{\text{type}_1}}{(\bar{\rho} + \bar{p})_{\text{type}_1}} - \frac{(\delta\rho)_{\text{type}_2}}{(\bar{\rho} + \bar{p})_{\text{type}_2}}. \quad (268)$$

Examples used below are the relative photon–baryon velocity potential, $\Psi_{v,\gamma B}$ and photon entropy per baryon perturbation, $\delta_{s_\gamma} = \frac{3}{4}\delta_\gamma - \delta_B$.

A.3 Tensor perturbations

For tensor perturbations, we have

$$\begin{aligned} g_{ij} &= {}^{(3)}g_{ij} = \bar{e}_i^I \bar{e}_j^J (\delta_{IJ} + h_{ij}^{(TT)}), \quad g_{00} = -\bar{N}^2, \quad g_{0i} = N_i = 0, \\ (\delta H) &= 0, \quad \sigma_j^i = -(K')_j^i = -\frac{1}{2\bar{N}}\dot{h}_{kj}^{(TT)}\delta^{ik}, \\ (\delta {}^{(3)}R')_j^i &= -{}^{(3)}\bar{\nabla}^2 \frac{1}{2}h_i^{(TT)j} + {}^{(3)}\bar{R} h_i^{(TT)j}, \quad \delta {}^{(3)}R = 0. \end{aligned} \quad (269)$$

The tensor mode is already gauge invariant. Only the anisotropic dynamical Einstein equations are needed: multiplying both sides by $2\bar{N}^2$ gives

$$\begin{aligned} \ddot{h}_i^{(TT)j} + \frac{\partial \ln(\bar{a}^3/\bar{N})}{\partial \tau} \dot{h}_i^{(TT)j} - \bar{N}^2 {}^{(3)}\bar{\nabla}^2 h_i^{(TT)j} + \frac{1}{3}\bar{N}^2 {}^{(3)}\bar{R} h_i^{(TT)j} \\ = 16\pi G_N \bar{N}^2 (\Pi_{\text{tot}})_i^j. \end{aligned} \quad (270)$$

With scalar fields only, there is no anisotropic stress, hence the gravity waves are freely propagating. Of course they can still be generated by quantum noise in the $h_{ij}^{(TT)}$ field. Anisotropic stresses from neutrinos and photons can lead to gravitational wave generation, but this is a very small effect. Cosmic strings decay by emitting gravitational waves, generated in response to their anisotropic stress.

B Transport theory in General Relativity

B.1 The distribution function and the BTE in GR

The theoretical framework used to calculate the anisotropies and distortions of the CMB is general relativistic polarized photon transport theory. Kinetic theory in general relativity was actively developed in the late sixties and early seventies (e.g., Ehlers 1971 and Stewart 1971). For the cosmological transport problem, we need a set of Boltzmann transport equations for single particle distribution functions. Due to the nonlocalizability of position and momentum, one must be careful in defining the distribution function. For flat cosmologies, the eigenmodes are plane waves, momenta are Fourier transform variables conjugate to positions, and a Wigner distribution function can be defined (in terms of a two-particle equal-time propagator). If the particles have spin (or polarization) labelled by s , then the Wigner distribution function is a matrix in spin space:

$$f_{s's}(\mathbf{q}, \mathbf{x}, \tau) \equiv \sum_{\mathbf{k}} e^{i\mathbf{k} \cdot \mathbf{x}} \langle a_{s', \mathbf{q}-\mathbf{k}/2}^\dagger(\tau) a_{s, \mathbf{q}+\mathbf{k}/2}(\tau) \rangle, \quad (271)$$

where $\langle \dots \rangle$ denotes a (nonequilibrium) ensemble average, the operator $a_{s, \mathbf{q}+\mathbf{k}/2}$ annihilates a particle with spin s of momentum $\mathbf{q} + \mathbf{k}/2$, and $a_{s', \mathbf{q}-\mathbf{k}/2}^\dagger$ creates a particle with spin s' of momentum $\mathbf{q} - \mathbf{k}/2$. The trace of $f_{s's}$ in spin space, $f_t(q^i, x^\alpha) = \frac{1}{2} \sum_s f_{ss}$, is the mean occupation number of the state of momentum q^i in the neighborhood of the spacetime point x^α . f_t defined this way is not a positive definite quantity and so the interpretation of f_t as phase space density is invalid.¹

Coherent effects – such as the modification of the photon propagator by collective plasma effects – must be taken into account by appropriately de-

¹A natural way to get a positive definite distribution is to discretize phase space into cells of size $(2\pi\hbar)^3$ and centers $(\mathbf{X}, \mathbf{Q})$. The uncertainty principle implies further localization within a cell is not possible. The photon field can be expanded in annihilation and creation operators $a_{s\mathbf{X}\mathbf{Q}}, a_{s\mathbf{X}\mathbf{Q}}^\dagger$, with associated wave functions $\langle \mathbf{x} | \mathbf{X}, \mathbf{Q} \rangle$ which are zero outside of the spatial part of the box and which are box-normalized plane waves, $\exp(i\mathbf{Q} \cdot \mathbf{x})$, inside. These form a complete orthonormal set. The relative degree to which the boxes are spatially elongated is at our disposal provided the quantum volume constraint is maintained. By shrinking the spatial directions at the expense of increasing the separation between wavenumbers one recovers the delta function wavefunctions of the position space representation of quantum mechanics; by shrinking the momentum directions one approaches the continuum plane wavefunctions of the momentum space representation. The occupation indices of each such fundamental phase space cell can be used to define the distribution function: $f_{ss'}(\mathbf{Q}, \mathbf{X}, \tau) = \langle a_{s'\mathbf{X}\mathbf{Q}}^\dagger(\tau) a_{s\mathbf{X}\mathbf{Q}}(\tau) \rangle$. Compared with the usual Wigner distribution, there are disadvantages (does not have a continuous dependence on position and momentum, boundary terms involving transport from one box to another are complicated) and advantages (emphasizes the fundamental graininess imposed by quantum mechanics on phase space, and coarse-graining of phase space only involves making the boxes of much larger volume than that required by the Heisenberg uncertainty principle). Both approaches give exact quantum evolution equations which reduce to the usual form of the Boltzmann transport equation in the classical limit.

finer quasiparticles which have these collective interactions included, but this is not of importance for the $\Delta T/T$ problem. In the classical limit – when spatial inhomogeneities of f_t and gravitational field curvature are both of long wavelength compared with the typical de Broglie wavelength of the particles, q^{-1} – localizability is a good approximation, f_t is positive definite, and the quantum evolution equation for f_t reduces to a Boltzmann transport equation. The transport model considers the particles propagating along geodesics in spacetime. The particles may undergo absorptions or emissions or scatterings at single points. For such a description of collisions to be valid it is also necessary that the interaction regime be small in spatial and temporal extent compared with the scale of inhomogeneity in $f_{s's}$. Also, in order for the equations to be closed off at the single-particle distribution level (rather than requiring e.g., a full Liouville equation or higher moments in a BBGKY hierarchy), the only correlations allowed to be explicitly included are those due to the particle statistics, Bose–Einstein or Fermi–Dirac. The equation for the evolution of the distribution function is

$$q^\alpha \frac{\partial f_{s_2 s_1}}{\partial x^\alpha} \Big|_q - \Gamma_{\alpha\beta}^i q^\alpha q^\beta \frac{\partial f_{s_2 s_1}}{\partial q^i} \Big|_x = q^0 \mathcal{S}_{s_2 s_1}[\mathbf{f}], \quad (272)$$

where $\mathcal{S}_{s_2 s_1}[f]$ is the source function. $f_{s's}$ is a general relativistic scalar under coordinate transformations of the position coordinates and of the momentum coordinates. $q^0 \mathcal{S}_{s_2 s_1}[f]$ also transforms as a general relativistic scalar, which conveniently allows transformation from one gauge to another.

Although eq. (272) is not manifestly covariant because the summation in the second term runs over spatial indices only, it is actually covariant – a consequence of the momentum being constrained to lie on the mass shell, $q^\alpha g_{\alpha\beta} q^\beta = -m^2$. Any other 3 parameters labelling the mass shell instead of the coordinate momenta q^i would also do. Thus, for the transport problem selecting a gauge involves choosing a spacetime coordinate system and a momentum space coordinate system, and these can be chosen relatively independently of each other if we wish. Coordinate momenta q^i are generally not very physically meaningful. It is usually better to use spatial momentum components relative to an orthonormal tetrad $e_\alpha^\mu(x^\mu)$: $q^I = e_\alpha^I q_{coord}^\alpha$, often along with further momentum-gauge transformations beyond this to simplify analytics or numerics, in particular one that makes the momentum a comoving one [195]. The transfer equation in the triad momentum variables¹ looks similar to eq. (272), and is easily obtained by appropriately transforming it:

$$q^a e_a[f] - \Gamma_{ab}^I q^a q^b \frac{\partial f}{\partial q^I} \Big|_x = q^n \mathcal{S}[f], \quad \text{where } e_a[f] \equiv e_a^\alpha \frac{\partial f}{\partial x^\alpha} \Big|_q, \quad (273)$$

and the Γ_{ab}^I are connection coefficients relative to the tetrad e_a , defined by the expansion $\nabla_{e_b} e_a = \Gamma_{ab}^c e_c$. To reduce this to a usable but quite general form, we

¹If the geodesic motion is $x^\alpha(\lambda)$, $q^I(\lambda)$, where λ is an affine parameter, then the geodesic equations $q^a = e_\alpha^a dx^\alpha/d\lambda$, $dq^I/d\lambda = -\Gamma_{ab}^I q^a q^b$ applied to $Df/d\lambda = (Df/d\lambda)_{coll}$, where $(Df/d\lambda)_{coll}$ describes the change in the distribution function as a result of local interactions, yields the transport equation.

use a little more of the machinery of differential geometry. The Γ_{ab}^c are often termed Ricci rotation coefficients (and denoted by ω_{ab}^c), and are related to the structure coefficients C_{ab}^c of the basis,²

$$[e_a, e_b] \equiv C_{ab}^c e_c, \quad \text{by } \Gamma_{ab}^c = -\frac{1}{2} g^{cd} \{e_d[g_{ab}] - e_b[g_{da}] - e_a[g_{db}]\} \\ - \frac{1}{2} [C_{ab}^c - g_{db} g^{cf} C_{fa}^d - g_{ad} g^{cf} C_{fb}^d]. \quad (274)$$

Let us first introduce an orthonormal basis, e_n, e_I , where $e_a^\alpha g_{\alpha\beta} e_a^\beta = \eta_{ab} = \text{diag}(-1, 1, 1, 1)$, so Γ_{ab}^c only involves the C_{ab}^c . If p^a is the momentum in the e_a basis, then $p^a p^b \Gamma_{ab}^I \frac{\partial f}{\partial p^I} = p^a p^b \eta_{ad} \delta^{IJ} C_{Jb}^d \frac{\partial f}{\partial p^I}$, often easier to calculate. The momentum $p \equiv (p^I \delta_{IJ} p^J)^{1/2}$ will redshift as the universe expands, so it is not the momentum we finally wish to work with. Since we have seen that the equations greatly simplify with the introduction of a comoving momentum, we introduce $q^I = \Omega p^I$, hence $q = \Omega p$, $\hat{q}^I = \hat{p}^I$ and $q^n = \sqrt{q^2 + m^2 \Omega^2}$, where the function $\Omega(\mathbf{x}, \tau)$ is at our disposal, except that it should be $\bar{a}(\tau)$ for the unperturbed case. The transformation of the action of the vector e_a on f from the space in which p is fixed to the space in which q is fixed is simply shown to be

$$e_a|_p[f_t] = e_a|_q[f_t] + e_a[\ln \Omega] q \partial f_t / \partial q.$$

Note that $q \partial f_t / \partial q = q \hat{q}^I \partial f_t / \partial q^I$. For the basis e_n, e_I , we have

$$C_{nI}^n = -C_{In}^n = e_I[\ln N], \quad C_{IJ}^n = 0, \\ C_{nI}^J = -C_{In}^J = e_I^J \left\{ e_n[e_I^i] + \frac{1}{N} e_I[N^i] \right\}, \\ C_{IJ}^K = e_I[e_J^j] e_j^K - e_J[e_I^j] e_j^K.$$

An example of the use of this is the computation of the acceleration 4-vector of the timelike hypersurfaces: $\mathcal{A}_n$ vanishes and $\mathcal{A}^I = \Gamma_{nn}^I = {}^{(3)}g^{IJ} C_{Jn}^n = {}^{(3)}g^{IJ} e_J[\ln N]$.

As in section A.1, we use the 3-space “conformally transformed” basis, $e_*^I = A^{-1} e^I$, $e_{*I} = A e_I$, with $A(\mathbf{x}, \tau)$ reducing to $\bar{a}(\tau)$ for the unperturbed case, but possibly inhomogeneous in the fluctuation case. This means the

²The commutator of the differential operator $[e_a, e_b]$ is defined by its action on a function f : $[e_a, e_b](f) = (e_a^\alpha e_{b,\alpha}^\beta - e_b^\alpha e_{a,\alpha}^\beta) \partial f / \partial x^\beta$. To define the sign conventions I use here, the curvature tensor is $R(e_a, e_b, e_c, e^d) = \langle e^d, R(e_a, e_b) e_c \rangle = R_{cab}^d$, hence $R_{cab}^d = e_a[\Gamma_{cb}^d] - e_b[\Gamma_{ca}^d] + \Gamma_{cb}^f \Gamma_{fa}^d - \Gamma_{ca}^f \Gamma_{fb}^d - C_{ab}^f \Gamma_{cf}^d$ where $C_{ab}^c = \Gamma_{ba}^c - \Gamma_{ab}^c$. Here the operator $R(e_a, e_b) = {}^{(4)}\nabla_{e_a} {}^{(4)}\nabla_{e_b} - {}^{(4)}\nabla_{e_b} {}^{(4)}\nabla_{e_a} + {}^{(4)}\nabla_{[e_a, e_b]}$. The Ricci tensor, scalar, Einstein tensor and Einstein equations are: $R_{cb} = R_{cab}^a$, $R = g^{cb} R_{cb}$, $G_{cb} = R_{cb} - \frac{1}{2} R g_{cb}$, $G_{cb} = 8\pi G_N T_{cb}$. The connection and curvature forms are $\omega_b^c \equiv \Gamma_{ba}^c e^a$, $\theta_c^d = d\omega_c^d + \omega_a^d \wedge \omega_c^a$, obeying $dg_{ab} = \omega_{ab} + \omega_{ba}$, where $\omega_{ab} \equiv g^{ac} \omega_b^c$, and the first and second Cartan equations, $de^a + \omega_b^a \wedge e^b = 0$ (or $\frac{1}{2}$ torsion) and $\theta_c^d = \frac{1}{2} R_{cab}^d e^a \wedge e^b$. Here $\wedge$ is the exterior product and d is the exterior derivative of forms. The latter 3 equations are all that is needed to compute connection coefficients and the curvature tensor for a metric in any basis, and is usually simpler than using direct Γ_{ab}^c calculation in a given basis.

momentum is $q^a = \Omega A \langle e_*^a, p \rangle$. As we shall see, it turns out to be most desirable to have $\Omega = A$ to ensure that there are no terms representing the redshifting of the radiation for the unperturbed background. It is this q^a and its q which is the inhomogeneous generalization of the comoving momentum introduced in section 3.1. Because of the flexibility in the spatial dependence of A , it is not unique. The way the basis change manifests itself is through $e_i^K e_n[e_J^i] = e_{*i}^K e_n[e_{*J}^i] - \delta_J^K e_n[\ln A]$. In terms of q and e_{*J} , the transport equation becomes

$$\begin{aligned} & \frac{q^a}{q^n} e_a[f_t] - q \frac{\partial f_t}{\partial q^I} \delta^{IJ} \left[\hat{q}_J e_n[\ln(A/\Omega)] + \frac{q^n}{q} \frac{1}{A} e_{*J}[\ln N] \right. \\ & \quad - \frac{q}{q^n} \frac{1}{A} \hat{q}_J \hat{q}^K e_{*K}[\ln \Omega] - \hat{q}_K e_{*i}^K \{ e_n[e_{*J}^i] + \frac{1}{N} e_{*J}[N^i] \} \\ & \quad \left. + \frac{q}{q^n} \hat{q}_K \hat{q}^M \frac{1}{A} e_{*i}^K \{ e_{*J}[e_{*M}^i] - e_{*M}[e_{*J}^i] \} \right] \\ & \quad + q \frac{\partial f_t}{\partial q^I} \frac{q}{q^n} \frac{1}{A} (\delta^{IJ} - \hat{q}^I \hat{q}^J) e_{*J}[\ln A] = \mathcal{S}[f_t]. \end{aligned} \quad (275)$$

B.2 Number, energy and momentum conservation equations

A first application of this equation is to derive the energy and momentum conservation equations for “type”-matter. Just as the stress energy tensor can be decomposed into $(\rho, J_{(e)}^I, p, \Pi_{IJ})_{\text{type}}$, so the number current 4-vector $\mathcal{J}_{\text{type}}^a$ of particles of a given type with respect to a flow U^a can be decomposed as $\mathcal{J}_{n_{\text{type}}}^a = n_{\text{type}} U^a + J_{(n_{\text{type}})}^a$, where $\perp_{ba} J_{(n_{\text{type}})}^a = 0$. Similarly the “type”-entropy 4-vector $\mathcal{J}_{(s_{\text{type}})}^a$ can be decomposed as $\mathcal{J}_{(s_{\text{type}})}^a = s_{(\text{type})} U^a + J_{(s_{\text{type}})}^a$. (In the comoving frame of an imperfect fluid moving with velocity U , $\mathcal{J}_{(s_{\text{type}})}^I = T^{-1} \mathcal{J}_{(e)_{\text{type}}}^I$, given by eq. (236).) If U is the flow of time, e_n^α , and we use the basis e_I , these various densities and currents are related to the distribution function by

$$\begin{aligned} \rho &= \sum_{\mathbf{qs}} \Omega^{-4} q^n f_t, \quad p = \frac{1}{3} \sum_{\mathbf{qs}} \Omega^{-4} \frac{q^2}{q^n} f_t, \\ n_{\text{type}} &= \sum_{\mathbf{qs}} \Omega^{-3} q^n f_t, \\ s &= - \sum_{\mathbf{qs}} \Omega^{-3} \{ f_t \ln f_t - (\pm)(1 \pm f_t) \ln(1 \pm f_t) \}, \\ J_{(e)}^I &= \sum_{\mathbf{qs}} \Omega^{-4} q^I f_t, \quad J_{(n)}^I = \sum_{\mathbf{qs}} \Omega^{-3} \frac{q^I}{q^n} f_t, \\ J_{(s)}^I &= - \sum_{\mathbf{qs}} \Omega^{-3} \frac{q^I}{q^n} \{ f_t \ln f_t - (\pm)(1 \pm f_t) \ln(1 \pm f_t) \}, \end{aligned} \quad (276)$$

$$\Pi^{IJ} = \sum_{\mathbf{qs}} \Omega^{-4} (\hat{q}^I \hat{q}^J - \frac{1}{3} \delta^{IJ}) \frac{q^2}{q^n} f, \quad \sum_{\mathbf{qs}} (\dots) \equiv \sum_{spin} \int \frac{d^3 \mathbf{q}}{(2\pi)^3} (\dots).$$

Here (+) is for bosons, (−) is for fermions. A sum over spins (or polarizations) is needed because of the way f_t has been defined. (For a general basis, the form for $\mathcal{J}_{n_{\text{type}}}^a$ involves $\Omega^{-3} (-^{(4)}g)^{1/2} q^a / (-q_n)$ and for T_{type}^{ab} involves $\Omega^{-4} (-^{(4)}g)^{1/2} q^a q^b / (-q_n)$, where q_n is the covariant time component of the momentum.)

Consider the limit of the BTE eq. (275) for nr -matter, for which $q^n \rightarrow m\Omega$. We take $\Omega = A = \bar{a}$ and ignore q/q^n terms. The BTE and the zeroth and first order moment equations w.r.t. $\mathbf{q}$ which give mass and momentum conservation are then:

$$\begin{aligned} e_n[f_t] + \frac{q^I}{m\bar{a}} e_I[f_t] - m e_{*J} [\ln N] \frac{\partial f_t}{\partial q^I} + e_{*i}^K \{e_n[e_{*J}^i] \\ + \frac{1}{N} e_{*J} [N^i]\} q_K \delta^{IJ} \frac{\partial f}{\partial q^I} = \mathcal{S}[f_t], \\ e_n[\rho \Omega^3] + \frac{1}{\bar{a}^2} e_{*I} [J_{(e)}^I \Omega^4] - \rho \Omega^3 e_{*i}^J \{e_n[e_{*J}^i] + \frac{1}{N} e_{*J} [N^i]\} = m \sum_{\mathbf{qs}} \mathcal{S}, \\ e_n[J_{(e)}^I \Omega^4] + \bar{a}^{-2} e_{*J} [(p \delta^{IJ} + \Pi^{IJ}) \Omega^5] + \rho \Omega^3 \delta^{IJ} e_{*J} [\ln N] - \Omega^4 (J_{(e)}^I e_{*i}^J \\ + J_{(e)}^K g_{KM} e_{*i}^M \delta^{IJ}) \{e_n[e_{*J}^i] + \frac{1}{N} e_{*J} [N^i]\} = \sum_{\mathbf{qs}} q^I \mathcal{S}. \end{aligned} \quad (277)$$

The nr -transport equation does indeed take the form of a Boltzmann equation with a gravitational force $-m \nabla \ln N$, with $\ln N$ the gravitational potential perturbation. The last term in eqs. (277) is a shearing term related to the extrinsic curvature. The general equation for semi-relativistic matter can also be obtained this way, but it is more easily derived from the ADM formulation of the energy and momentum conservation laws for type-matter:

$$\begin{aligned} e_n[\rho] - K(\rho + p) - (K')^{IJ} \Pi_{IJ} + {}^{(3)}\nabla_I J_{(e)}^I \\ + 2J_{(e)}^I {}^{(3)}\nabla_I \ln N = \sum_{\mathbf{ps}} p^n \mathcal{S}, \\ e_n[J_{(e)}^I] - K J_{(e)}^I - 2(K')^{IJ} \Pi_{IJ} + {}^{(3)}\nabla_J (p {}^{(3)}g^{IJ} + \Pi^{IJ}) \\ + ((\rho + p) {}^{(3)}g^{IJ} + \Pi^{IJ}) {}^{(3)}\nabla_J \ln N + J_{(e)}^J {}^{(3)}\nabla_J N^I = \sum_{\mathbf{ps}} p^I \mathcal{S}. \end{aligned}$$

The physical interpretation of the different terms is clear.

The perturbed energy and momentum conservation equations for nr -type particles follow from eqs. (277). More generally, we shall keep in the terms of order $\bar{p}/\bar{\rho}$ to have a generally valid result:

$$\frac{\bar{\rho}_{\text{type}}}{3(\bar{\rho} + \bar{p})_{\text{type}}} \bar{e}_n[\delta_{\text{type}}] + 3\bar{e}_n[\varphi] - {}^{(3)}\bar{\nabla}^2 (\Psi_{v,\text{type}} + \Psi_\sigma)$$

$$+ \bar{H} \frac{(\delta p - \bar{p}\delta)_{\text{type}}}{(\bar{\rho} + \bar{p})_{\text{type}}} = \frac{\sum_{\mathbf{p}s} p^n \mathcal{S}_{\text{type}} - p^n \bar{\mathcal{S}}_{\text{type}} (1 + \delta_{\text{type}})}{(\bar{\rho} + \bar{p})_{\text{type}}}, \quad (278)$$

$$\begin{aligned} & \frac{\bar{\rho}_{\text{type}}}{(\bar{\rho} + \bar{p})_{\text{type}}} e_n \left[\left(1 + \frac{\bar{p}_{\text{type}}}{\bar{\rho}_{\text{type}}} \right) \Psi_{v,\text{type}} \right] - 3 \bar{H} \frac{\bar{p}_{\text{type}}}{\bar{\rho}_{\text{type}}} \Psi_{v,\text{type}} \\ &= \nu + \frac{(\delta p)_{\text{type}}}{(\bar{\rho} + \bar{p})_{\text{type}}} + \frac{\left(\frac{2}{3} {}^{(3)}\bar{\nabla}^2 + \frac{1}{3} {}^{(3)}\bar{R} \right) \bar{p}_{\text{type}} \pi_{t,\text{type}}}{(\bar{\rho} + \bar{p})_{\text{type}}} \\ & - \frac{{}^{(3)}\bar{\nabla}^{-2} {}^{(3)}\bar{\nabla}_I \sum_{\mathbf{p}s} p^n \left(\mathcal{S}_{\text{type}} \frac{p^I}{p^n} + \bar{\mathcal{S}}_{\text{type}} {}^{(3)}\bar{\nabla}^I \Psi_{v,\text{type}} \right)}{(\bar{\rho} + \bar{p})_{\text{type}}}. \quad (279) \end{aligned}$$

The perturbed energy conservation equations for the total energy and momentum are the same, except that the sums over sources $\mathcal{S}_{\text{tot}}$ vanish: the total energy and momentum are conserved.

The nonrelativistic limit of eq. (279) is $\bar{p}_{\text{type}}/\bar{\rho}_{\text{type}} \rightarrow 0$, $p^n \rightarrow m_{nr}$. The energy equation, eq. (278), is handled by writing $\rho_{nr} = n_{nr}(m_{nr} + \epsilon_{nr})$, where m_{nr} , n_{nr} and ϵ_{nr} are the mass, number density and thermal energy per nr -type particle. Terms of zeroth order in m_{nr}^{-1} give the number conservation equation and terms of first order give the thermal energy conservation law, which is just $d\epsilon_{nr} + p_{nr}dn_{nr}^{-1} = T_{nr}ds_{nr}$, where ds_{nr} denotes the entropy generation in the nr -matter in time $d\tau$, T_{nr} the temperature. These laws are:

$$\bar{e}_n[\delta_{n,nr}] + 3\bar{e}_n[\varphi] - {}^{(3)}\bar{\nabla}^2(\Psi_{v,nr} + \Psi_\sigma) = \bar{n}_{nr}^{-1} \sum_{\mathbf{p}s} (\delta\mathcal{S}_{nr} - \bar{\mathcal{S}}_{nr}\delta_{n,nr}), \quad (280)$$

$$\begin{aligned} & \bar{e}_n[\delta\epsilon_{nr}] - \bar{p}_{nr}\bar{e}_n[\delta_{n,nr}] + 3\bar{H}\bar{n}_{nr}^{-1}((\delta p)_{nr} - \bar{p}_{nr}\delta_{n,nr}) \\ &= \bar{n}_{nr}^{-1} \sum_{\mathbf{p}s} \frac{p^2}{2m_{nr}} (\delta\mathcal{S}_{nr} - \bar{\mathcal{S}}_{nr}\delta_{n,nr}). \quad (281) \end{aligned}$$

$$\begin{aligned} & \bar{e}_n[\Psi_{v,nr}] = \nu + \frac{(\delta p)_{nr}}{m_{nr}\bar{n}_{nr}} + \frac{\left(\frac{2}{3} {}^{(3)}\bar{\nabla}^2 + \frac{1}{3} {}^{(3)}\bar{R} \right) \bar{p}_{nr} \pi_{t,nr}}{m_{nr}\bar{n}_{nr}} \\ & - \bar{n}_{nr}^{-1} {}^{(3)}\bar{\nabla}^{-2} {}^{(3)}\bar{\nabla}_I \sum_{\mathbf{p}s} \left[\frac{p^I}{m_{nr}} \delta\mathcal{S}_{nr} + \left(\frac{p^I}{m_{nr}} + {}^{(3)}\bar{\nabla}^I \Psi_{v,nr} \right) \bar{\mathcal{S}}_{nr} \right]. \quad (282) \end{aligned}$$

For cold dark matter we only need eq. (280) and eq. (282), and this is all we need for baryons as well if the baryonic pressure and heating can be neglected, which is the case if we only wish to follow the development of primary anisotropies.

B.3 The transport of extremely relativistic particles

We now turn back to the transport equation to apply it to radiative transfer. Instead of using q^I , we shall change to $q, \hat{q}^I$ and define a derivative with respect to $\hat{q}^I = \hat{q}_I$ by

$$\frac{\partial}{\partial \hat{q}^I} \equiv \hat{q}_I \frac{\partial}{\partial q} + q^{-1} \frac{\partial}{\partial \hat{q}^I}. \quad (283)$$

The reason for this separation is that while terms involving derivatives with respect to $\hat{q}^I$ are very relevant for the bending of light, i.e., lensing, they are not relevant for most issues in primary and secondary anisotropy development (they can contribute if there is large scale mean curvature). We shall call the source associated with the first term $\mathcal{S}_{tSW}$ and the source associated with the second $\mathcal{S}_{tbend}$. Note that $\hat{q}_I \partial / \partial \hat{q}^I = 0$. Instead of repeating the Boltzmann equation, we shall write this in terms of the Δ_t notation introduced in section 3.1:

$$1 + \Delta_t \equiv (q/T_{c*}) / \ln(f_t^{-1} \pm 1), \quad (+) \text{ BE}, \quad (-) \text{ FD},$$

$$e_n[\Delta_t] + \frac{q}{q^n} \hat{q}^I \frac{1}{A} e_{*I}[\Delta_t] = \bar{N}^{-1} (\mathcal{G}_{tSW} + \mathcal{G}_{tbend} + \mathcal{G}_{tC}), \quad (284)$$

$$\bar{N}^{-1} \mathcal{G}_{tSW} = \left[1 + \Delta_t - \frac{\partial \Delta_t}{\partial \ln q} \right] \left[\hat{q}_K \hat{q}^J e_{*i}^K \{ e_n[e_{*J}^i] + \frac{1}{N} e_{*J}[N^i] \} \right. \\ \left. - \frac{q^n}{q} \hat{q}^J \frac{1}{A} e_{*J}[\ln N] + \frac{q}{q^n} \hat{q}^J \frac{1}{A} e_{*J}[\ln \Omega] - e_n[\ln(A/\Omega)] \right], \quad (285)$$

$$\bar{N}^{-1} \mathcal{G}_{tbend} = \frac{\partial \Delta_t}{\partial \hat{q}^I} (\delta^{IJ} - \hat{q}^I \hat{q}^J) \left[\frac{q^n}{q} \frac{1}{A} e_{*J}[\ln N] \right. \\ \left. - \frac{q}{q^n} \frac{1}{A} e_{*J}[\ln A] - \hat{q}_K e_{*i}^K \{ e_n[e_{*J}^i] + \frac{1}{N} e_{*J}[N^i] \} \right. \\ \left. + \frac{q}{q^n} \hat{q}_K \hat{q}^M \frac{1}{A} e_{*i}^K \{ e_{*M}[e_{*J}^i] - e_{*M}[e_{*J}^i] \} \right] \quad (286)$$

$$\bar{N}^{-1} \mathcal{G}_{t,source} \equiv \frac{\mathcal{S}[f_t] (1 + \Delta_t)^2}{(q/T_{c*})(f_c + \Delta f_t)(1 \pm (f_c + \Delta f_t))}. \quad (287)$$

For light massive neutrinos and photons whose spectrum is frequency dependent, it is better to use either $(1 + \Delta_t)^{-1}$ or $\ln(f_t^{-1} \pm 1)$, which is akin to a dimensionless generalized chemical potential, for the transport.

For massless *er*-particles, the components of the stress-energy tensor are related to Δ_t by

$$\left\{ \frac{\rho_{er}}{\bar{\rho}_{er}}, \frac{p_{er}}{\bar{p}_{er}}, \frac{J_{(e)er}^I}{\bar{\rho}_{er}}, \frac{\Pi^{IJ}}{\bar{\rho}_{er}} \right\} = \int \frac{d\Omega_{\hat{q}}}{4\pi} (1 + \Delta_t)^4 \{ 1, 1, \hat{q}^I, \hat{q}^I \hat{q}^J - \frac{1}{3} \delta^{IJ} \}.$$

The radiation brightness perturbation is defined to be $\bar{\rho}_{er}^{-1} d\rho_{er} / (d\Omega_{\hat{q}}/4\pi)$.

In [195, 2], I used $\Omega = A = \bar{a}$, conformal time, $\bar{N} = \bar{a}$, and a triad orthogonal to linear order in the metric perturbation $h_{ij} = \bar{a}^{-2} ({}^{(3)}g_{ij} - {}^{(3)}\bar{g}_{ij})$, where ${}^{(3)}\bar{g}_{ij}$ is the unperturbed (flat) spatial metric:

$$e_{*i}^I = \delta_{*i}^I + \frac{1}{2} h_I^I, \quad e_{*I}^i = \delta_{*I}^i - \frac{1}{2} h_I^i, \quad {}^{(3)}\bar{g}_{ij} = \bar{a}^2 \delta_{ij}. \quad (288)$$

Raising and lowering of indices is here done with respect to $\delta_{IJ} = e_I \cdot e_J$. We now concentrate on massless particle transport ($q^n = q$) in a flat unperturbed Universe for which the bending source is of second order, take $\Omega = A = \bar{a}$ and assume Δ_t is q -independent, as for Thomson scattering of a Planck distribution.

To linear order in $h_{\alpha\beta}$ and Δ_t , eq. (285) can be written as

$$\begin{aligned}\mathcal{G}_{tSW} &= \bar{N} \left(\hat{q}_K \hat{q}^J e_{*i}^K \{e_n[e_{*J}^i] + \frac{1}{N} e_{*J}[N^i]\} - \hat{q}^J \frac{1}{A} e_{*J}[\ln N] \right) \\ &= -\hat{q}^i \hat{q}^j \frac{1}{2} \dot{h}_{ij} + \hat{q}^i \hat{q}^j \partial_j h_{0i} + \frac{1}{2} \hat{q}^j \partial_j h_{00} \quad (= \dot{\Delta}_t + \hat{q}^i \partial_i \Delta_t). \quad (289)\end{aligned}$$

B.4 momentum space gauge transformations

There are two kinds of gauge transformations that operate on f_t , coordinate and momentum. For Thomson scattering problems, we usually restrict ourselves to classes of momenta for which $\partial f_t / \partial \tau$ vanishes in the unperturbed state, so f_t is a function only of q ; in that case, δf_t changes only under momentum gauge transformations. However, since we usually tie the momentum variable choice to the coordinate system choice (using a triad e^I perpendicular to the flow of time), the two are intimately related. We now discuss the general situation, where we allow the new momenta to be arbitrary functions of the old: $q_{\text{new}}^I(q_{\text{old}}^J)$. This accompanies the transformations of time, $\tau_{\text{new}} = \tau_{\text{old}} + T$, and space, $x_{\text{new}}^i = x_{\text{old}}^i + L^i$, coordinates.

The class of momentum transformations we have been discussing so far are conformal transformations of an orthonormal basis, hence $q_{\text{new}}^a = \Omega L_b^a q_{\text{old}}^b$, where L_b^a is a Lorentz transformation. We can therefore identify a velocity vector $\mathbf{v}$, a gamma factor $\gamma = (1 - \mathbf{v} \cdot \mathbf{v})^{-1/2}$ and a rotation matrix R_J^I such that

$$q_{\text{new}}^I = \frac{\Omega_{\text{new}}}{\Omega_{\text{old}}} R_J^I (q_{\text{old}}^J - \gamma q_{\text{old}}^n v^J + (\gamma - 1) \mathbf{q}_{\text{old}} \cdot \hat{v} v^J). \quad (290)$$

The old hypersurface as seen on the new hypersurface is moving with velocity $\mathbf{v}$. To linear order, we have the following transformations

$$\mathbf{q}_{\text{new}} = (1 + \delta \ln \Omega) R(\mathbf{q}_{\text{old}} - q_{\text{old}}^n \mathbf{v}), \quad (291)$$

$$q_{\text{new}} = q_{\text{old}} + q_{\text{old}} \delta \ln \Omega - q_{\text{old}}^n \hat{q}_{\text{old}} \cdot \hat{v} v, \quad (292)$$

$$\hat{q}_{\text{new}} = R \left(\hat{q}_{\text{old}} - \frac{q_{\text{old}}^n}{q_{\text{old}}} (\mathbf{v} - \hat{q}_{\text{old}} \cdot \mathbf{v} \hat{q}_{\text{old}}) \right), \quad (293)$$

$$\delta f_{t\text{new}} = \delta f_{t\text{old}} - q \frac{\partial \bar{f}}{\partial q} \frac{q_{\text{new}} - q_{\text{old}}}{q_{\text{old}}} - \left[\frac{\partial \bar{f}}{\partial \tau} T \right] \quad (294)$$

$$= \delta f_{t\text{old}} - q \frac{\partial \bar{f}}{\partial q} \left(\delta \ln \Omega - \frac{q_{\text{old}}^n}{q_{\text{old}}} \hat{q}_{\text{old}} \cdot \hat{v} v \right) - \left[\frac{\partial \bar{f}}{\partial \tau} T \right], \quad (295)$$

$$\begin{aligned}\Delta_{t\text{new}} &= \Delta_{t\text{old}} + \frac{q_{\text{new}} - q_{\text{old}}}{q_{\text{old}}} \\ &= \Delta_{\text{old}} + \left(\delta \ln \Omega - \frac{q_{\text{old}}^n}{q_{\text{old}}} \hat{q}_{\text{old}} \cdot \hat{v} v \right) + \left[\frac{\partial \bar{f} / \partial \tau}{\partial \bar{f} / \partial \ln q} T \right]. \quad (296)\end{aligned}$$

Notice that it is only the redshifting associated with the conformal factor or relative flow that enters in the transformation of f . If we restrict ourselves to the class of comoving momenta, then the terms in square brackets vanish.

We know how the velocity $\mathbf{v}$ and the scale factor transform ($\bar{a}_{\text{new}} = \bar{a}_{\text{old}} + \bar{H}\bar{N}T$). Although we have restricted ourselves to momenta that have Ω reducing to $\bar{a}$ in the unperturbed case, we have some freedom in deciding how $(\bar{a}^{-1}\Omega)$ transforms. Therefore, for the combined coordinate and gauge transformation of the radiation distribution function for scalar and tensor perturbations we have

$$\begin{aligned}\Delta_{t\text{new}}^{(S)} &= \Delta_{t\text{old}}^{(S)} + \bar{H}\bar{N}T + \hat{q}^I{}^{(3)}\bar{\nabla}_I\bar{N}T + \left[\ln \left(\frac{(\bar{a}^{-1}\Omega)_{\text{new}}}{(\bar{a}^{-1}\Omega)_{\text{old}}} \right) \right]^{(S)}, \\ \Delta_{t\text{new}}^{(T)} &= \Delta_{t\text{old}}^{(T)} + \left[\ln \left(\frac{(\bar{a}^{-1}\Omega)_{\text{new}}}{(\bar{a}^{-1}\Omega)_{\text{old}}} \right) \right]^{(T)}.\end{aligned}\quad (297)$$

For the most common choice for Ω , namely $\Omega = \bar{a}$, we see that, as expected, the angle average of $4\Delta_t^{(S)}$ transforms as a density perturbation and the first moment with respect to $\hat{q}^I$ transforms as a velocity, while all higher moments, including that for the anisotropic stress, are gauge invariant. And $\Delta_t^{(T)}$ is gauge invariant. Looking at eq. (260), we see that the following quantity is gauge invariant,

$$\Delta_t^{(S)} - \bar{H}\Psi_\sigma - \hat{q}^I\bar{e}_I[\Psi_\sigma] = \Delta_{tL}^{(S)}, \quad (298)$$

as are many other combinations. Equation (298) relates the distribution function in the longitudinal gauge to that in the synchronous gauge.

A pure momentum gauge transformation with $\bar{a}^{-1}\Omega_{\text{old}} = 1$ and $\bar{a}^{-1}\Omega_{\text{new}} = e^\nu$ gives $\Delta_{\text{new}} = \Delta_{\text{old}} + \nu$. This turns out to be a more relevant combination for the longitudinal gauge. The synchronous gauge combination is:

$$\Delta_{tS}^{(S)} + \ddot{\psi} - \hat{q}^I\partial_I\dot{\psi} = \Delta_{tL}^{(S)} + \nu_L. \quad (299)$$

When we solve the transport equations in the synchronous gauge, it is this quantity which free-streams after the photons have decoupled [88].

The momentum gauge transformation can be quite decoupled from coordinate transformations: it is worthwhile to show explicitly the remarkable flexibility that it allows. We can discuss this entirely in terms of $\bar{a}^{-1}\Omega$ which we now allow to be an arbitrary function of $q, \hat{q}^I$. In particular, it can be expanded in spherical harmonics to induce the following transformation:

$$\ln(\bar{a}^{-1}\Omega)_{\text{new}} = \ln(\bar{a}^{-1}\Omega)_{\text{old}} + \sum_{\ell m} z_{\ell m}(q_{\text{old}}, \mathbf{x}, \tau) Y_{\ell m}(\hat{q}_{\text{old}}), \quad (300)$$

$$\Delta_{\text{new}} = \Delta_{\text{old}} + \sum_{\ell m} z_{\ell m}(q_{\text{old}}) Y_{\ell m}(\hat{q}). \quad (301)$$

We are therefore allowed to perform gauge transformations on Δ_t beyond $\ell = 0, 1$ if we wish, although these are not connected to coordinate transformations. Indeed, it appears to be possible to use the momentum transformation to completely remove the distribution function perturbation. Of course, tensor, octopole and higher multipoles in the momentum gauge transformation

modify the transport operator: $\mathcal{G}_{tSW}$ transforms as well and if we allow an order ℓ term to appear in Δ_t , the $\hat{q}^j \partial_j \Delta_t$ term in the transport operator will induce a term of order $\ell+1$ in $\hat{q}$ in $\mathcal{G}_{tSW}$. Since $\mathcal{G}_{tSW}$ and the Compton source function have terms that are at most quadratic in $\hat{q}$, it would seem wise not to induce terms cubic and higher order in $\hat{q}$. This restricts the class of momentum transformations on Δ_t to have only $\ell = 0, 1$ terms. Consider how we would get the combination eq. (299) in the synchronous gauge: we would make a pure momentum gauge transformation, $\ln(\bar{a}^{-1}\Omega)_{\text{new}} = \psi - \hat{q}^I \partial_I \psi$. In practice one doesn't usually think of it this way. Rather one takes the transport operator and $\mathcal{G}_{tSW}$ and shuffles terms from the *right-hand side* to the *left-hand side* if it looks convenient to do so. That is how we decided that the combination $\Delta_{tS} + \dot{\psi} - \hat{q}^I \partial_I \psi$ was useful computationally [88].

If we change the momentum variables for one species, but not another, then the interpretation becomes more complicated. For example, we should require that such physically meaningful quantities as the entropy per baryon perturbation $\frac{3}{4}\delta_\gamma - \delta_B$ and the relative velocity $\mathbf{v}_\gamma - \mathbf{v}_B$ be gauge invariant under the combination of spacetime and momentum coordinate changes. However, all species present will have distribution functions, and they can all be transformed. Thus, for example, just as the photon density transforms to $\delta_\gamma + 4\nu$ under $\bar{a}^{-1}\Omega = e^\nu$, so the baryon density transforms to $\delta_B + 3\nu$.

For the flat unperturbed case, we can do a Fourier expansion of the distribution function and the Sachs–Wolfe source. In the frame in which $\mathbf{k}$ is taken to be along the 3-axis and (θ, ϕ) are the polar angles, the Sachs–Wolfe source terms under the standard momentum gauge choice are

$$(\Omega = \bar{a}) \text{ scalar: } \mathcal{G}_{tSW}^{(S)} = -ik\mu\nu - \dot{\varphi} - \mu^2 k^2 \bar{a}^{-1} \Psi_\sigma, \quad \mu = \hat{k} \cdot \hat{q}, \quad (302)$$

$$\begin{aligned} \text{tensor: } \mathcal{G}_{tSW}^{(T)} &= -\frac{1}{2} \hat{q}^i \hat{q}^j \dot{h}_{ij}^{TT} \\ &= -(1 - \mu^2) \left(\tilde{\mathcal{G}}_{tSW}^{(T+)} \cos(2\phi) + \tilde{\mathcal{G}}_{tSW}^{(T\times)} \sin(2\phi) \right), \\ \tilde{\mathcal{G}}_{tSW}^{(T\epsilon)} &= \frac{1}{2} \dot{h}_{(T\epsilon)}, \quad \dot{h}_{(T+)} = \frac{1}{2} (\dot{h}_{11} - \dot{h}_{22}), \quad \dot{h}_{(T\times)} = \dot{h}_{12} \end{aligned} \quad (303)$$

There is of course no effect on the polarization components. Under a further pure momentum transformation, the distribution function and Sachs–Wolfe scalar terms transform to:

$$\Omega = \bar{a} \left(1 + \nu + \frac{\partial \bar{a}^{-1} \Psi_\sigma}{\partial \tau} - \hat{q}^i \partial_i \bar{a}^{-1} \Psi_\sigma \right), \quad (304)$$

$$\tilde{\Delta}_t^{(S)} = \Delta_t^{(S)} + \nu + \frac{\partial \bar{a}^{-1} \Psi_\sigma}{\partial \tau} - \hat{q}^i \partial_i \bar{a}^{-1} \Psi_\sigma, \quad (305)$$

$$\tilde{\mathcal{G}}_{tSW}^{(S)} = \dot{\nu} - \dot{\varphi} + \frac{\partial^2 \bar{a}^{-1} \Psi_\sigma}{\partial \tau^2}, \quad (306)$$

while the tensor terms remain invariant.

C Polarized transport for Thomson scattering

C.1 The polarization matrix and Stokes parameters

The off-diagonal components $f_{s_2 s_1}$, $s_2 \neq s_1$, of the distribution function contain phase information, describing the probability amplitude for propagation from a state of spin s_1 to a state of spin s_2 . For photons, there are two polarizations, hence a 2×2 “polarization matrix” [258] transverse to $\hat{q} \otimes \hat{q}$ is required for photons in the direction $\hat{q}$. Consider linear polarization. If we expand the polarization distribution function in terms of the basis consisting of the Pauli matrices $\sigma^{(i)}$, $i = 1, 2, 3$ and the identity $\sigma^{(0)} = \text{diag}(1, 1)$,

$$(f_{ss'}) = \sum_{\mu=0}^3 \frac{1}{2} f_{(\mu)} \sigma^{(\mu)}, \quad (307)$$

then the 4 real distribution functions $f_{(\mu)}$ correspond to conventional Stokes parameters, except that they are defined for distribution functions, as described in section 3:

$$f_{(0)} = f_t, \quad f_{(1)} = f_U, \quad f_{(2)} = f_V, \quad f_{(3)} = f_Q. \quad (308)$$

By using polarization vectors one can define an object in the combined position and momentum space, $\mathbf{f}$, which has properties similar to a spatial tensor of rank two for fixed momentum. Consider a photon travelling in the direction $\hat{q}$ and two polarization vectors $\varepsilon_{1,2}$ perpendicular to $\hat{q}$ and to each other. The $\varepsilon_{\mathbf{A}}$ will be functions of $\hat{q}$ and possibly of $\mathbf{x}$ or $\mathbf{k}$. To make a tensor out of the 2×2 matrix, $f_{(\mu)} \sigma^{(\mu)}/2$, we use the tensor product basis

$$\begin{aligned} \mathcal{E}_t &= \frac{1}{2}(\varepsilon_1 \otimes \varepsilon_1 + \varepsilon_2 \otimes \varepsilon_2), & \mathcal{E}_Q &= \frac{1}{2}(\varepsilon_1 \otimes \varepsilon_1 - \varepsilon_2 \otimes \varepsilon_2), \\ \mathcal{E}_U &= \frac{1}{2}(\varepsilon_1 \otimes \varepsilon_2 + \varepsilon_2 \otimes \varepsilon_1), & \mathcal{E}_V &= -\frac{1}{2}i(\varepsilon_1 \otimes \varepsilon_2 - \varepsilon_2 \otimes \varepsilon_1), \\ \text{i.e., } \mathcal{E}_{(\mu)} &= \sum_{A,B=1,2} \frac{1}{2} \sigma_{(\mu)}^{AB} \varepsilon_{\mathbf{A}} \otimes \varepsilon_{\mathbf{B}}. \end{aligned} \quad (309)$$

For observations, the basis $\varepsilon_A(\hat{q}, \mathbf{x}, \tau)$ would, for given $\hat{q}$, be defined with some axis convention on the celestial sphere; the tensor $\mathbf{f} = \sum_{\mu=0}^3 f_{(\mu)} \mathcal{E}_{(\mu)}$ is independent of polarization basis orientation, with $f_{(\mu)}$ transforming under rotation of the polarization basis in a complementary way to $\mathcal{E}_{(\mu)}$. It is useful to also use a polarization basis whose orientation is defined with respect to eigenmode variables in the expansion eq. (147). For the flat case, a wavenumber $\mathbf{k}$ can be used to label the eigenfunctions, $\varepsilon_A(\hat{q}, \hat{k}, \tau)$ can be a function of $\hat{k}$, independent of $\mathbf{x}$, and a mode expansion can be made:

$$\mathbf{f} = w \sum_{\mathcal{M}\mathbf{k}} f_{(\mu)}^{(\mathcal{M})} \mathcal{E}_{(\mu)} Q_{\mathbf{k}\mathcal{M}}(\mathbf{x}, \tau) a_{\mathbf{k}\mathcal{M}} + \text{cc}. \quad (310)$$

We can also expand $\mathbf{f}$ in the basis $\{e_I\}$ of the time hypersurfaces, $\mathbf{f} = f^{IJ} e_I \otimes e_J$, which makes the spatial tensor aspect manifest. Just as ${}^{(3)}g_{ij}$ is expanded in

scalar, vector, and tensor modes, so can f_{ij} . For scalar perturbations, we project onto δ_{ij} and $\hat{k}_i \hat{k}_j - \frac{1}{3} \delta_{ij}$. As we show below, we can choose $\varepsilon_2 \perp \hat{k}$ as well as $\perp \hat{q}$, which implies $f_U^{(S)} = 0$ and $f_V^{(S)} = 0$. For tensor perturbations, we project using $E_{ij}^{(T\epsilon)}$ of eq. (170):

$$f_{ij}^{(T)} = w \sum_{(\mu)=t,Q,U,V} \sum_{\epsilon=+,\times} \sum_{\mathbf{k}} \tilde{f}_{(\mu)}^{(T\epsilon)} \frac{E^{(T\epsilon)} \cdot \mathcal{E}_{(\mu)}}{\mathcal{E}_{(\mu)} \cdot \mathcal{E}_{(\mu)}} E_{ij}^{(T\epsilon)} e^{i\mathbf{k} \cdot \mathbf{x}} a_{\mathbf{k}(T\epsilon)} + \text{cc}. \quad (311)$$

The quantities $\tilde{f}_{(\mu)}^{(T\epsilon)}$ or equivalently $\tilde{\Delta}_{(\mu)}^{(T\epsilon)}$ are the natural mode functions for tensor perturbations. These are evaluated in section C.6.

Thomson scattering is conservative, hence in the comoving frame of the baryons, the photon energy out equals the photon energy in. The scattering function that enters the Boltzmann transport equation can then be written as

$$\begin{aligned} \mathcal{S}_{s_2 s_1}(\mathbf{x}, \tau, q, \hat{q}) = & \\ & - \sum_{s'_1 s'_2} \left(\int d\Omega_{\hat{q}'} R_{s_2 s_1; s'_2 s'_1}(\mathbf{x}, \tau, q; \hat{q} \rightarrow \hat{q}') f_{s_2 s_1}(\mathbf{x}, \tau, q, \hat{q}) \right. \\ & \left. - \int d\Omega_{\hat{q}'} R_{s'_2 s'_1; s_2 s_1}(\mathbf{x}, \tau, q; \hat{q}' \rightarrow \hat{q}) f_{s'_2 s'_1}(\mathbf{x}, \tau, q, \hat{q}') \right). \end{aligned}$$

Denote the total scattering rate (per unit conformal time) by

$$\tau_C^{-1} \equiv n_e \sigma_T a = \sum_{s'_1 s'_2} \int d\Omega_{\hat{q}'} R_{s_2 s_1; s'_2 s'_1}(\mathbf{x}, \tau, q; \hat{q} \rightarrow \hat{q}') \quad (312)$$

and define a phase function by

$$P_{s'_2 s'_1; s_2 s_1}(\mathbf{x}, \tau; \hat{q}' \rightarrow \hat{q}) \equiv 4\pi \tau_C R_{s'_2 s'_1; s_2 s_1}(\mathbf{x}, \tau, q; \hat{q}' \rightarrow \hat{q}) \quad (313)$$

(independent of the magnitude of q for Thompson scattering). Instead of proceeding with the polarization matrix language, let us go over into the Stokes parameter language, noting that we can expand any symmetric matrix in spin space in terms of $\sigma^{(\mu)}$,

$$\mathcal{S}_{s_2 s_1} = \mathcal{S}_{(\mu)} \frac{\sigma^{(\mu)}}{2}; \quad (314)$$

in particular, we can expand the source function. The phase function $P_{(\mu)}^{(\nu)}$ maps the distribution from a $\hat{q}'$ -orthogonal system to a $\hat{q}$ -orthogonal system. We can then write

$$\begin{aligned} \mathcal{S}_{(\mu)}(\mathbf{x}, \tau, q, \hat{q}) = & -\tau_C^{-1} \left(f_{(\mu)}(\mathbf{x}, \tau, q, \hat{q}) \right. \\ & \left. + \int \frac{d\Omega_{\hat{q}'}}{4\pi} P_{(\mu)}^{(\nu)}(\mathbf{x}, \tau, \hat{q}' \rightarrow \hat{q}) f_{(\nu)}(\mathbf{x}, \tau, q, \hat{q}') \right). \end{aligned} \quad (315)$$

Our goal is therefore to calculate $P_{(\mu)}^{(\nu)}$, or equivalently the spatial tensor map

$$\mathbf{P} = \sum_{\mu, \nu=0}^3 P_{(\mu)}^{(\nu)} \mathcal{E}_{(\mu)}(\hat{q}, \mathbf{x}, \tau) \otimes \mathcal{E}_{(\nu)}(\hat{q}, \mathbf{x}, \tau) \quad (316)$$

expressed in terms of sky orientation (or via mode expansions).

The calculation of $P_{(\mu)}^{(\nu)}$ is done through a sequence of “rotations” of the Stokes parameters which progressively take us: from (1) a linear polarization basis $E'_{1,2}$ perpendicular to the photon direction $\hat{q}'$ before the scattering and referred to the sky reference frame; through (2) a linear polarization basis $\varepsilon'_{1,2}$ in a plane perpendicular to $\hat{q}'$, which, for convenience also has ε'_2 perpendicular to $\hat{k}$; into (3) a polarization basis $e'_{1,2}$ in a plane $\perp \hat{q}'$, and also $e'_2 \perp \hat{q}$, a natural basis for action on the distribution function by the scattering phase matrix, with the result re-expressed in terms of a new polarization basis $e_{1,2}$ spanning a plane $\perp \hat{q}$ with $e_2 \perp \hat{q}'$; through (4) a linear polarization basis $\varepsilon_{1,2}$ in a plane perpendicular to $\hat{q}$ with ε_2 perpendicular to $\hat{k}$ as well; and, finally, into (5) a linear polarization basis $E_{1,2}$ in a plane perpendicular to $\hat{q}$ referred to the sky reference frame. The transformations are all designed to get the distribution function into the correct form for step (3), in which the familiar action of Thomson scattering of light linearly polarized in a direction perpendicular and parallel to the *scattering plane* (that spanned by $\hat{q}'$ and $\hat{q}$) can be performed. The bases in steps (2) and (4) are suited to the free transport between scatters, since they are a natural polarization basis for the independent modes of the system. The rotations (1), (2), (3) leave $\mathbf{f}(\hat{q}', \mathbf{x}, \tau)$ invariant and the rotations (3), (4), (5) leave $\mathbf{f}(\hat{q}, \mathbf{x}, \tau)$ invariant, with the entire action of the scatter expressible as the transformation step from $\hat{q}'$ to $\hat{q}$, in terms of the mapping $\mathbf{P}(\hat{q}' \rightarrow \hat{q})$:

$$\begin{aligned} \mathbf{P}(\hat{q}' \rightarrow \hat{q}) = & \frac{3}{4}(1 + (\hat{q}' \cdot \hat{q})^2) \{ (e_2(\hat{q}) \otimes e_2(\hat{q})) \otimes (e'_2(\hat{q}') \otimes e'_2(\hat{q}')) \} \\ & + \frac{3}{4} 2\hat{q}' \cdot \hat{q} \frac{1}{2} \{ (e_1(\hat{q}) \otimes e_2(\hat{q})) \otimes (e'_2(\hat{q}') \otimes e'_1(\hat{q}')) \\ & + (e_2(\hat{q}) \otimes e_1(\hat{q})) \otimes (e'_1(\hat{q}') \otimes e'_2(\hat{q}')) \} . \end{aligned} \quad (317)$$

This relatively simple expression demonstrates the utility of the $\mathbf{f}$ approach, although for it to be usable $e'_{\mathbf{A}}$ and $e_{\mathbf{A}}$ must be expressed in terms of the mode-bases $\varepsilon'_{\mathbf{A}}$ and $\varepsilon_{\mathbf{A}}$ and sky-bases $E'_{\mathbf{A}}$ and $E_{\mathbf{A}}$, which is where the work lies. Chandrasekhar [200] develops the Stokes parameter equations in his section on Rayleigh scattering, which has the same angular scattering dependence as Thomson scattering, by doing these rotations, but using a more classical language and approach.

The full sequence of operations can be expressed in terms of a total phase tensor

$$P_{(\mu)}^{(\nu)} = [P_5]_{(\mu)}^{(\alpha)} [P_4]_{(\alpha)}^{(\beta)} [P_{scat}]_{(\beta)}^{(\gamma)} [P_2]_{(\gamma)}^{(\delta)} [P_1]_{(\delta)}^{(\nu)} , \quad (318)$$

acting on the distribution function $f_{(\nu)}(\mathbf{x}, \tau, \hat{q}')$. However, since the computational method to solve the transport equation uses the modes of the system, $f_{(\delta)}(\mathbf{k}, \tau, \hat{q}')$, we actually do not need to do step (1).

In linear perturbation theory for an Einstein–deSitter Universe, the modes are plane waves, labelled by the comoving wavevector $\mathbf{k}$. A linear polarization basis in which ε'_2 is perpendicular to $\hat{k}$ as well as $\hat{q}'$ is ([88], Appendix 5)

$$\varepsilon'_2 = \frac{\hat{k} \times \hat{q}'}{(1 - (\hat{k} \cdot \hat{q}')^2)^{1/2}}, \quad \varepsilon'_1 = \hat{q}' \times \varepsilon'_2 = \frac{\hat{k} - (\hat{k} \cdot \hat{q}')\hat{q}'}{(1 - (\hat{k} \cdot \hat{q}')^2)^{1/2}}. \quad (319)$$

Thus $\{\varepsilon'_1, \varepsilon'_2, \hat{q}'\}$ is an orthonormal triad for the incoming photon state. For given $\hat{k}$ and $\hat{q}'$, the incoming Stoke's parameters are in this coordinate system. Similarly, $\{\varepsilon_1, \varepsilon_2, \hat{q}\}$ with $\hat{q}$ replacing $\hat{q}'$ in eq. (319) is an appropriate triad for the outgoing (scattered) photon state, but after the polarizing action of the scatter is taken into account.

In the scattering frame, we define

$$e'_1 = \frac{\hat{q} - \hat{q} \cdot \hat{q}' \hat{q}'}{(1 - (\hat{q} \cdot \hat{q}')^2)^{1/2}}, \quad e'_2 = e'_1 \times \hat{q}' = \frac{\hat{q} \times \hat{q}'}{(1 - (\hat{q} \cdot \hat{q}')^2)^{1/2}}. \quad (320)$$

Thus, $e'_1 = \hat{q}' \times e'_2$ and, very importantly, e'_2 is $\perp \hat{q}$, i.e., is perpendicular to the scattering plane. By interchanging $\hat{q}'$ and $\hat{q}$ we get the outgoing triad $\{e_1, e_2, \hat{q}\}$, with polarization basis differing from the incoming one by sign changes:

$$e_2 \equiv -e'_2, \quad e_1 = \hat{q} \times e_2. \quad (321)$$

The angular dependence of Thomson scattering on the Stokes parameters is described by the phase tensor

$$\begin{aligned} [P_{scat}]_{(0)}^{(0)} &= [P_{scat}]_{(3)}^{(3)} = \frac{3}{2}(1 + (\hat{q} \cdot \hat{q}')^2)/2, \\ [P_{scat}]_{(3)}^{(0)} &= [P_{scat}]_{(0)}^{(3)} = -\frac{3}{2}(1 - (\hat{q} \cdot \hat{q}')^2)/2, \\ [P_{scat}]_{(1)}^{(1)} &= [P_{scat}]_{(2)}^{(2)} = \frac{3}{2}\hat{q} \cdot \hat{q}'. \end{aligned} \quad (322)$$

The rest of the components vanish, thanks to the particular e'_1, e'_2 basis choice with e'_2 perpendicular to the scattering plane. This gives eq. (317).

But we wish to use the incoming mode-basis, $\varepsilon'_{1,2}$ and outgoing basis $\varepsilon_{1,2}$. A rotation about the direction $\hat{q}'$ by an angle ψ' takes $\varepsilon'_{1,2}$ into $e'_{1,2}$, where

$$\cos \psi' = \varepsilon'_2 \cdot e'_2 = \varepsilon'_1 \cdot e'_1 = \frac{\hat{k} \cdot \hat{q} - \hat{k} \cdot \hat{q}' \hat{q} \cdot \hat{q}'}{(1 - (\hat{q} \cdot \hat{q}')^2)^{1/2} (1 - (\hat{k} \cdot \hat{q}')^2)^{1/2}}. \quad (323)$$

The effect of the basis change on $f_{(\mu)}(\mathbf{k}, \hat{q}', \tau)$ is encoded in the action of the 2×2 rotation matrix

$$e^{i\psi' \sigma^{(2)}} = \cos(\psi') \sigma^{(0)} + \sin(\psi') i \sigma^{(2)} = \begin{bmatrix} \cos(\psi') & \sin(\psi') \\ -\sin(\psi') & \cos(\psi') \end{bmatrix} \quad (324)$$

acting on the left of the polarization matrix and its inverse (adjoint) acting on the right:

$$\begin{aligned}\frac{1}{2}([P_2]_{(\beta)}^{(\mu)} f_{(\mu)}) \sigma^{(\beta)} &= e^{i\psi' \sigma^{(2)}} \frac{1}{2} f_{(\mu)}(\mathbf{k}, \hat{q}', \tau) \sigma^{(\mu)} e^{-i\psi' \sigma^{(2)}}, \\ [P_2]_{(0)}^{(0)} &= [P_2]_{(2)}^{(2)} = 1, \\ [P_2]_{(3)}^{(3)} &= [P_2]_{(1)}^{(1)} = \cos(2\psi'), \quad [P_2]_{(1)}^{(3)} = -[P_2]_{(3)}^{(1)} = \sin(2\psi').\end{aligned}$$

The rest of the $[P_2]_{(\beta)}^{(\mu)}$ vanish. The rotation by angle ψ from the triad $\{e_1, e_2, \hat{q}\}$ to the triad $\{\varepsilon_1, \varepsilon_2, \hat{q}\}$ gives a phase tensor $[P_4]_{(\mu)}^{(\nu)}$ identical in form to $[P_2]_{(\mu)}^{(\nu)}$ if we replace ψ' by $-\psi$, where $\cos\psi$ is similar to eq. (323) with $\hat{q}$ and $\hat{q}'$ interchanged.

We now have all of the ingredients to get $P_{(\alpha)}^{(\delta)} = [P_4]_{(\alpha)}^{(\beta)} [P_{scat}]_{(\beta)}^{(\gamma)} [P_2]_{(\gamma)}^{(\delta)}$. To make the form useful, we need to express $\hat{q}, \hat{q}', \hat{k}$ in some coordinate basis. Let us choose polar coordinates with $\hat{k}$ the pole and $\hat{q} = (\theta, \phi)$ and $\hat{q}' = (\theta', \phi')$. $P_{(\mu)}^{(\nu)}$ is then a function of $\mu = \hat{k} \cdot \hat{q}$, $\mu' = \hat{k} \cdot \hat{q}'$, and $\hat{q} \cdot \hat{q}' = \sqrt{1 - \mu^2} \sqrt{1 - (\mu')^2} \cos(\phi - \phi') + \mu\mu'$. The phase tensor can be expanded in terms of $\cos(m(\phi - \phi'))$, $\sin(m(\phi - \phi'))$, where $m = 0, 1, 2$ terms appear. Thus we have a sequence of products such as $\cos(m\phi) \cos(m\phi')$ and $\cos(m\phi) \sin(m\phi')$. The conventional approach is to expand the incoming distribution function (or equivalently the temperature fluctuation $\Delta_{(\mu)}$) in $\cos(m\phi')$ and $\sin(m\phi')$ terms and the outgoing distribution in $\cos(m\phi)$ and $\sin(m\phi)$ terms; i.e., into scalar and tensor terms, and vector terms denoted by “vec” which we ignore:

$$\begin{aligned}\Delta_t(\hat{q}, \mathbf{k}) &= \Delta_t^{(S)}(\mu) + \text{vec} + \Delta_t^{(T+)}(\mu) \cos(2\phi) + \Delta_t^{(T\times)}(\mu) \sin(2\phi) \\ \Delta_Q(\hat{q}, \mathbf{k}) &= \Delta_Q^{(S)}(\mu) + \text{vec} + \Delta_Q^{(T+)}(\mu) \cos(2\phi) + \Delta_Q^{(T\times)}(\mu) \sin(2\phi) \\ \Delta_U(\hat{q}, \mathbf{k}) &= \Delta_U^{(S)}(\mu) + \text{vec} - \Delta_U^{(T+)}(\mu) \sin(2\phi) + \Delta_U^{(T\times)}(\mu) \cos(2\phi) \\ \Delta_V(\hat{q}, \mathbf{k}) &= \Delta_V^{(S)}(\mu) + \text{vec} - \Delta_V^{(T+)}(\mu) \sin(2\phi) + \Delta_V^{(T\times)}(\mu) \cos(2\phi)\end{aligned}\tag{325}$$

In the same way, we can also expand the source function $\mathcal{S}_{(\mu)}$ for $f_{(\mu)}$ – or $\mathcal{G}_{(\mu)}$ for $\Delta_{(\mu)}$, defining $\mathcal{G}_{(\mu)}^{(S)}(\mu), \mathcal{G}_{(\mu)}^{(V\ c, s)}(\mu), \mathcal{G}_{(\mu)}^{(T+, \times)}(\mu)$. The reason for the different sin and cos combinations for $\Delta_U^{(T+)}, \Delta_Q^{(T+)}$ is that the phase tensor expansion couples + to + and $\times$ to $\times$ but not + to $\times$: i.e., the modes are independent. Using the 3-tensor Δ and the 3-tensor map $\mathbf{P}$, we do not go through this intermediate step of defining $\Delta_{(\mu)}^{(T\{+, \times\})}$, but rather go directly to variables $\tilde{\Delta}_{(\mu)}^{(T\{+, \times\})}$, which have a further μ dependence removed from them. For $\Delta_{(\mu)}^{(S)}$ there is no difference.

To calculate the polarization a detector would observe, we must choose a fixed frame on the sky, say Galactocentric coordinates. Since $-\hat{q}$ points outward in the radial direction, the two polarization vectors on the sky $E_1, E_2 = E_1 \times \hat{q}$

form an orthonormal basis for the celestial sphere. An angle $\psi_{\mathbf{k}}$ defines the rotation to $\{\varepsilon_1, \varepsilon_2\}$. It is the angle between the fixed E_1 and $\mathbf{k}_\perp$, where $\mathbf{k}_\perp = \mathbf{k} - \mathbf{k} \cdot \hat{q} \hat{q}$, is, as we look out upon a specific spot on the celestial sphere, the projection of $\mathbf{k}$ onto it. The phase tensor $[P_5]_{(\mu)}^{(\nu)}$ is identical in form to $[P_4]_{(\mu)}^{(\nu)}$ with ψ replaced by $\psi_{\mathbf{k}}$ which acts on $\Delta_{(\mu)}(\mathbf{k}, \hat{q}, \tau)$ to give

$$\begin{aligned}\Delta_t(\hat{q}, \text{here, now}) &= \int \frac{d^3 \mathbf{k}}{(2\pi)^3} \Delta_t(\mathbf{k}, \hat{q}, \tau_0), \\ \Delta_Q(\hat{q}, \text{here, now}) &= \int \frac{d^3 \mathbf{k}}{(2\pi)^3} (\Delta_Q(\mathbf{k}, \hat{q}, \tau_0) \cos(2\psi_{\mathbf{k}}) + \Delta_U \sin(2\psi_{\mathbf{k}})), \\ \Delta_U(\hat{q}, \text{here, now}) &= \int \frac{d^3 \mathbf{k}}{(2\pi)^3} (-\Delta_Q(\mathbf{k}, \hat{q}, \tau_0) \sin(2\psi_{\mathbf{k}}) + \Delta_U \cos(2\psi_{\mathbf{k}})), \\ \Delta_V(\hat{q}, \text{here, now}) &= \int \frac{d^3 \mathbf{k}}{(2\pi)^3} \Delta_V(\mathbf{k}, \hat{q}, \tau_0). \\ \cos(2\psi_{\mathbf{k}}) &= -E_{2i} E_{2j} \frac{(\hat{k}_i - \hat{k} \cdot \hat{q} \hat{q}_i)(\hat{k}_j - \hat{k} \cdot \hat{q} \hat{q}_j) - \epsilon_{imn} \epsilon_{jrs} \hat{q}_m \hat{q}_r \hat{k}_n \hat{k}_s}{(1 - (\hat{k} \cdot \hat{q})^2)}, \\ \sin(2\psi_{\mathbf{k}}) &= -E_{2i} E_{2j} \frac{(\hat{k}_j - \hat{k} \cdot \hat{q} \hat{q}_j) \epsilon_{imn} \hat{q}_m \hat{k}_n}{(1 - (\hat{k} \cdot \hat{q})^2)},\end{aligned}\tag{326}$$

where the summation convention on repeated indices has been used and ϵ_{imn} is the completely antisymmetric Levi-Cevita symbol. Because of the $\mathbf{k}$ dependence of the phases implicit in $\Delta_Q(\hat{q}, \mathbf{k}, \tau_0)$, etc. we cannot do the $\psi_{\mathbf{k}}$ integration in eq. (326). A strategy for making small angle polarization maps using this formula and knowledge of the polarization power spectrum is described in [88].

C.2 Scalar perturbation source terms

C.2.1 Thomson source functions

For the scalar components of the phase function, we have

$$\begin{aligned}[P_S]_{(0)}^{(0)} &= \frac{3}{8}[3 - \mu^2 - (\mu')^2 + 3\mu^2(\mu')^2], \\ [P_S]_{(0)}^{(3)} &= \frac{3}{8}[3\mu^2 - 1][(\mu')^2 - 1], \quad [P_S]_{(3)}^{(0)} = \frac{3}{8}[\mu^2 - 1][3(\mu')^2 - 1], \\ [P_S]_{(3)}^{(3)} &= \frac{3}{8}[3 - 3\mu^2 - 3(\mu')^2 + 3\mu^2(\mu')^2], \quad [P_S]_{(2)}^{(2)} = \frac{3}{4}\mu\mu',\end{aligned}$$

with the rest vanishing. In terms of an expansion of $\Delta_t^{(S)}$ and $\Delta_Q^{(S)}$ in angular moments, $\Delta_{t\ell}^{(S)}$ and $\Delta_{Q\ell}^{(S)}$, with respect to Legendre polynomials [88]

$$\Delta_{t,Q,U,V}^{(S)}(\hat{q}, \mathbf{k}, \tau) = \sum_{\ell} (2\ell + 1) (-i)^{\ell} \Delta_{t,Q,U,V}^{(S)}(k, \tau) P_{\ell}(\hat{q} \cdot \hat{k}),\tag{327}$$

we have

$$\tau_C \mathcal{G}_{tC}^{(S)} = -\Delta_t^{(S)} + \Delta_{t0}^{(S)} - \frac{1}{2} P_2(\hat{k} \cdot \hat{q}) (\Delta_{t2}^{(S)} + \Delta_{Q2}^{(S)} + \Delta_{Q0}^{(S)}), \quad (328)$$

$$\tau_C \mathcal{G}_{QC}^{(S)} = -\Delta_Q^{(S)} + \frac{1}{2} (1 - P_2(\hat{k} \cdot \hat{q})) (\Delta_{Q0}^{(S)} + \Delta_{t2}^{(S)} + \Delta_{Q2}^{(S)}), \quad (329)$$

$$\tau_C \mathcal{G}_{UC}^{(S)} = 0, \quad (330)$$

$$\tau_C \mathcal{G}_{VC}^{(S)} = -\Delta_V^{(S)} + \frac{3}{4} \hat{k} \cdot \hat{q} \Delta_{V1}^{(S)}. \quad (331)$$

This equation was derived in the comoving baryon gauge, but the transformation of $\Delta_t^{(S)} - \Delta_{t0}^{(S)}$ to a frame in which the baryons are moving with velocity $\mathbf{v}_B^{(S)}$ can be done using eq. (296), which only modifies eq. (328):

$$\tau_C \mathcal{G}_{tC}^{(S)} = -\Delta_t^{(S)} + \Delta_{t0}^{(S)} + \hat{q} \cdot \mathbf{v}_B^{(S)} - \frac{1}{2} P_2(\mu) (\Delta_{t2}^{(S)} + \Delta_{Q2}^{(S)} + \Delta_{Q0}^{(S)}). \quad (332)$$

In eq. (332), the source term proportional to $\Delta_{t2}^{(S)}$ arises because of the angular dependence of Thomson scattering. This quadrupole anisotropy is also responsible for the generation of polarization. The Sachs–Wolfe source term $\mathcal{G}_{tSW}^{(S)}$ is given by eq. (302); $\mathcal{G}_{Q,U,V,SW}^{(S)}$ all vanish. Since $\Delta_V^{(S)} = 0$ in the early Universe and there is no coupling through $\mathcal{G}_{VC}^{(S)}$ nor through gravity to excite it, it remains zero and an evolution equation for V is unnecessary. Although $\Delta_U^{(S)}(\hat{q}, \mathbf{k}, \tau)$ also vanishes, hence the power spectrum $d\mathcal{C}_{U\ell}^{(S)}/d\ln k = 0$, $\Delta_U^{(S)}(\hat{q}, \mathbf{x}, \tau)$ does not vanish since it appears when one rotates from the polarization basis fixed by $\hat{q}, \hat{k}$ to one defined relative to sky coordinates: $\Delta_U^{(S)}(\hat{q}, \mathbf{x}, \tau)$ is a random field determined from the nonzero power spectrum $\frac{1}{2} d\mathcal{C}_{Q\ell}^{(S)}/d\ln k$. (See eq. (326).)

C.2.2 The moment equations for photons

The moment equations are explicitly (for flat universes, see section C.4 for nonflat modifications):

$$\begin{aligned} \ell = 0 \quad & \dot{\Delta}_{t0}^{(S)} + k\Delta_{t1}^{(S)} = -\dot{\varphi} - \frac{1}{3} k^2 \bar{a}^{-1} \Psi_\sigma, \\ \ell = 1 \quad & \dot{\Delta}_{t1}^{(S)} - k\left(\frac{1}{3}\Delta_{t0}^{(S)} - \frac{2}{3}\Delta_{t2}^{(S)}\right) \\ & = k\frac{1}{3}\nu - n_e \sigma_T a \left(\Delta_{t1}^{(S)} - \frac{1}{3}\hat{k} \cdot i\mathbf{v}_B^{(S)}\right), \\ \ell = 2 \quad & \dot{\Delta}_{t2}^{(S)} - k\left(\frac{2}{5}\Delta_{t1}^{(S)} - \frac{3}{5}\Delta_{t3}^{(S)}\right) \\ & = \frac{2}{15} k^2 \bar{a}^{-1} \Psi_\sigma - n_e \sigma_T a \frac{1}{10} (9\Delta_{t2}^{(S)} - \Delta_{Q0}^{(S)} - \Delta_{Q2}^{(S)}), \\ \ell \geq 3 \quad & \dot{\Delta}_{t\ell}^{(S)} - k\left(\frac{\ell}{2\ell+1}\Delta_{t(\ell-1)}^{(S)} - \frac{\ell+1}{2\ell+1}\Delta_{t(\ell+1)}^{(S)}\right) = -n_e \sigma_T a \Delta_{t\ell}^{(S)}. \end{aligned} \quad (333)$$

The moment equations for the polarization are:

$$\ell = 0 \quad \dot{\Delta}_{Q0}^{(S)} + k\Delta_{Q1}^{(S)} = -n_e \sigma_T a \frac{1}{2} (\Delta_{Q0}^{(S)} - \Delta_{Q2}^{(S)} - \Delta_{t2}^{(S)}),$$

$$\begin{aligned}
\ell = 1 & \quad \dot{\Delta}_{Q1}^{(S)} - k(\frac{1}{3}\Delta_{Q0}^{(S)} - \frac{2}{3}\Delta_{Q2}^{(S)}) = -n_e\sigma_T a \Delta_{Q1}^{(S)}, \\
\ell = 2 & \quad \dot{\Delta}_{Q2}^{(S)} - k(\frac{2}{5}\Delta_{Q1}^{(S)} - \frac{3}{5}\Delta_{Q3}^{(S)}) \\
& \quad = -n_e\sigma_T a \frac{1}{10}(9\Delta_{Q2}^{(S)} - \Delta_{Q0}^{(S)} - \Delta_{t2}^{(S)}), \\
\ell \geq 3 & \quad \dot{\Delta}_{Q\ell}^{(S)} - k\left(\frac{\ell}{2\ell+1}\Delta_{Q(\ell-1)}^{(S)} - \frac{\ell+1}{2\ell+1}\Delta_{Q(\ell+1)}^{(S)}\right) = -n_e\sigma_T a \Delta_{Q\ell}^{(S)}.
\end{aligned} \tag{334}$$

We can rewrite the $\ell = 0, 1$ photon and neutrino moment equations using photon–fluid potentials:

$$\frac{1}{4}\dot{\delta}_\gamma + \dot{\varphi} + \frac{1}{3}k^2\bar{a}^{-1}(\Psi_{v,\gamma} + \Psi_\sigma) = 0, \tag{335}$$

$$\bar{a}^{-1}\dot{\Psi}_{v,\gamma} - \bar{H}\Psi_{v,\gamma} - (\frac{1}{4}\delta_\gamma + \nu) + \frac{1}{6}k^2\pi_{t,\gamma} = -n_e\sigma_T\Psi_{v,\gamma B}, \tag{336}$$

$$\text{relative velocity potential: } \Psi_{v,\gamma B} \equiv \Psi_{v,\gamma} - \Psi_{v,B}. \tag{337}$$

The photon density, velocity potential, isotropic pressure $((\delta p)_\gamma)$ and anisotropic stress $(\pi_{t,\gamma})$ perturbations are related to the low order moments by

$$\Delta_{t0}^{(S)} = \frac{\delta_\gamma}{4} = \frac{(\delta p)_\gamma}{4\bar{p}_\gamma}, \quad \Delta_{t1}^{(S)} = \frac{k\Psi_{v,\gamma}}{3\bar{a}}, \quad \Delta_{t2}^{(S)} = k^2\frac{\pi_{t,\gamma}}{12}. \tag{338}$$

Under $(\Omega = \bar{a})$ -gauge transformations, δ_γ and $\Psi_{v,\gamma}$ can change, but $\pi_{t,\gamma}$, $\Delta_{t\ell}^{(S)}$ for $\ell \geq 2$, $\Delta_{t,\ell \geq 2}^{(S)} \equiv \Delta_t^{(S)} - \Delta_{t0}^{(S)} + 3i\hat{q} \cdot \hat{k}\Delta_{t1}^{(S)}$, and $\Delta_{Q\ell}^{(S)}$ for $\ell \geq 0$ do not.

C.2.3 CDM and baryon transport

These are coupled to the equations for the other types of matter present. The equations for cold dark matter and for the baryons are of the form of eqs. (280), (282), with the proper Thomson scattering coupling included in the latter case.

$$\text{CDM:} \quad \frac{1}{3}\dot{\delta}_{cdm} + \dot{\varphi} + \frac{1}{3}k^2\bar{a}^{-1}(\Psi_{v,cdm} + \Psi_\sigma) = 0, \tag{339}$$

$$\bar{a}^{-1}\dot{\Psi}_{v,cdm} = \nu, \tag{340}$$

$$\text{baryons:} \quad \frac{1}{3}\dot{\delta}_B + \dot{\varphi} + \frac{1}{3}k^2\bar{a}^{-1}(\Psi_{v,B} + \Psi_\sigma) = 0, \tag{341}$$

$$\bar{a}^{-1}\dot{\Psi}_{v,B} = \nu + n_e\sigma_T\bar{a}\frac{4}{3}\frac{\rho_\gamma}{\rho_B}\bar{a}^{-1}\Psi_{v,\gamma B}. \tag{342}$$

Overall momentum conservation of the photon–baryon fluid determines the form of the Compton drag. The baryon pressure and anisotropic stress from electron–ion viscosity have been neglected. In dealing with the combined photon plus baryon system, as well as $\Psi_{v,\gamma B}$, it is useful to consider the equations for the entropy per baryon and for the momentum current of the combined $(\gamma + B)$ -fluid:

$$\text{entropy per baryon: } \delta_{s_\gamma} \equiv \frac{3}{4}\delta_\gamma - \delta_B, \tag{343}$$

$$\dot{\delta}_{s\gamma} + k^2 \bar{a}^{-1} \Psi_{v,\gamma B} = 0, \quad (344)$$

$$\begin{aligned} \bar{a}^{-1} \dot{\Psi}_{v,\gamma B} - \bar{H}(y_B \Psi_{v,\gamma B} + \Psi_{v,(\gamma+B)}) \\ = -(y_B \tau_C)^{-1} \bar{a}^{-1} \Psi_{v,\gamma B} + \frac{1}{4}(\delta_\gamma - \frac{2}{3}k^2 \pi_{t,\gamma}), \end{aligned} \quad (345)$$

$$(\gamma + B) \text{ vel. potential: } \Psi_{\gamma+B} = \frac{(\bar{\rho}_\gamma + \bar{p}_\gamma) \Psi_{v,\gamma} + \bar{\rho}_B \Psi_{v,B}}{(\bar{\rho}_\gamma + \bar{p}_\gamma + \bar{\rho}_B)}, \quad (346)$$

$$\bar{\rho}_B \bar{a}^{-1} \partial_\tau \left[\left(\frac{4}{3} \frac{\bar{\rho}_\gamma}{\bar{\rho}_B} + 1 \right) \Psi_{v,(\gamma+B)} \right] = (\delta p)_\gamma - \frac{2}{3} k^3 \bar{p}_\gamma \pi_{t,\gamma} + \left(\frac{4}{3} \bar{\rho}_\gamma + \bar{\rho}_B \right) \nu,$$

$$\text{where } y_B \equiv \frac{\bar{\rho}_B}{\bar{\rho}_\gamma + \bar{p}_\gamma + \bar{\rho}_B} = \left(\frac{4}{3} \frac{\bar{\rho}_\gamma}{\bar{\rho}_B} + 1 \right)^{-1}. \quad (347)$$

The entropy generation equation, eq. (344), takes the form of a conservation law. The combined momentum current dissipates only because of the viscous anisotropic stresses in $\pi_{t,\gamma} = \pi_{t,(\gamma+B)}$. In computations, we evolve $\Psi_{v,\gamma B}$ instead of $\Psi_{v,\gamma}$. We shall see that $\Psi_{v,\gamma B}$ goes to zero linearly with τ_C at high redshift (section C.3.1). For isocurvature baryon perturbations, $\delta_{s\gamma}$ is a useful variable to solve for [216], while for isocurvature CDM perturbations, the manifestly gauge invariant $\frac{3}{4}\delta_\gamma - \delta_{cdm}$ and $\delta_B - \delta_{cdm}$ are useful for small k , but not for large k [215].

C.2.4 The transport of massless neutrinos

Massless neutrinos and any other freely-streaming extremely relativistic particles (denoted by *fer*) obey the same transport equation as for the photons, except there is of course no Thomson scattering source, and usually the spin (or mixing of neutrino types as in oscillations) does not need to be treated. If they are stable, then, once neutrino scattering is negligible when the temperature drops below an MeV, only the Sachs–Wolfe source term needs to be included. The initial conditions at time τ_i (assumed to be $\ll \tau_{eq}$, i.e., safely in the *er*-dominated regime) for the neutrino distribution function are found by expanding it to order $k\tau_i$, which implies only terms up to $\ell = 3$ are needed, $\Delta_{fer} \approx \sum_{\ell=0}^3 (2\ell+1)(-i)^\ell \Delta_{fer,\ell} P_\ell(\hat{q} \cdot \hat{k})$. For example, for adiabatic perturbations in the synchronous gauge the initial conditions are

$$\begin{aligned} \Delta_{fer}(k, \mu, \tau_i) &\approx \frac{1}{4} \delta_{fer,i} (\beta + 3(1-\beta)\mu^2)(1 - ik\mu\tau_i/3), \\ \Delta_{fer,0}(k, \tau_i) &\approx \frac{1}{4} \delta_{fer,i}, \\ \bar{a}^{-1} \Psi_{v,fer}(k, \tau_i) &= 3\Delta_{fer,1}/k = \frac{9-4\beta}{60} \delta_{fer,i} \tau_i, \\ \Delta_{fer,2} &= \frac{1}{12} k^2 \pi_{t,fer} = -\frac{1-\beta}{10} \delta_{fer,i}, \quad \Delta_{fer,3}/k = -\frac{1-\beta}{70} \delta_{fer,i} \tau_i, \\ \beta &\equiv \frac{5\bar{\rho}_\gamma(\tau_i) + 9\bar{\rho}_{fer,tot}(\tau_i)}{15\bar{\rho}_\gamma(\tau_i) + 19\bar{\rho}_{fer,tot}(\tau_i)}. \end{aligned} \quad (348)$$

Here $\bar{\rho}_{fer,tot}(\tau_i)$ is the total density of extremely relativistic particles that are freely streaming at time τ_i . For three relativistic neutrino species, $\bar{\rho}_{fer,tot}(\tau_i)/$

$\bar{\rho}_\gamma(\tau_i) = 3 \times 2 \times (7/16) \times (4/11)^{4/3} = 0.6813$ and $\beta = 0.3984$. The starting time is assumed to be $\ll \tau_{eq}$, i.e., safely in the *er*-dominated regime. These initial conditions contrast with those for the tightly coupled photons: with $\tau_C(\tau_i) \approx 0$, we have $\Delta_{\gamma,t,0} = \frac{1}{4}\delta_{\gamma,i}$, $\Psi_{v,\gamma B,i} \approx 0$, $\bar{a}^{-1}\Psi_{v,\gamma} \approx \bar{a}^{-1}\Psi_{v,B} \approx \frac{1}{12}\delta_{\gamma,i}\tau_i$, and $\Delta_{\gamma,t,\ell} \approx 0$ for $\ell \geq 2$. As well, $\delta_{s_\gamma} \approx 0$, hence $\delta_{B,i} \approx \frac{3}{4}\delta_{\gamma,i}$. For CDM, the defining condition of the synchronous gauge is $\Psi_{v,cdm} = 0$, and the density perturbation starts off the same as that for baryons, $\delta_{cdm,i} = \frac{3}{4}\delta_{\gamma,i}$. The initial conditions for the metric variables in this gauge are $\dot{\varphi} = -\frac{1}{2}\beta\delta_{\gamma,i}\tau_i^{-1}$, $k^2\bar{a}^{-1}\Psi_\sigma = -\frac{3}{2}(1-\beta)\delta_{\gamma,i}\tau_i^{-1}$. (The initial conditions for the relativistic neutrinos follow from expanding the past-history integration, eq. (350) given below.)

C.2.5 Hot and warm dark matter transport

For scalar perturbations and hot or warm dark matter

$$\begin{aligned} \partial_\tau[\Delta_{hdm}] + i\frac{q}{q^n}k\hat{q} \cdot \hat{k} \frac{\bar{N}}{\bar{a}}\Delta_{hdm} &= (\mathcal{G}_{hdm\,SW} + \mathcal{G}_{hdm\,curv}), \\ \mathcal{G}_{hdm\,SW}^{(S)} &= -i\frac{q^n}{q}\hat{q} \cdot \mathbf{k}\nu - \dot{\varphi} - (\hat{q} \cdot \mathbf{k})^2\bar{a}^{-1}\Psi_\sigma, \quad q^n = \sqrt{q^2 + m^2\bar{a}^2}. \end{aligned} \quad (349)$$

It is the semi-relativistic stage, when q/q^n is not simply unity or $q/(m\bar{a})$, that creates the difficulty. Thus, it is perhaps worthwhile to make a brief aside on the numerical methods used in [134, 2, 233, 253, 259, 255, 260, 261, 262] to solve collisionless damping equations for semi-relativistic particles. Just as for photons, a hierarchy of moment equations can be written for $\Delta_{hdm,\ell}, \Delta_{wdm,\ell}$. For massless neutrinos, the moment expansion became our preferred method in [134]. For hot and warm dark matter, the number of equations to be solved is the product of the number of multipoles that are being followed times the number of momentum groups, which then must be summed over with appropriate weights to get the neutrino stress-energy tensor for the source side of Einstein's equations. In [195], we described an efficient Gauss-Legendre integration method using as the integration variable g , where $dg = -(x\partial\bar{f}/\partial x)x^2dx$, $x \equiv q/(\bar{a}\bar{T}_{hdm})$, which gives all momentum groups significant weights in the energy group sum: 24 groups give accurate results. The moment expansion with truncation is called the “*P-N*” method, and it requires many moments to give accurate results and so is expensive. However, with a suitable boundary condition in ℓ space, Lithwick and I [261] showed that high precision can be achieved with just 20 moments for high k and 10 moments for low k , a very modest numerical cost. One can also lower the number of multipoles to 2 in the very nonrelativistic regime. This seems to be the best approach to this problem now [260, 261, 262]. Another method is to discretize the BTE in angle as well as energy, which reduces the problem to a set of ODEs with total number equal to the number of momentum groups times the number of angular bins. Durrer [259] used this “*S-N*” method to solve for massive neutrino transport.

In Bond and Szalay [195] and subsequent work on hot and hot/cold models [134, 233, 2], we adopted a history-integration method. Since this is very

different than the moment approach I shall discuss it in a little detail. We are not interested in the detailed angular distribution of neutrinos as we are for photons, but only in the density, pressure, velocity, anisotropic stress and the action these have upon the metric variables. The transport equation can be integrated and low order moments taken, which expresses the result in terms of momentum integrals of spherical Bessel functions with momentum-dependent arguments:

$$\begin{aligned}\Delta_{hdm,\ell}(q, k, \tau) = & \mathcal{D}_{hdm,\ell}(q, k, \tau) - \int_{\tau_i}^{\tau} d\tau' \left\{ \dot{\varphi}(k, \tau') j_{\ell}(k\Delta\eta) \right. \\ & + k^2 \bar{a}^{-1} \Psi_{\sigma}(k, \tau') \left(\frac{2\ell(\ell+1)-1}{(2\ell-1)(2\ell+3)} j_{\ell}(k\Delta\eta) \right. \\ & - \frac{(\ell-1)\ell}{(2\ell-1)(2\ell+1)} j_{\ell-2}(k\Delta\eta) - \frac{(\ell+1)(\ell+2)}{(2\ell+1)(2\ell+3)} j_{\ell+2}(k\Delta\eta) \Big) \\ & \left. - k \frac{q^n}{q} \nu(k, \tau') \left(\frac{\ell}{(2\ell+1)} j_{\ell-1}(k\Delta\eta) - \frac{\ell+1}{(2\ell+1)} j_{\ell+1}(k\Delta\eta) \right) \right\},\end{aligned}\tag{350}$$

$$\begin{aligned}\Delta\eta = & \eta(q, \tau) - \eta(q, \tau'), \quad \text{where } \eta(q, \tau) = \int_0^{\tau} d\tau' \frac{q}{\sqrt{q^2 + m^2 \bar{a}^2}}, \\ \bar{\eta} = & \int_0^{\tau} d\tau' \frac{\langle q \rangle}{\sqrt{\langle q \rangle^2 + m^2 \bar{a}^2}}, \quad \langle q^m \rangle = \frac{\int q^{2+m} dq \bar{f}}{\int q^2 dq \bar{f}}, \quad m = 1, 2, \dots, \\ \frac{\langle q \rangle}{\bar{a} T_{hdm}} = & \frac{7\zeta_4}{2\zeta_3} = 3.151 \text{ fermions}, \quad \frac{3\zeta_4}{\zeta_3} = 2.701 \text{ bosons}.\end{aligned}$$

The explicit numbers for the average momenta assume unperturbed $\bar{f} = (e^{q/(\bar{a}\bar{T})} \pm 1)^{-1}$ distributions for the light fermions and bosons. Recall that for light neutrinos $\bar{a}\bar{T}_{m\nu} = 1.95$ K. The $\mathcal{D}_{hdm,\ell}(q, k, \tau)$ describe the evolution of the initial conditions. For example, in the synchronous gauge for an adiabatic mode we have $\nu = 0$ and [195]

$$\begin{aligned}\mathcal{D}_{hdm,0}(q, k, \tau) = & \frac{1}{4} \delta_{hdm}(q, k, \tau_i) [j_0(k\Delta\eta_i) - 2(1-\beta)j_2(k\Delta\eta_i)], \\ \mathcal{D}_{hdm,1}(q, k, \tau) = & \frac{1}{4} \delta_{hdm}(q, k, \tau_i) \left[\frac{9-4\beta}{5} (j_1(k\Delta\eta_i) \right. \\ & \left. + \frac{1}{9} k \tau_i j_0(k\Delta\eta_i)) - \frac{6}{5} (1-\beta) j_3(k\Delta\eta_i) \right], \\ \Delta\eta_i \equiv & \eta(q, \tau) - \eta(q, \tau_i).\end{aligned}\tag{351}$$

The complication in this equation (350) is the integral over past time τ' of the metric variables, turning the metric ODEs into integro-differential equations, not by itself a great numerical problem, but for speed some care is needed to efficiently yet fully sample the past history. In [261] we show adaptive (Romberg) integration makes this method competitive in accuracy and

numerical cost with the moment method. Even with less efficient sampling, the speedup in the [134] neutrino code, which was also applied to hot/cold hybrid models in [2, 233], was considerable. Past-history approaches are now also being used to great advantage for rapid computation of $\Delta_{t,\ell}$ for the radiation [306]. The Sachs-Wolfe metric part of this is similar to eq. (350), except the optical depth $\exp[-\zeta_C]$ enters in the obvious way. The Compton terms associated with the source $\mathcal{G}_{tC}^{(S)}$ have similar j_ℓ expansions, but now the low order moments of $\Delta_t^{(S)}$ enter into the integral. One of the features of the past-history method is that one does not have to calculate $\Delta_{t,\ell}^{(S)}$ at every ℓ , whereas this is necessary because of the way the equations are coupled for the moment method.

For the hot or warm dark matter, one can also save by shifting into $P - 1$ equations once the particles are strongly nonrelativistic and the wavenumber is much below the Jeans length, $k_{Jhdm}(a) = (4\pi G_N \bar{\rho}_{nr} \bar{a}^2 / c_{s,hdm}^2)^{1/2}$, $\sim \bar{a}^{1/2}$, where $c_{s,hdm} \sim \bar{a}^{-1}$ is the adiabatic sound speed. I shall sketch this since it exercises some of the equations derived earlier. The energy and momentum conservation laws for hdm , eqs. (278), (279) with no source terms, $\mathcal{S}_{hdm} = 0$, are generally valid of course, but to close off the equations, a model is required for the pressure fluctuation $(\delta p)_{hdm}$ and the anisotropic stress $\pi_{t,hdm}$. This relation is complex because of collisionless damping, but at late times the hdm obeys an equation similar to cdm . To roughly model residual effects of the random velocity dispersion, we can introduce fudge factors akin to variable Eddington factors to close off the hierarchy:

$$\begin{aligned}
(\delta p)_{hdm} &= C_{p,hdm} \bar{p}_{hdm} \delta_{hdm}, \\
\left(\frac{2}{3} \overline{\nabla}^2 + \frac{1}{3} {}^{(3)}\overline{R}\right) \pi_{t,hdm} &= C_{\pi_t,hdm} \delta_{hdm}, \\
\dot{\delta}_{hdm} + 3\dot{\varphi} + k^2 \bar{a}^{-1} (\Psi_{v,hdm} + \Psi_\sigma) + 3\bar{H} (C_{p,hdm} - 1) \frac{3}{5} c_{s,hdm}^2 \delta_{hdm} &= 0, \\
\bar{a}^{-1} \dot{\Psi}_{v,hdm} &= \nu + (C_{p,hdm} + C_{\pi_t,hdm}) \frac{3}{5} c_{s,hdm}^2 \delta_{hdm}, \\
\text{“adiabatic sound” speed: } c_{s,hdm}^2 &\equiv \frac{5}{3} \frac{\bar{p}_{hdm}}{\bar{\rho}_{hdm}} = \frac{5}{3} \frac{\langle q^2 \rangle}{3m_{hdm}^2 \bar{a}^2}, \\
\frac{c_{s,hdm}}{\frac{\langle q \rangle_{hdm}}{m_{hdm} \bar{a}}} &= \left(\frac{20\eta_5 \eta_3}{27\eta_4^2} \right)^{1/2} = 0.85 \text{ fermions}, \quad \left(\frac{20\zeta_5 \zeta_3}{27\zeta_4^2} \right)^{1/2} = 0.89 \text{ bosons}; \\
P - 1 \text{ eqs. are used for } c_{s,hdm}(a) &< \text{TOL}_{vnr}, \quad k < \text{TOL}_J \quad k_{Jhdm}(a), \\
\text{and } (C_{p,hdm} + C_{\pi_t,hdm}) &\text{ is set to } 5/3.
\end{aligned} \tag{352}$$

The fudge factor is arbitrary: the 5/3 choice is arranged so that it is the adiabatic sound speed, $c_{s,hdm}$, rather than the isothermal sound speed, because compressing neutrinos that are gravitationally bound would be better approximated this way. (The Riemann eta and zeta values are η_j , ζ_j .) The pdV term on the neutrino energy density, $\sim \bar{H} (C_{p,hdm} - 1) c_{s,hdm}^2 \delta_{hdm}$, is not important. Although one could get more sophisticated by better modelling $C_{p,hdm}$, $C_{t,hdm}$ and thus the damping (see [300] for a nice nr analytic model

of neutrino-damping), it is better to deal with the damping by the full past-history integration or with a hierarchy of moments. Thus the tolerance factors are chosen to be quite conservative. The “very nonrelativistic” tolerance factor, TOL_{vnr} , should be quite small (< 0.05) and the Jeans tolerance factor, TOL_J , should be at most a tenth.

C.3 Numerically useful regimes for scalar perturbations

C.3.1 Tight-coupling, shear viscosity and thermal diffusion

Tight coupling equations adequately approximate the hierarchy prior to a (k -dependent) redshift $z_{tc}(k)$. These are obtained by first developing a two-fluid treatment of the photon–baryon interaction, which is adequate provided the Compton timescale τ_C is short compared with all other timescales in the problem, in particular, the light-crossing time across half a wavelength πk^{-1} , and the Hubble time at that epoch. In the [134, 88] code, we choose z_{tc} to be at least 2000, and also required that $k\tau_C < 0.01$ and $\bar{H}\bar{a}\tau_C < 0.01$ to remain in tight coupling: the results are insensitive to considerable relaxation of this criterion.

Two-fluid equations are obtained from the infinite hierarchy of moment equations (eq. 333) by setting $\Delta_{t3}^{(S)}$ to zero in the $\ell = 2$ equation, thereby truncating the hierarchy, and neglecting $\dot{\Delta}_{t2}^{(S)}$. The polarization is also assumed to change quickly enough so that $\Delta_{Q0}^{(S)}$ and $\Delta_{Q2}^{(S)}$ are in the steady state found by setting the right-hand side of the $\ell = 0$ and $\ell = 2$ equations to zero:

$$\Delta_{Q0}^{(S)} = \frac{5}{4}\Delta_{t2}^{(S)}, \quad \Delta_{Q2}^{(S)} = -\frac{1}{4}\Delta_{t2}^{(S)}. \quad (353)$$

Thus, the $\ell = 2$ equation fixes the anisotropic stress (total quadrupole anisotropy):

$$\frac{1}{6}\pi_{t,\gamma} = k^{-2}2\Delta_{t2}^{(S)} = \frac{4}{15f_\eta}\tau_C\bar{a}^{-1}(\Psi_{v,\gamma} + \Psi_\sigma) \quad (354)$$

$$= \frac{4}{15f_\eta}\tau_C\bar{a}^{-1}(\Psi_{v,(\gamma+B)} + \Psi_\sigma) - \frac{4}{15f_\eta}\tau_C c_{s(\gamma+B)}^2\bar{a}^{-1}\Psi_{v,\gamma B},$$

$$f_\eta = \frac{3}{4} \text{ with polarization, } f_\eta = \frac{9}{10} \text{ without,}$$

$$f_\eta = 1 \text{ isotropic, no polarization,}$$

$$\text{shear viscosity: } \eta_\gamma = \frac{4}{15f_\eta}\rho_\gamma\bar{a}\tau_C,$$

$$\text{sound speed: } c_{s(\gamma+B)}^2 = \frac{c^2}{3} \frac{1}{1 + \frac{3}{4}\frac{\rho_B}{\rho_\gamma}},$$

$$\text{thermal diffusion: } \kappa_\gamma = \frac{4}{3}\frac{\rho_\gamma}{T_\gamma}\bar{a}\tau_C.$$

Identifying this result with the form of eq. (236) for scalar perturbations gives the photon’s shear viscosity η_γ . The photon kinematic viscosity is $\eta_\gamma/(\bar{\rho}_\gamma + \bar{p}_\gamma)$,

hence is $(5f_\eta)^{-1}\bar{a}\tau_C$. Weinberg's classic text [264, 265] gives the $f_\eta = 1$ result. (He restricted himself to the approximation that Compton scattering was angle and polarization independent.) The bulk viscosity for photons vanishes. To determine the thermal diffusion coefficient, we must identify terms in these equations with the defining relation for κ_γ , eq. (236), which involves the fluid acceleration as well as the temperature gradient. The temperature gradient is a projected one, so that it has no component in the direction of the fluid's velocity U ; i.e., it reduces to a spatial gradient in the fluid's comoving frame. In the frame defined by the time-surface velocity e_n^α , it picks up a time component: $\perp_{(U)I}^b ({}^{(4)}\nabla_b[\ln T] + \mathcal{A}_{(U)b})$ is ${}^{(3)}\nabla_I[\ln T] + U^n \bar{e}_n[U_I]$. For scalar photon perturbations, this reduces to ${}^{(3)}\nabla_I[\frac{1}{4}\delta_\gamma + \bar{H}\Psi_{v,\gamma}]$; thus,

$$(J_{(e)\gamma})_I = -\kappa_\gamma \bar{T}_\gamma {}^{(3)}\nabla_I(\frac{1}{4}\delta_\gamma + \bar{H}\Psi_{v,\gamma} + \nu - \bar{a}^{-1}\dot{\Psi}_\gamma)$$

defines the combination we are looking for. With appropriate multiplication by $\bar{\rho}_\gamma$ to relate to the energy conservation equation, we get Weinberg's [265] result for κ_γ : the photon entropy per unit volume times τ_C . Of course, it is unaffected by f_η which arises in the anisotropic shear stress. Indeed, the potential for $(J_{(e)\gamma})_I$ is just $\frac{4}{3}\bar{\rho}_\gamma(\Psi_{v,\gamma B} + (\bar{a}\tau_C)^{-1}\frac{1}{6}k^2\pi_{t,\gamma})$, i.e., basically $\Psi_{v,\gamma B}$, but with the anisotropic stress contribution to it removed.

The two-fluid equations are obtained from the $\ell = 0$ and $\ell = 1$ equations, and the baryon mass and momentum conservation equations. They are eqs. (341), (342), (335) and the $\dot{\Psi}_{v,\gamma B}$ equation, (345), with $\pi_{t,\gamma}$ substituted into it. Alternatively one can use the $\dot{\delta}_{s_\gamma}$ and $\dot{\Psi}_{v,(\gamma+B)}$ equations instead of the $\dot{\delta}_B$ and $\dot{\Psi}_{v,B}$ equations.

The tight coupling equations are a one-fluid ($\gamma + B$) approximation in which the two-fluid character is encoded in diffusion and viscosity coefficients. They are obtained by creating a "constitutive relation" for $\Psi_{v,\gamma B}$ by expanding $y_B\tau_C\bar{a}^{-1}\dot{\Psi}_{v,\gamma B}$ in eq. (345) in powers of $y_B\tau_C$. Even if one is only interested in first order τ_C effects in the evolution equations, the expansion of $\Psi_{v,\gamma B}$ must go to quadratic order:

$$\begin{aligned} \bar{a}^{-1}\Psi_{v,\gamma B} &= -y_B k^2 \tau_C^2 \frac{4}{15f_\eta} \bar{a}^{-1}(\Psi_{v,B} + \Psi_\sigma) \\ &\quad + y_B \tau_C (\frac{1}{4}\delta_\gamma + \bar{H}\Psi_{v,B}) [1 - (3 - y_B + p_e)y_B \tau_C \bar{H}\bar{a}] \\ &\quad + y_B^2 \tau_C^2 ((q+1)(\bar{H}\bar{a})^2 \bar{a}^{-1}\Psi_{v,B} + \frac{1}{3}k^2 \bar{a}^{-1}\Psi_{v,B} + \bar{a}(\delta H)), \quad (355) \\ (\gamma + B) \text{ kinematic shear viscosity: } &= \frac{\eta_\gamma}{\frac{4}{3}\rho_\gamma + \rho_B} = \frac{3}{5f_\eta} c_{s(\gamma+B)}^2 \bar{a}\tau_C, \\ (\gamma + B) \text{ sound damping rate: } \Gamma &= \frac{1}{2}\bar{a}^{-1}k^2 c_{s(\gamma+B)}^2 \tau_C \left(\frac{4}{5f_\eta} + \frac{3\bar{\rho}_B}{4\bar{\rho}_\gamma} y_B \right), \end{aligned}$$

$$\text{where } q = -\frac{d\ln(\bar{H}\bar{a})}{d\ln\bar{a}}, \quad p_e \equiv -d\ln Y_e/d\ln a.$$

The term in $\bar{a}^{-1}\Psi_{v,\gamma B}$ of order $y_B\tau_C^2$ is from the shear viscosity, while the $y_B\tau_C$ and $(y_B\tau_C)^2$ terms together are the thermal diffusion contributions (the

nonnegligible $(y_B \tau_C)^2$ terms come from the fluid acceleration which enters the thermal diffusion expression). Three tight coupling equations are then to be solved: for entropy generation, $\dot{\delta}_{s,\gamma}$, and for $(\gamma + B)$ mass and momentum conservation. In practice, we solve the $\dot{\delta}_\gamma$ equation and the $\dot{\Psi}_{v,B}$ equation, (342), instead of (346).

We saw in section 5.2.1 that the transfer equation can be recast in terms of $\mathcal{T}_C \tilde{\Delta}_t$, with $\mathcal{T}_C$ the Compton transparency, and leading sources $\mathcal{V}_C \frac{1}{4} \tilde{\delta}_\gamma$, $-\tilde{q}^i {}^{(3)}\nabla_i \mathcal{V}_C \bar{a}^{-1} \tilde{\Psi}_{v,B}$, and the integrated Sachs–Wolfe term. Because the source term $\tilde{\mathcal{G}}_{tSW}^{(S)}$ (eq. (304)) vanishes if the gravitational potential is constant, it is sometimes useful to regroup the $\ell = 0, 1$ photon transport equations, and the 2-fluid and tight-coupling equations following from it, to exploit this in searching for analytic solutions:

$$\begin{aligned} \frac{1}{4} \partial_\tau [\tilde{\delta}_\gamma] + \frac{1}{3} k^2 \bar{a}^{-1} \tilde{\Psi}_{v,\gamma} &= \tilde{\mathcal{G}}_{tSW}^{(S)}, \\ \partial_\tau [\bar{a}^{-1} \tilde{\Psi}_{v,\gamma}] + \frac{1}{4} \tilde{\delta}_\gamma + \frac{1}{6} k^2 \pi_{t,\gamma} &= -n_e \sigma_T \Psi_{v,\gamma B}. \end{aligned} \quad (356)$$

(To put all density and velocity perturbations on the same footing, we should also transform the momenta of baryons, massive and massless neutrinos, cold, warm and hot dark matter to $\tilde{q}^I$, which yields gauge-invariant quantities; e.g., $\frac{1}{3} \tilde{\delta}_B = \frac{1}{3} \delta_B + \nu + \partial_\tau [\bar{a}^{-1} \Psi_\sigma]$, $\tilde{\Psi}_{v,B} \equiv \Psi_{v,B} + \Psi_\sigma$.)

A WKB analysis of these equations gives the usual damped oscillation behavior, e.g., [130, 265, 264, 8]. For a mode of wavenumber k , we have

$$\frac{1}{4} \tilde{\delta}_\gamma \propto (1 - y_B)^{1/4} e^{-i \int k c_{s,(\gamma+B)} d\tau} e^{-\int \Gamma \bar{a} d\tau}$$

(if metric terms are ignored). The $4/(5f_\eta)$ is from viscous shear, while the smaller $\frac{3\bar{\rho}_B}{4\bar{\rho}_\gamma} y_B$ part is due to the thermal diffusion. (Without the acceleration correction, $\dot{\Psi}_{v,\gamma B}$, in eq. (355), the correct y_B multiplier in the diffusion term of the damping rate is not obtained.) The Silk damping scale factor σ_D defined in [2] is related to Γ by $\sigma_D^2 = (k\tau_{dec})^{-2} \int_0^{\tau_{dec}} \Gamma \bar{a} d\tau$. The integrand $\sim \bar{a}^{p_e+5/2} d \ln \bar{a}$ is a very steeply rising function, so to truncate the integral at precisely a_{dec} (defined by eq. (71)) will provide only a rough estimation of the overall factor; assuming the region near decoupling dominates so we take $p_e \approx p_{e,dec}$, which is also a crude approximation,

$$\sigma_D^2 \approx \frac{1}{15f_\eta} \frac{1}{(p_e + \frac{5}{2})(p_e + 2)} \frac{(1 + 15f_\eta \bar{\rho}_B y_B / (16\bar{\rho}_\gamma))}{(1 + 3\bar{\rho}_B / (4\bar{\rho}_\gamma))}. \quad (357)$$

With $f_\eta = 3/4$ and $7 \lesssim p_{e,dec} \lesssim 12$ (from fig. 3(c)), we have $0.02 \lesssim \sigma_D \lesssim 0.03$ in the small Ω_B limit. For the low Ω_B values inferred from nucleosynthesis, the shear viscosity is by far the dominant damping term, and $\sigma_D \propto f_\eta^{-1/2}$: the inclusion of polarization therefore results in a 10% increase in σ_D over the value that is obtained if polarization is not included. Thus, apart from the intrinsic interest in polarization [266, 134, 88], it is clearly important to include it because of the enhanced damping.

The photon–baryon fluid equations are coupled to those for CDM, the Boltzmann equations for massless neutrinos and massive neutrinos or warm dark matter if applicable, and the metric equations to obtain Ψ_σ and $\dot{\varphi}$. In a typical run, we begin evolution when the waves are far outside the horizon and also in the relativistic dominated regime (so the initial conditions can be integrated analytically). For the photons and baryons, we start off with the tight coupling equations. For massless neutrinos, we solve for typically 40 moments, and shut them off once the energy density becomes negligible. After $z_{tc}(k)$, we solve the full moment hierarchy equations up to some $\ell_{max}(\tau)$, which we increase according to an algorithm based on a monitor of the radiation power in high ℓ -modes (ℓ_{max} scales with $k\tau$.) Special care must be taken with the computational procedure and time-stepping through recombination. We either integrate the full equations forward to the present – which can often mean we are just generating Bessel functions by ODE solvers, not the most straightforward nor accurate method, or, for many models and wavenumbers, we can do this more accurately in a single step using free-streaming equations. Before turning to these, it is worthwhile to note how well one can do with just two fluid or tight coupling equations.

Two-fluid and tight coupling equations dominated theoretical explorations of the sixties, seventies and even into the eighties, with the notable exception of the expansion of the transport equation in angular bins by Peebles and Yu [130] and unpublished earlier work using a moment expansion by Bardeen. For example, among others, [267, 268, 143] used two-fluid models to calculate transfer functions for matter, for which it is often quite accurate, and to estimate CMB anisotropies, which Seljak [143] has recently shown to come reasonably close to a full transport solution. Tight coupling equations have not only been used to begin full transport calculations when the equations are very stiff. They have also been used successfully in calculating transfer functions, estimating Silk damping by WKB solutions for baryon-dominated models, and have also been extensively used to make estimates for $\mathcal{G}_0, \mathcal{G}_1$ used in the approximate equations of section 5.1, e.g., [131, 2, 269, 270, 271]. Keeping only the tight coupling equations to lowest order in τ_C has a solution which can be expressed in terms of hypergeometric functions [269] plus a special solution of the inhomogeneous equation driven by the metric variables. Since hypergeometric functions are not very useful for calculations, searching for WKB solutions is usually more profitable. In [2], I used the approximation of (1) tiny ρ_B , (2) a constant gravitational potential ν_L through photon decoupling, a limiting case for CDM-dominated universes, to elucidate the role that the Sachs–Wolfe effect, electron bulk flows and photon compression had on the development of anisotropy in both adiabatic and isocurvature models. Doroshkevich, Starobinsky and collaborators included finite ρ_B effects. None of our results could be viewed as particularly accurate, especially at high ℓ . Hu and Sugiyama [271] explicitly included models for the metric variables in the finite but small ρ_B cases and showed that one can use the $\mathcal{G}_0, \mathcal{G}_1$ results to get the spectra even at high ℓ to within about 10% accuracy. Of course, if back action of the photon

and baryons upon the metric variables becomes important, this will not work so well. These semi-analytical methods still require computation of Bessel functions to large order to get the $\mathcal{C}_\ell$'s, but this is quite fast. However, with current computing power, full Boltzmann transport calculation of an entire model runs very quickly, measured in hours on DEC alphas for enough k -mode coverage to get good accuracy.

C.3.2 Free-streaming

Either by direct rearrangement of the transfer equation or by the use of inhomogeneous momentum transformations of the sort used to get eq. (304), we can rearrange terms in the transport equation to give the modified distribution and source

$$\begin{aligned}\tilde{\Delta}_t &= \Delta_t + \nu + \partial_\tau(\bar{a}^{-1}\Psi_\sigma) - \hat{q}^{i(3)}\bar{\nabla}_i\bar{a}^{-1}\Psi_\sigma - n_e\sigma_T\Psi_{v,\gamma B}, \\ \tilde{\mathcal{G}}_{tSW}^{(S)} + \tilde{\mathcal{G}}_{tC}^{(S)} &= \dot{\nu} - \dot{\varphi} + \frac{\partial^2\bar{a}^{-1}\Psi_\sigma}{\partial\tau^2} - \frac{\partial n_e\sigma_T\Psi_{v,\gamma B}}{\partial\tau} \\ &\quad - n_e\sigma_T\bar{a}\frac{1}{2}P_2(\hat{k}\cdot\hat{q})(\Delta_{t2} + \Delta_{Q2} + \Delta_{Q0}) - n_e\sigma_T\bar{a}\Delta_{t,\ell\geq 2}.\end{aligned}\quad (358)$$

The distribution function perturbation $\tilde{\Delta}_t$ is gauge invariant. (Taking the $n_e\sigma_T\Psi_{v,\gamma B}$ term into the modified distribution function is not really important, since it is expected to fall to zero quite quickly after photon decoupling, especially for normal recombination. With this form, the Thomson source falls even more quickly to zero.) In [88], we identified this variable, expressed in terms of $\bar{a}\delta H$ for use when $\dot{\varphi}$ was negligible, as the one of relevance for free-streaming (but without the $n_e\sigma_T\Psi_{v,\gamma B}$ term.)

If the metric source terms become small beyond some time $\tau_s(k)$ (redshift $z_s(k)$), the radiation free-streams:

$$\tilde{\Delta}_t(\mathbf{k}, \hat{q}, \tau) = e^{-i\hat{q}\cdot\mathbf{k}(\tau-\tau_s(k))}\tilde{\Delta}_t(\mathbf{k}, \hat{q}, \tau_s(k)). \quad (359)$$

The numerical output at redshift $z_s(k)$ is $\Delta_{t\ell}(k, \tau_s)$, including multipoles up to $L_{max}(\tau_s)$, from which $\tilde{\Delta}_{t\ell}$ can be constructed. Expanding the plane waves $e^{-i\mu k(\tau-\tau_s(k))}$ in terms of Legendre polynomials and $j_\ell(k(\tau-\tau_s(k)))$ and integrating over μ gives a direct relation involving Clebsch–Gordon coefficients that allows one to get $\Delta_{t\ell}(k, \tau_0)$ at the present time τ_0 in a single step (Eq. (4.5) of [88]):

$$\tilde{\Delta}_{t\ell}(\mathbf{k}, \tau_0) = \sum_{\ell' L} (-1)^{(L+\ell'-\ell)/2} \langle \ell 0 \ell' 0 | L 0 \rangle^2 (2\ell' + 1) j_L(k\chi_s) \tilde{\Delta}_{t\ell'}(\mathbf{k}, \tau_s). \quad (360)$$

The Clebsch–Gordon coefficients, $\langle \ell m \ell' m' | L M \rangle$, use standard notation (e.g., deShalit and Feshbach [263]). Note that $|\ell - \ell'| \leq L \leq \ell + \ell'$ and L must be even (odd) if $\ell + \ell'$ is even (odd). The spherical Bessel sum has to go to very high L , to $L_{max}(\tau_s) + L_{max}(\tau_0)$. The former may be only a few hundred, but

the latter will be at least 3000. Spherical Bessel functions can be evaluated to $\ell \sim 6000$ and higher with accuracy using Miller's method on a recursion relation.

As we have seen in eq. (261), $\nu_L - \varphi_L = \ddot{\psi}_S - \varphi_S$. The time derivatives of these terms go to zero when the anisotropic stress becomes negligible (so $\nu_L = -\varphi_L$) and the gravitational potential becomes constant: this occurs for $\Omega_{nr} = \Omega = 1$ universes well after radiation-matter equality τ_{eq} . We would typically take $z_s(k) \sim 100$ for standard CDM models with normal recombination, although for accuracy long waves are integrated to the present, which is a trivial computational burden; in reionized models there is persistent damping down to low redshifts so a z_s near or at the present is needed for accuracy for high k as well. (The streaming formula can be modified to take the dominant damping effect into account.) The potential terms $\dot{\nu}_L - \dot{\varphi}_L$ are nonzero at late times if the universe becomes vacuum-dominated [110], but these effects have little influence on high k 's, so although for low k 's one evolves the equations forward to the present, for high k 's one can still use the free-streaming prescription.

As a last aspect of this free-streaming, we described “small-angle” approximations in section 5.1.4 that have been used to speed up the evaluation of correlation functions and power spectra in the past; they are not used anymore for primary anisotropies because the techniques and computing are well in place for doing full Boltzmann transport. A conceptually useful way of thinking of the free-streaming transport which connects to section 5.1 is to treat the radiation pattern itself as the source, with a delta function visibility at some time τ_s :

$$\mathcal{G}(q, \hat{q}, \mathbf{k}, \tau) = \mathcal{V}(\tau) \tilde{\Delta}_t(q, \hat{q}, \mathbf{k}, \tau_s), \quad \mathcal{V}(\tau) = \delta(\tau - \tau_s).$$

With $\mathcal{G}$ so defined, one just applies eq. (129) with either $\mathcal{P}_{\overline{\mathcal{G}}}(k; \chi_s, 0)$ being proportional to $\langle |\tilde{\Delta}_t(k_\ell, \mu_\ell, \tau_s)|^2 \rangle$, where $\mu_\ell \equiv k_\parallel / k_\ell$ and $k_\ell \equiv \sqrt{\frac{Q^2}{\mathcal{R}(\chi_s)^2} + k_\parallel^2}$ — a DSZ [131] style approximation — or one isotropizes, with $\mathcal{P}_{\overline{\mathcal{G}}}$ proportional to $W_t^2(k; \tau_s) = \sum_{\ell \geq 2} (2\ell + 1) \langle |\tilde{\Delta}_{t\ell}(k, \tau_s)|^2 \rangle$, a nearly-conserved quantity which is what the second approximation method exploits. For example, the way it was used in [88] for CDM-type models was to integrate the Boltzmann equations down to redshift $z_s = 200$ or so, construct $W_t^2(k; \tau_s)$, then use eq. (129) to get $\mathcal{C}_\ell$ (e.g., fig. 7 of [88]).

C.4 Modifications with mean curvature

In the seventies and eighties, when approximate methods were still being heavily used for anisotropy calculations, it was usual to free-stream the radiation from an early time when the curvature was unimportant to now using flat model results, but with an angle-distance relation appropriate for the curved model, eq. (130). The results for open CDM and isocurvature baryon models were then used to constrain parameters with data from the small and intermediate angle CMB experiments of the time, e.g., [134, 135, 216, 243].

Now the calculations are being done precisely. When there is mean curvature, one cannot expand in plane waves. The modes $Q_{k\mathcal{M}}$ are eigenfunctions, $-\bar{a}^2 {}^{(3)}\bar{\nabla}^2 Q_{k\mathcal{M}} = k^2 Q_{k\mathcal{M}}$, of the background Laplacian. Although plane waves are not solutions for curved FRW spaces, spherical waves $\propto Y_{\ell m}$ are solutions, with radial wavefunctions $X_{k\ell}(\chi/d_{\text{curv}})$ which go to spherical Bessel functions $j_\ell(k\chi)$ when k is large compared with d_{curv}^{-1} , and which, like Bessel functions, can be generated by solving various recursion relations.¹ This suggests multipole expansions are indeed the way to try to solve the equations. One wants this to be as close to the flat case as possible. Let us define a polynomial of order ℓ by $\text{poly}_\ell(x\mu, x^2) \equiv x^\ell P_\ell(\mu)$. In the curved case, just as in the flat case, we can write²

$$\begin{aligned} \Delta^{k\mathcal{M}}(\mathbf{x}, \tau, \hat{q}) &= \sum_{\ell} (2\ell + 1) (\Delta_\ell(k, \tau) / \kappa_{\ell,t}) (-k)^{-\ell} \text{poly}_\ell(\hat{q}^i {}^{(3)}\bar{\nabla}_i, \bar{a}^2 {}^{(3)}\bar{\nabla}^2) Q_{k\mathcal{M}}(\mathbf{x}), \\ \kappa_{\ell,t} &= \prod_{\ell'=0}^{\ell} \kappa_{\ell'}, \quad \kappa_\ell^2 \equiv 1 - (\ell^2 - 1) \frac{{}^{(3)}\bar{R}\bar{a}^2}{6k^2}, \quad \kappa_0 = \kappa_1 = 1. \end{aligned} \quad (361)$$

The Δ_ℓ correspond to $\Delta_{\{t,Q,U,V\},\ell}^{(S)}$, $\Delta_{\text{erv},\ell}$, etc. and ${}^{(3)}\bar{R}\bar{a}^2 = \pm 6d_{\text{curv}}^{-2}$. The product of κ_ℓ 's in the denominator helps to regulate the hierarchy of moment equations in the presence of curvature [303, 304, 305]. When we express the hierarchy equations for neutrinos and photons in terms of $\Delta_\ell(k, \tau) / \kappa_{\ell,t}$ they remain the same as for the flat case, e.g., eqs. (333), except an effective source term is added to the right-hand side:

$$\mathcal{G}_{\text{curv},\ell} = k \frac{\ell + 1}{2\ell + 1} \frac{\ell(\ell + 2) {}^{(3)}\bar{R}\bar{a}^2}{6k^2} \Delta_{t(\ell+1)}^{(S)} / \kappa_{\ell,t}. \quad (362)$$

The ${}^{(3)}\bar{R}$ corrections to the metric equations must be included as well of course. For numerical solution, one should rewrite the equations explicitly in terms of $\Delta_\ell(k, \tau)$. In that case, the $\mathcal{G}_{\text{curv},\ell}$ is absorbed into the left hand side, with κ_ℓ terms now appearing in the coupling of $\Delta_{(\ell-1)}$, $\Delta_{(\ell+1)}$ to $\dot{\Delta}_\ell$:

$$\dot{\Delta}_\ell - k \left(\frac{\ell}{2\ell + 1} \kappa_\ell \Delta_{(\ell-1)} - \frac{\ell + 1}{2\ell + 1} \kappa_{\ell+1} \Delta_{(\ell+1)} \right) = \text{usual RHS}. \quad (363)$$

¹For open FRW universes the spectrum of the Laplacian has $kd_{\text{curv}} \geq 1$, and the radial functions are

$$X_{k\ell}(x) = \sqrt{\frac{\pi}{2}} (kd_{\text{curv}})^\ell P_{i\beta - \frac{1}{2}}^{-(\ell + \frac{1}{2})}(\cosh(x)) \sinh^{-1/2}(x), \quad x \equiv \chi/d_{\text{curv}},$$

where the $P_\nu^\mu(x)$ are associated Legendre functions (e.g., [133]).

²This expansion format suggests that we define generalized potentials $U_{t\ell}$ by $\Delta_t = \sum_{\ell} (-)^\ell \text{poly}_\ell(\hat{q}^i {}^{(3)}\bar{\nabla}_i, \bar{a}^2 {}^{(3)}\bar{\nabla}^2) U_{t\ell} / \kappa_{\ell,t}$, so that $U_{t1} = \bar{a}^{-1} \Psi_{v,\gamma}$, $U_{t2} = 5\kappa_{2,t} \pi_{t,\gamma} / 12$ and the higher ℓ equations become more like the energy and momentum conservation laws.

(Sources in the $\ell = 2$ equation also have to be multiplied by κ_2 , and, to the extent they explicitly involve Δ_ℓ , rewritten with the correct $\kappa_{\ell,t}$ factors.) Because the angle-distance relation for open universes results in the typical ℓ associated with a given wavenumber being much larger than in a flat universe, being able to free-stream from an early time to the present is very useful to speed up numerical evaluations, but this is difficult to make efficient, unlike in the flat case [305]. The full numerical problem for open universes was first tackled by Mike Wilson [133], was picked up again by [138], and, more recently, by [245, 244, 305] for open CDM models and by [287] for texture and other isocurvature seed models; closed models are addressed in [304, 305].

In the absence of knowing what the generation mechanism is for the fluctuations, it is usual in cosmology to consider “natural” spectral shapes such as power laws. What complicates matters is that the phase space for curved universes goes like $\beta^2 d\beta$, where $\beta \equiv \sqrt{(kd_{curv})^2 - 1}$ for scalar perturbations and $\sqrt{(kd_{curv})^2 - 3}$ for tensor perturbations, with the spectrum of β going from 0 to ∞ . (In closed models, $\beta \equiv \sqrt{(kd_{curv})^2 + 1}$ for scalar perturbations, $\sqrt{(kd_{curv})^2 + 3}$ for tensor perturbations, with $\beta > 0$ but in this case the β spectrum is discrete.) It is unclear *a priori* whether the power law should be in kd_{curv} , β , volume or another combination.

In inflation models with mean curvature, if the generation mechanism is the usual zero point quantum fluctuations in scalar or gravity wave fields, the equations of sections 6.3 and 6.2.5 describe the development. In [245, 244], it was shown that $d_{curv}^{-1}(kd_{curv})^{-2}(3 + (kd_{curv})^2)((kd_{curv})^2 - 1)^{-1/2}$ is an inflation-inspired analogue of the k^1 Harrison–Zeldovich energy density spectrum for flat Universes. This looks complicated but has a very simple physical interpretation: just as for the flat case, this translates to equal power per decade of wavenumber in the gravitational potential. Thus, it is advantageous to use power per logarithmic waveband to express this. Actually the scale independence is in the gauge invariant variables ζ or φ_{com} (section 6.2.3), which are $\propto \Phi_N \equiv -\Phi_H$, the gravitational potential, on large scales. With tilt ν_s , $\mathcal{P}_\zeta(k) \sim (kd_{curv})^{\nu_s}$ is suggested by the absence of curvature effects explicitly appearing in the equation for scalar field perturbations, eq. (176). The analogue for tensor perturbations for which curvature corrections explicitly appear in the gravitational wave evolution equation, eq. (169), is $\mathcal{P}_{GW}(k) \sim ((kd_{curv})^2 - 2)^{\nu_t/2}$. In realistic inflation models there are further small corrections near $\beta = 0$ [305].

C.5 Lensing

Even though one usually linearizes in the metric variables to treat gravitational lensing in cosmological contexts, in transport theory it is a nonlinear process: $\mathcal{G}_{tbend}$ involves the transverse derivative to the instantaneous direction of the photon path, $-\partial\Delta_t/\partial\hat{q}^I F^I$, where F^I is a linear combination of the perturbed metric variables, $\nu, \varphi, \Psi_\sigma$. What complicates this is that under linear gauge transformations, Δ_t can get new components $\propto \hat{q}^J V_J$, where V_J involves metric

components; thus terms $F^I V_I$ of quadratic order in the metric components are induced. The situation can be clarified by recognizing that, in the absence of interactions with matter, the Boltzmann equation is just a book-keeping device saying that the mean photon occupation number (or phase space density) is conserved along photon trajectories and the photon trajectories can be solved with linearized potentials. As expected $d\hat{q}^I/d\tau = F^I$.

The expressions for the angular power spectrum derived in this section are meant to exercise some of the machinery and approximations given previously in these appendices. The relationship between $C^{lens}(\varpi)$ and $C_\ell^{no-lens}$ is equivalent to an expression given by Seljak [280] whose numerical results are described below; see also [275].

It is customary (e.g., [273]) to work in the longitudinal gauge for lensing, with metric variables $\nu_L = \Phi_N$ and $\varphi_L \rightarrow -\Phi_N$ once anisotropic stress can be neglected, so one's Newtonian insight into the potential Φ_N can be applied. In terms of these variables,

$$\mathcal{G}_{\text{tbend}} = \frac{\partial \Delta_t}{\partial \hat{q}^I} (\delta^{IJ} - \hat{q}^I \hat{q}^J) e_{*J} [\nu_L - \varphi_L]. \quad (364)$$

To relate this to the equations of motion, the expressions in the footnote in section B.1 are evaluated using the Ricci rotation coefficients eq. (274). For each geodesic there is an affine parameter λ “clocking” changes. We can also measure changes by transforming to conformal time $\tau(\lambda)$ or, as is done here, to comoving radial distance $\chi(\lambda)$ which is set to zero (as is λ) at the end of the photon trajectory; i.e., here, at $\mathbf{x}_0$, and now, at τ_0 . In terms of the photon momenta q^I that gives us the gauge-invariant $\tilde{\Delta}_t$ variable (i.e., with $\ln \Omega = \ln \bar{a} + \nu_L$, $\ln A = \ln \bar{a} + \varphi_L$), the geodesic equations are

$$\begin{aligned} \frac{1}{N} \frac{d \ln q}{d\tau} &= \bar{e}_n [\nu_L - \varphi_L] \\ \frac{1}{N} \frac{d \hat{q}^I}{d\tau} &= -(\delta^{IJ} - \hat{q}^I \hat{q}^J) \frac{1}{\bar{a}} e_{*J} [\nu_L - \varphi_L]. \\ \frac{dx^i}{d\tau} &= \frac{N}{A} \left(\frac{q}{q^n} e_{*I}^i \hat{q}^I + A e_n^i \right) \rightarrow e^{\nu_L - \varphi_L} \delta_I^i \hat{q}^I, \\ \frac{d\tau}{d\lambda} &= -\frac{q^n}{N\Omega} \rightarrow -e^{-2\nu_L} q, \quad d\chi/d\lambda \rightarrow e^{-(\nu_L + \varphi_L)} q. \end{aligned} \quad (365)$$

(Note that a surface of constant conformal time is not a surface at a fixed comoving distance in this gauge when one takes the perturbations into account.) The photon position as it meanders back and forth under the action of the metric obeys

$$\begin{aligned} x^I &\approx r^I - s^I, \quad r^I = x_0^I - \hat{q}_0^I \chi, \\ s^I &= -\int_0^{\chi_s} d\chi \int_0^\chi d\chi' (\delta^{IJ} - \hat{q}^I \hat{q}^J) e_{*J} [\nu_L - \varphi_L]. \end{aligned} \quad (366)$$

There are many similarities to the Zeldovich approximation, with the unperturbed photon trajectory r^I like the unperturbed (Lagrangian space) position,

with the deviation from that trajectory s^I like the displacement field, and with the true trajectory x^I like the “true” (Eulerian space) position. One can use the same methods for solving this problem as is used to map from Lagrangian space to Eulerian space in 3D cosmology. A flat Universe has been assumed. Thus we can use Fourier transform methods to find the solution. For example the correlation function at time τ_0 can be expressed in terms of the radiation pattern on the surface a distance χ_s away by

$$\begin{aligned} C^{lens}(\varpi) &= \langle \tilde{\Delta}_t(\hat{q}_0, \mathbf{x}_0, \tau_0) \tilde{\Delta}_t(\hat{q}'_0, \mathbf{x}_0, \tau_0) \rangle \\ &\approx \sum_{\mathbf{k}} e^{-i\mathbf{k} \cdot (\hat{q}_0 - \hat{q}'_0)\chi_s} \langle e^{-i\mathbf{k} \cdot (\mathbf{s} - \mathbf{s}')} \tilde{\Delta}_t(\mathbf{k}, \tau_s, \hat{k} \cdot \hat{q}_s) \tilde{\Delta}_t^*(\mathbf{k}, \tau_s, \hat{k} \cdot \hat{q}'_s) \rangle. \end{aligned} \quad (367)$$

Here as usual $\varpi = \hat{q}_0 - \hat{q}'_0$, $\hat{q}_s$ and $\hat{q}'_s$ are the directions of the photons at χ_s , and $\Delta \mathbf{s} \equiv \mathbf{s} - \mathbf{s}'$. The ensemble-average encompasses the statistics of both the radiation pattern at χ_s and the distribution of the clumped matter lying between χ_s and us which is responsible for the bending. In practice it will be an excellent approximation to assume they are statistically independent of each other. As a further simplification along the lines of the “small angle approximations” described in section C.3.2, we replace $\langle \tilde{\Delta}_t(\mathbf{k}, \tau_s, \hat{k} \cdot \hat{q}_s) \tilde{\Delta}_t^*(\mathbf{k}, \tau_s, \hat{k} \cdot \hat{q}'_s) \rangle$ by the DSZ approximation, $\langle |\tilde{\Delta}_t(\mathbf{k}, \tau_s, \bar{\mu})|^2 \rangle$. In the usual DSZ approximation, $\bar{\mu} = \hat{k} \cdot (\hat{q}_s + \hat{q}'_s)/2$. In principle the average lensed polar direction, $(\hat{q}_s + \hat{q}'_s)/2$, could be shifted considerably from the unperturbed direction $(\hat{q}_0 + \hat{q}'_0)/2$ on the sky. Still, as a first approximation we replace $\bar{\mu}$ by its ensemble average, $\bar{\mu} = \hat{k} \cdot (\hat{q}_0 + \hat{q}'_0)/2$, invoking [88] who showed that one still gets a good approximation by going one step beyond DSZ by isotropizing $\langle |\tilde{\Delta}_t(\mathbf{k}, \tau_s, \bar{\mu})|^2 \rangle$.

For small angles we can also use a Fourier transform approximation to the power spectrum, utilizing a split into components transverse and parallel to the average line of sight, which sets the unlensed 2D wavenumber to be $\mathbf{Q}_0 = \mathbf{k}_{\perp\chi_s}$:

$$C^{lens}(\varpi) \approx \int \frac{d^2 \mathbf{Q}_0}{(2\pi)^2} e^{i\mathbf{Q}_0 \cdot \varpi} \langle e^{-i\mathbf{Q}_0 \cdot (\mathbf{s} - \mathbf{s}')/\chi_s} \rangle \frac{2\pi}{Q_0^2} C_{\ell_0}^{no-lens}. \quad (368)$$

As usual, $Q_0 = |\mathbf{Q}_0| = \ell_0 + \frac{1}{2}$. The statistical average $\langle e^{-i\mathbf{Q}_0 \cdot \Delta \mathbf{s}/\chi_s} \rangle$ is the characteristic function for the random variable $\mathbf{Q}_0 \cdot \Delta \mathbf{s}$, expressible in terms of all of the connected N -point correlation function of it. A subject which is interesting to explore is the extent to which non-Gaussian features will manifest themselves. To date the papers have focussed on simplified approximations to get an idea of the magnitude of the effect. The leading term for this average is a Gaussian approximation, $\exp[-\frac{1}{2} \sum_{AB} Q_{0A} Q_{0B} \langle \Delta s_A \Delta s_B \rangle / \chi_s^2]$, where $A, B = 1, 2$ for the two components of the transverse vector. If the separation $|\varpi|$ is small, then Δs_A can be expanded in terms of the “shear tensor” $\varepsilon_{AB} = -\partial \Delta s_A / \partial (\chi_s \varpi^B)$. (Strain tensor rather than shear tensor is the more appropriate name.) For the basis of the illustration of this section, we shall just consider the isotropized version of $\langle (\mathbf{Q}_0 \cdot \Delta \mathbf{s})^2 \rangle$, i.e., $\frac{1}{2} Q_0^2 \langle \Delta \mathbf{s} \cdot \Delta \mathbf{s} \rangle$, which

I define to be $\frac{1}{2}Q_0^2\varepsilon^2\varpi^2\chi_s^2$. In the small angle limit of the isotropized version, $\varepsilon^2 = \frac{1}{2}\varepsilon_{AB}\varepsilon^{AB}$.

We can use Fourier methods to determine the rms displacement. In the $\varphi_L = -\Phi_N$ limit, [280, 273]

$$\frac{1}{2}\varepsilon^2\varpi^2 \approx \int \frac{d\chi}{\chi_s} \int \frac{d^2\mathbf{Q}_0}{(2\pi)^2} (1 - e^{i\mathbf{Q}_0 \cdot \varpi}) \frac{2\pi^2}{Q_0} 4\mathcal{P}_{\Phi_N} \left(\frac{Q_0}{\chi_s} \right) (1 - \chi/\chi_s)^2. \quad (369)$$

In the small ϖ limit, ε is ϖ -independent as expected. For this constant ε case, the Fourier transform of the correlation function can be done explicitly:

$$\mathcal{C}_\ell^{lens} = \int d^2\mathbf{Q}_0 \frac{\exp[-|\mathbf{Q} - \mathbf{Q}_0|^2/(Q_0^2\varepsilon^2)]}{\pi(Q_0^2\varepsilon^2)} \frac{Q^2}{Q_0^2} \mathcal{C}_{\ell_0}^{no-lens}. \quad (370)$$

$\mathbf{Q}$ is the lensed angular wavenumber and $Q = \ell + \frac{1}{2}$. The total power is conserved – the logarithmic integrals of $\mathcal{C}_\ell^{lens}$ and $\mathcal{C}_{\ell_0}^{no-lens}$ are the same – but it is rearranged via the convolution, which is a smoothing in ℓ -space. If $\varepsilon(\varpi)$ is changing slowly with angular scale ϖ , $\varepsilon(Q_0^{-1})$ is reasonable to use. Seljak [280] has used realistic gravitational potential power spectra – linear theory on large scales with a good approximation to nonlinear effects on small scales, thereby enhancing the lensing effect – to estimate ε . A rough fit, covering a range from arcsecond scales up to tens of arcminutes, is $\varepsilon(\theta) \sim 0.2 - 0.03 \ln(\theta/1')$ for a CDM model with $\sigma_8 = 1$ and $\sim 0.14 - 0.03 \ln(\theta/1')$ for a $\Omega_{vac} = 0.8$ model. Thus the spread around ℓ_0 , $\Delta\ell/\ell_0 \sim \varepsilon$, is not very large, ~ 0.2 at a few arcminutes, less for larger scales; note also that ε is changing slowly with θ , with a local power law index $\lesssim 0.2$ for arcminute scales, so the constant ε approximation is not even too bad. The net effect is that the higher Doppler peaks and troughs are smoothed out enough so that one must take the lensing effect into account in some happy future where we have an extremely well determined $\mathcal{C}_\ell$.

C.6 Tensor perturbation source terms

As we saw in eq. (311), the natural variables to use for tensor perturbations are $\tilde{\Delta}_{\{t,U,V,Q\}}^{(T\epsilon)}$ defined by the expansion

$$\Delta_{ij}^{(T)} = w \sum_{(\mu)=t,Q,U,V} \sum_{\epsilon=+,\times} \sum_{\mathbf{k}} \tilde{\Delta}_{(\mu)}^{(T\epsilon)} \frac{E^{(T\epsilon)} \cdot \mathcal{E}_{(\mu)}}{\mathcal{E}_{(\mu)} \cdot \mathcal{E}_{(\mu)}} E_{ij}^{(T\epsilon)} e^{i\mathbf{k} \cdot \mathbf{x}} a_{\mathbf{k}(T\epsilon)} + cc.$$

The polarization basis for $\mathbf{k}$ -modes is eq. (319), with $\hat{q}$ replacing $\hat{q}'$:

$$\varepsilon_2 = (-\sin\phi, -\cos\phi, 0), \quad \varepsilon_1 = (-\mu\cos\phi, -\mu\sin\phi, \sqrt{1-\mu^2}). \quad (371)$$

To determine $\tilde{\Delta}_{t,U,V,Q}^{(T\epsilon)}$, we need the 2×4 transformation matrix of inner products

$$\left(\frac{E^{(T\{+,\times\})} \cdot \mathcal{E}_{\{t,U,V,Q\}}}{\mathcal{E}_{\{t,U,V,Q\}} \cdot \mathcal{E}_{\{t,U,V,Q\}}} \right)$$

$$\begin{aligned}
&= \begin{bmatrix} -(1-\mu^2)\cos(2\phi) & (1+\mu^2)\cos(2\phi) & 2\mu\sin(2\phi) & 0 \\ -(1-\mu^2)\sin(2\phi) & (1+\mu^2)\sin(2\phi) & -2\mu\cos(2\phi) & 0 \end{bmatrix}; \\
\text{e.g., } \frac{E^{(T+)} \cdot \mathcal{E}_Q}{\mathcal{E}_Q \cdot \mathcal{E}_Q} &= (\varepsilon_1)_1(\varepsilon_1)_1 - (\varepsilon_2)_1(\varepsilon_2)_1 - (\varepsilon_1)_2(\varepsilon_1)_2 + (\varepsilon_2)_2(\varepsilon_2)_2 \\
&= (1+\mu^2)\cos(2\phi). \tag{372}
\end{aligned}$$

Note that there is no $\tilde{\Delta}_V^{(T\epsilon)}$. One can also expand the source functions $\mathcal{G}^{(T)}_{ij,C}$ and $\mathcal{G}^{(T)}_{ij,SW}$ in modes:

$$\begin{aligned}
&\mathcal{G}^{(T)}_{ij\{C,SW\}} \\
&= w \sum_{(\mu)} \sum_{\epsilon} \sum_{\mathbf{k}} \tilde{\mathcal{G}}^{(T\epsilon)}_{(\mu)\{C,SW\}} \frac{E^{(T\epsilon)} \cdot \mathcal{E}_{(\mu)}}{\mathcal{E}_{(\mu)} \cdot \mathcal{E}_{(\mu)}} E^{(T\epsilon)}_{ij} e^{i\mathbf{k} \cdot \mathbf{x}} a_{\mathbf{k}(T\epsilon)} + \text{cc.}
\end{aligned}$$

The evaluation of $\tilde{\mathcal{G}}^{(T\epsilon)}_{(\mu)SW}$ is simple, with the result eq. (207): $\tilde{\mathcal{G}}^{(T\epsilon)}_{tSW} = \frac{1}{2}\dot{h}_{(T\epsilon)}$, with the rest vanishing.

To get the Thomson scattering source functions eq. (208) for $\tilde{\mathcal{G}}^{(T\epsilon)}_{(\mu)C}$ is more work. A straightforward route is to isolate the $\cos(2(\phi - \phi'))$, $\sin(2(\phi - \phi'))$ terms in the phase tensor components $[P]^{(\mu)}_{(\nu)}$. Let us denote the perturbation variables in an expansion in $\cos(2\phi)$ and $\sin(2\phi)$ by $\Delta^{(T\epsilon)}_{(\mu)}$, $\mathcal{G}^{(T\epsilon)}_{(\mu)\{C,SW\}}$, without the *tilde*. The relation to the *tilde* variables is

$$\Delta_t^{(T\epsilon)} \equiv -(1-\mu^2)\tilde{\Delta}_t^{(T\epsilon)}, \quad \Delta_Q^{(T\epsilon)} = (1+\mu^2)\tilde{\Delta}_Q^{(T\epsilon)}, \quad \Delta_U^{(T\epsilon)} = -2\mu\tilde{\Delta}_U^{(T\epsilon)},$$

and similarly for $\mathcal{G}^{(T\epsilon)}_{(\mu)SW}$ and $\mathcal{G}^{(T\epsilon)}_{(\mu)C}$, which is given by

$$\begin{aligned}
\tau_C \mathcal{G}_{tC}^{(T\epsilon)} &= -\Delta_t^{(T\epsilon)} + (1-\mu^2)\Upsilon_{(T\epsilon)}, \quad \epsilon = +, \times, \\
\tau_C \mathcal{G}_{QC}^{(T\epsilon)} &= -\Delta_Q^{(T\epsilon)} - (1+\mu^2)\Upsilon_{(T\epsilon)}, \\
\tau_C \mathcal{G}_{UC}^{(T\epsilon)} &= -\Delta_U^{(T\epsilon)} + 2\mu\Upsilon_{(T\epsilon)}, \\
\tau_C \mathcal{G}_{VC}^{(T\epsilon)} &= -\Delta_V^{(T\epsilon)}, \\
\Upsilon_{(T\epsilon)} &\equiv \frac{3}{8} \int \frac{1}{2} d\mu' \left[\frac{1}{2}(1-(\mu')^2)\Delta_t^{(T\epsilon)} - \frac{1}{2}(1+(\mu')^2)\Delta_Q^{(T\epsilon)} + \frac{1}{2}2\mu'\Delta_U^{(T\epsilon)} \right].
\end{aligned} \tag{373}$$

Although the derivation of $[P]^{(\mu)}_{(\nu)}$ was done in the comoving baryon gauge, the tensor terms $\Delta^{(T\epsilon)}_{(\mu)}$ are all gauge invariant, so are valid in any gauge.

The classical route is to have the form of $[P]^{(\mu)}_{(\nu)}$ compel one to first transform to variables $\Delta^{(T\epsilon)}_{(\mu)}$ which take out the $\cos(2\phi)$, $\sin(2\phi)$, then be compelled by the form of eqs. (373) to introduce the Polnarev combinations $\tilde{\Delta}^{(T\epsilon)}_{(\mu)}$. Note that $\Delta_V^{(T\epsilon)}$ obeys pure damping equations with no source terms, hence remains

unexcited by gravitational waves, and so vanishes identically, a result which follows directly in the *tilde* representation.

The $\tilde{\Delta}_{t,U,V,Q}^{(T\epsilon)}$ obey the simplified transfer equations

$$\frac{\partial}{\partial \tau} \tilde{\Delta}_t^{(T\epsilon)} + \hat{q} \cdot \nabla \tilde{\Delta}_t^{(T\epsilon)} = \frac{1}{2} \dot{h}_{(T\epsilon)} - \tau_C^{-1} \tilde{\Delta}_t^{(T\epsilon)} + \tau_C^{-1} \Upsilon_{(T\epsilon)}, \quad (374)$$

$$\frac{\partial}{\partial \tau} \tilde{\Delta}_Q^{(T\epsilon)} + \hat{q} \cdot \nabla \tilde{\Delta}_Q^{(T\epsilon)} = -\tau_C^{-1} \tilde{\Delta}_Q^{(T\epsilon)} + \tau_C^{-1} \Upsilon_{(T\epsilon)}, \quad \tilde{\Delta}_Q^{(T\epsilon)} = \tilde{\Delta}_U^{(T\epsilon)},$$

$$\Upsilon_{(T\epsilon)} \equiv \frac{3}{8} \int \frac{1}{2} d\mu' [\frac{1}{2}(1 - (\mu')^2)^2 \tilde{\Delta}_t^{(T\epsilon)} + \frac{1}{2}(1 + 6(\mu')^2 + (\mu')^4) \tilde{\Delta}_Q^{(T\epsilon)}],$$

$$\Upsilon_{(T\epsilon)} = \frac{1}{10} \tilde{\Delta}_{t0}^{(T\epsilon)} + \frac{1}{7} \tilde{\Delta}_{t2}^{(T\epsilon)} + \frac{3}{70} \tilde{\Delta}_{t4}^{(T\epsilon)} + \frac{3}{5} \tilde{\Delta}_{Q0}^{(T\epsilon)} - \frac{6}{7} \tilde{\Delta}_{Q2}^{(T\epsilon)} + \frac{3}{70} \tilde{\Delta}_{Q4}^{(T\epsilon)}.$$

As for scalar perturbations, these two transfer equations are solved by expanding in Legendre polynomials, $P_\ell(\mu)$ [140]. The moment equations are identical in form to those for scalar perturbations, except that only the $\ell = 0$ equations have nonzero sources for both $\tilde{\Delta}_{t,Q\ell}^{(T\epsilon)}$: Higher moments have only the usual $\tau_C^{-1} \tilde{\Delta}_{t\ell}^{(T\epsilon)}$ damping and grow only as a result of the flux from lower ℓ 's through the $\hat{q} \cdot \hat{k} \tilde{\Delta}_{t,Q}^{(T\epsilon)}$ propagation term. The $\ell = 0$ source feeding the development of total anisotropy is $\frac{1}{2} \dot{h}_{(T\epsilon)} + \tau_C^{-1} \Upsilon_{(T\epsilon)}$. The polarization growth is fed by $\tau_C^{-1} \Upsilon_{(T\epsilon)}$ in the $\ell = 0$ equation.

Given $\dot{h}_{(T\epsilon)}$, there is an exact solution for $\tilde{\Delta}_t^{(T\epsilon)} - \tilde{\Delta}_Q^{(T\epsilon)}$, which is a free-streaming solution including damping (associated with the Thomson depth ζ_C). The polarization is quite small [141], so this is also a good approximation for $\tilde{\Delta}_t^{(T\epsilon)}$, the solution when the $\Upsilon_{(T\epsilon)}$ feed is neglected:

$$\tilde{\Delta}_t^{(T\epsilon)} \approx \tilde{\Delta}_t^{(T\epsilon)} - \tilde{\Delta}_Q^{(T\epsilon)} = \int_0^{\tau_0} e^{-\zeta_C(\tau)} d\tau e^{-i\mathbf{k} \cdot \hat{q}\chi} \frac{1}{2} \dot{h}_{(T\epsilon)}(\tau), \quad (375)$$

$$\tilde{\Delta}_{t\ell}^{(T\epsilon)} \approx \tilde{\Delta}_{t\ell}^{(T\epsilon)} - \tilde{\Delta}_{Q\ell}^{(T\epsilon)} = \int_0^{\tau_0} e^{-\zeta_C(\tau)} d\tau j_\ell(k\chi) \frac{1}{2} \dot{h}_{(T\epsilon)}(\tau). \quad (376)$$

Although working with the $+$, $\times$ quantities has some advantages, for derivations it is useful to use the expansion

$$\Delta_t^{(T)} = -\sqrt{\frac{16\pi}{15}} \tilde{\Delta}_t^{(TG)} Y_{22} + \text{cc}, \quad \tilde{\Delta}_t^{(TG)} \equiv \frac{1}{\sqrt{2}} (\tilde{\Delta}_t^{(T+)} - i\tilde{\Delta}_t^{(T\times)}). \quad (377)$$

Here cc denotes complex conjugate. This explicitly shows that an $\ell = 2$ tensor component is the leading term coming out of gravity waves, whereas for scalar modes there are (gauge dependent) $\ell = 0$ and $\ell = 1$ terms. We also introduce the notation $h_G \equiv (h_{(T+)} - ih_{(T\times)})/\sqrt{2}$ for the analogous gravity wave contribution. To go from $\tilde{\Delta}_t^{(TG)}$ to multipole components on the sky and the angular

power spectrum, we make use of

$$\begin{aligned}
Y_{\ell\pm 2}(\mu, 0) &= \sum_m \mathcal{D}_{\pm 2m}^{(\ell)}(\hat{k}) Y_{\ell m}(\hat{q}), \quad \mu = \hat{k} \cdot \hat{q}, \\
(2\ell + 1)Y_{2\pm 2}(\mu, 0)P_\ell(\mu) \\
&= \sum_{\ell'} \sqrt{5}\sqrt{(2\ell' + 1)} \langle 20\ell'0|\ell 0 \rangle \langle 2 \pm 2\ell' - (\pm 2)|\ell 0 \rangle Y_{\ell'\pm 2}(\mu, 0) \\
&= \sqrt{5} \sum_{\ell'} \frac{1}{2} \sqrt{\frac{3}{2}} \sqrt{(2\ell' + 1)} \sqrt{(\ell' - 1)\ell'(\ell' + 1)(\ell' + 2)} \\
&\quad \times \left[\frac{\delta_{\ell, \ell'+2}}{(2\ell' + 1)(2\ell' + 3)} - 2 \frac{\delta_{\ell, \ell'}}{(2\ell' - 1)(2\ell' + 3)} + \frac{\delta_{\ell, \ell'-2}}{(2\ell' - 1)(2\ell' + 1)} \right].
\end{aligned}$$

$\mathcal{D}_{\pm 2m}^{(\ell)}(\alpha, \beta, \gamma)$ denotes the irreducible rotation tensor of rank ℓ for a rotation with Euler angles α, β, γ , with here $\alpha = 0$ and β, γ the polar angles of $\hat{k}$. The $\langle \ell m \ell' m' | L M \rangle$ are Clebsch–Gordon coefficients [263]. Thus the multipole coefficients are

$$\begin{aligned}
a_{t, \ell m}^{(T)} &= \sum_{\mathbf{k}} \mathcal{D}_{\pm 2m}^{(\ell)}(\hat{k}) \sqrt{(2\ell + 1)} \sqrt{(\ell - 1)\ell(\ell + 1)(\ell + 2)} \\
&\quad \times \left[\frac{\tilde{\Delta}_{t, \ell-2}^{(TG)}}{(2\ell - 1)(2\ell + 1)} + 2 \frac{\tilde{\Delta}_{t, \ell}^{(TG)}}{(2\ell - 1)(2\ell + 3)} + \frac{\tilde{\Delta}_{t, \ell+2}^{(TG)}}{(2\ell + 1)(2\ell + 3)} \right] + \text{cc}
\end{aligned}$$

and the differential angular power spectrum is

$$\begin{aligned}
\frac{d\mathcal{C}_{t\ell}^{(T)}}{d \ln k} &= \ell(\ell + 1) \left(1 - \frac{1}{\ell^2}\right) \left(1 + \frac{2}{\ell}\right) \frac{k^3}{2\pi^2} \left\langle \left| \frac{\tilde{\Delta}_{t, \ell-2}^{(TG)}}{\left(1 - \frac{1}{2\ell}\right)\left(1 + \frac{1}{2\ell}\right)} \right. \right. \\
&\quad \left. \left. + 2 \frac{\tilde{\Delta}_{t, \ell}^{(TG)}}{\left(1 - \frac{1}{2\ell}\right)\left(1 + \frac{3}{2\ell}\right)} + \frac{\tilde{\Delta}_{t, \ell+2}^{(TG)}}{\left(1 + \frac{1}{2\ell}\right)\left(1 + \frac{3}{2\ell}\right)} \right|^2 \right\rangle. \quad (378)
\end{aligned}$$

When we use the brick wall approximation for $e^{-\zeta_C(\tau)}$, unity after recombination, zero before, in eq. (376) we obtain the Abbott and Wise [225] approximation for tensor mode microwave background fluctuations. Keeping the full $e^{-\zeta_C(\tau)}$ improves the approximation. Obtaining the power spectra for the polarization is more complex because the multiplying functions going from $\Delta_{Q,U}^{(T\epsilon)}$ to the $\tilde{\Delta}_Q^{(T\epsilon)}$ variables are not simply $Y_{2\pm 2}$.

References

- [1] Ya.B. Zeldovich and R.A. Sunyaev, 1980, *Ann. Rev. Astron. Ap.* **18**, 537; R.A. Sunyaev and Ya.B. Zeldovich, 1981, *Space Physics Reviews* **1**, 1.

- [2] J.R. Bond, 1988, in: W.G. Unruh, ed., *The Early Universe*, Proc. NATO Summer School, Vancouver Is., Aug. 1986 (Reidel, Dordrecht).
- [3] J.R. Bond, 1989, The formation of cosmic structure, in: A. Astbury et al., eds., *Frontiers in Physics – From Colliders to Cosmology* (World Scientific, Singapore) p. 182.
- [4] G. Efstathiou, 1990, in: A. Heavens, A. Davies and J. Peacock, eds., *Physics of the Early Universe*, Scottish Universities Summer School Publications.
- [5] J.R. Bond, 1994, Cosmic Structure Formation and the Background Radiation, in: T. Padmanabhan, ed., *Proceedings of the IUCAA Dedication Ceremonies*, Pune, India, Dec. 28–30, 1992 (Wiley).
- [6] J.R. Bond, 1994, Testing inflation with the cosmic background radiation, in: M. Sasaki, ed., *Relativistic Cosmology*, Proc. 8th Nishinomiya-Yukawa Memorial Symposium (Academic Press) astro-ph/9406075.
- [7] M. White, D. Scott and J. Silk, 1994, *Ann. Rev. Astron. Ap.* **32**, 319.
- [8] P.J.E. Peebles, 1993, *Principles of Physical Cosmology* (Princeton University Press, Princeton, NJ), especially chapters 6 and 24.
- [9] R.B. Partridge, 1994, *3 K: The Cosmic Microwave Background Radiation* (Cambridge University Press, Cambridge).
- [10] The Evolution of the Universe, Dahlem Workshop Report ES 19, ed. G. Borner and S. Gottlober, (John Wiley and Sons); *e.g.*, J.R. Bond 1997, pp. 199–223, *Implications of the Background Radiation for Cosmic Structure Formation*
- [11] J.C. Mather et al., 1994, *Ap. J. Lett.* **420**, 439.
- [12] D.J. Fixsen et al., 1996, *Ap. J.* **473**, 576.
- [13] A.A. Penzias and R.W. Wilson, 1965, *Ap. J.* **142**, 419.
- [14] R. Weiss, 1980, *Ann. Rev. Astron. Ap.* **18**, 489.
- [15] H.P. Gush, M. Halpern and E. Wishnow, 1990, *Phys.Rev. Lett.* **65**, 537.
- [16] T.F. Howell and J.R. Shakeshaft, 1966, *Nature* **210**, 1318; 1967, **216**, 753.
- [17] A. Kogut, 1991, in: S.S. Holt, C.L. Bennett and V. Trimble, eds., *After the First Three Minutes*, AIP Conference Proceedings, Vol. 222 (AIP, New York).
- [18] D.G. Johnson and D. Wilkinson, 1987, *Ap. J. Lett.* **313**, L1.

- [19] M. Bensadoun et al., 1993, *Ap. J.* **409**, 1.
- [20] G. Sironi et al., 1990 *Ap. J.* **357**, 301; 1994, 1995, in *Astrophys. Lett. and Comm.*, **32**.
- [21] S. Staggs and D. Wilkinson, 1995, in *Astrophys. Lett. and Comm.*, **32**.
- [22] M. Bersanelli et al., 1995, in *Astrophys. Lett. and Comm.*, **32**.
- [23] G. Herzberg, 1945, *Molecular Spectra and Molecular Structure* (Prentice Hall, New York).
- [24] P. Thaddeus, 1972, *Ann. Rev. Astron. Ap.* **10**, 305.
- [25] D.M. Meyer and M. Jura, 1985, *Ap. J.* **297**, 119.
- [26] K. Roth, D.M. Meyer and I. Hawkins, 1993, *Ap. J. Lett.* **413**, L67.
- [27] P. Crane, 1989, *Ap. J.* **346**, 136; 1995, in *Astrophys. Lett. and Comm.* **32**; E. Palazzi, 1990, *Ap. J.* **357**, 14.
- [28] D.P. Woody and P.L. Richards, 1981, *Ap. J.* **248**, 18.
- [29] H.P. Gush, 1981, *Phys. Rev. Lett.* **47**, 745.
- [30] J.B. Peterson, P.L. Richards and T. Timusk, 1985, *Phys. Rev. Lett.* **55**, 332.
- [31] T. Matsumoto, S. Hayakawa, H. Matsuo, H. Murakami, S. Sato, A.E. Lange and P.L. Richards, 1988, *Ap. J.*, **329**, 567.
- [32] R. Sachs and A. Wolfe, 1967, *Ap. J.* **147**, 73.
- [33] A. Kompaneets, 1957, *Sov. Phys. JETP* **4**, 730.
- [34] Ya.B. Zeldovich and R.A. Sunyaev, 1969, *Ap. Space Sci.* **4**, 301.
- [35] R.J. Gould, 1984, *Ap. J.* **285**, 275.
- [36] C. Burigana, L. Danese and G. DeZotti, 1991, *Astron. Ap.* **246**, 49.
- [37] Yu.E. Lyubarskii and R.A. Sunyaev, 1983, *Astron. Ap.* **123**, 171.
- [38] R.A. Windhorst, E.B. Fomalont, R.B. Partridge and J.D. Lowenthal 1993, *Ap. J.* **405**, 498; R.A. Windhorst et al. 1995, *Nature* **375**, 471.
- [39] C.G.T. Haslam, H. Stoffel, C.J. Salter, and W.E. Wilson 1982, *Astron. Ap. Supp.* **47**, 1.
- [40] P. Reich and W. Reich 1988, *Astron. Ap. Supp.* **74**, 7.
- [41] M. Jones 1996, preprint

- [42] J.R. Bond, B.J. Carr and C.J. Hogan, 1991, *Ap. J.* **367**, 420 [BCH2].
- [43] E.L. Wright et al., 1991, *Ap. J.* **381**, 200.
- [44] F.-X. Desert, F. Boulanger and J.L. Puget, 1990, *Astron. Ap.* **237**, 215.
- [45] J.L. Puget and A. Leger, 1989, *Ann. Rev. Astron. Ap.* **27**, 161.
- [46] J.S. Mathis, W. Rumpl and K.H. Nordsieck, 1977, *Ap. J.* **217**, 425.
- [47] W.J. Barnes, 1994, MIT Thesis, A Model of Galactic Dust and Gas from FIRAS.
- [48] W. Reach et al., 1995, *Ap. J.* **451**, 188.
- [49] J.-L. Puget et al., 1996, *Astron. Ap.* **308**, L5.
- [50] J. Dorscher and T. Henning, 1995, *Astron. Ap. Reviews* **6**, 271.
- [51] S.-H. Kim, P.G. Martin and P.D. Hendry, 1994, *Ap. J.* **422**, 164.
- [52] E.L. Wright, 1987, *Ap. J.* **320**, 818.
- [53] D. Layzer and R. Hively, 1973, *Ap. J.* **179**, 361
- [54] F. Hoyle, 1980, *Steady State Cosmology Revisited*, (University College Cardiff Press, Cardiff, Wales).
- [55] F. Hoyle, G.R. Burbidge and J.V. Narlikar, 1993, *Ap. J.* **410**, 437; 1994, *M.N.R.A.S.* **267**, 1007.
- [56] N.C. Rana, 1979, *Ap. Space Sci.* **66**, 173; 1980 *Ap. Space Sci.* **71**, 123; 1981, *M.N.R.A.S.* **197**, 1125.
- [57] E.L. Wright, 1982, *Ap. J.* **255**, 401; I. Hawkins and E.L. Wright, 1988, *Ap. J.* **324**, 46.
- [58] E.L. Wright et al., 1994, *Ap. J.* **420**, 450.
- [59] P. Guhathakurta and B.T. Draine, 1989, *Ap. J.* **345**, 230.
- [60] P.J.E. Peebles, 1968, *Ap. J.* **153**, 1.
- [61] Ya.B. Zeldovich, V.G. Kurt and R.A. Sunyaev, 1969, *Sov. Phys. JETP* **28**, 146.
- [62] D.R. Bates and A. Dalgarno, 1962, in: D.R. Bates, ed., *Atomic and Molecular Processes* (Academic Press) p. 245.
- [63] W.J. Boardman, 1964, *Ap. J. Suppl.* **9**, 185.
- [64] D.E. Osterbrock, 1974, *Astrophysics of Gaseous Nebulae* (Freeman, San Francisco).

- [65] J.H. Krolik, 1990, Ap. J. **353**, 21.
- [66] B.J. Carr, J.R. Bond and W.D. Arnett, 1984, Ap. J. **277**, 445.
- [67] W.H. Press and P. Schechter, 1974, Ap. J. **187**, 425.
- [68] J.R. Bond and S. Myers, 1996, The peak-patch picture of cosmic catalogues I: algorithms; II: validation, Ap. J. Supp. **103**, 1; IV: analytic methods, preprint.
- [69] M. Tegmark, J. Silk and A. Blanchard, 1994, Ap. J. **420**, 484.
- [70] M. Fukugita and M. Kawasaki, 1994, M.N.R.A.S. **269**, 563.
- [71] P.J.E., Peebles, 1987, Ap. J. **277**, L1.
- [72] J. Bartlett and A. Stebbins, 1991, Ap. J. **371**, 8.
- [73] T.P. Walker, G. Steigman, D.N. Schramm, K.A. Olive and H.S. Kang, 1991, Ap. J. **376**, 51; M.S. Smith, L.H. Kawano and R.A. Malaney, 1993, Ap. J. Supp. **85**, 219.
- [74] S. Ikeuchi, K. Tomisaki and J.P. Ostriker, 1983, Ap. J. **265**, 583.
- [75] J.J. Levin, K. Freese and D.N. Spergel, 1992, Ap. J. **389**, 464.
- [76] J.P. Ostriker, C. Thompson and E. Witten, 1986, Phys. Lett. **B180**, 231.
- [77] J.P. Ostriker and C. Thompson, 1987, Ap. J. Lett. **323**, L97.
- [78] C. Thompson, 1993, private communication.
- [79] M.G. Hauser et al., 1991, in: S.S. Holt, C.L. Bennett and V. Trimble, eds., *After the First Three Minutes*, AIP Conference Proceedings, Vol. 222, (AIP, New York) p. 161.
- [80] M.G. Hauser, 1995, in *Unveiling the Cosmic Infrared Background*, ed. E. Dwek, AIP Conference Proceedings, (AIP, New York); 1995, IAU Symposium 168, *Examining the Big Bang and Diffuse Background Radiations*, ed. M. Kafatos and Y. Kondo (Kluwer, Dordrecht).
- [81] J.R. Bond, B.J. Carr and C.J. Hogan, 1986, Ap. J. **306**, 428.
- [82] J.R. Bond and S. Myers, 1993, in: M. Shull and H. Thronson, eds., *The Evolution of Galaxies and their Environment*, Proceedings of the Third Teton Summer School, NASA Conference Publication 3190, p. 21.
- [83] A. Franceschini, L. Toffolatti, P. Mazzei, L. Danese and G. De Zotti, 1991, Astron. Ap. Suppl. **89**, 285; G. De Zotti et al., 1995, in: Dust, Molecules and Backgrounds: from Laboratory to Space, Planetary and Space Science, in press.

- [84] C. Bennett et al., 1994, Ap. J., **436**, 423.
- [85] C. Bennett et al., 1996, Ap. J. Lett., **464**, 1; and 4-year DMR references therein.
- [86] A. Kogut et al., 1996, Ap. J. Lett. **464**, 5.
- [87] G. Hinshaw et al., 1996, Ap. J. Lett. **464**, 17.
- [88] J.R. Bond and G. Efstathiou, 1987, M.N.R.A.S. **226**, 655.
- [89] J.R. Bond, 1995, Astrophys. Lett. and Comm., **32**, 63.
- [90] G.F. Smoot et al., 1992, Ap. J. Lett. **396**, L1.
- [91] R. Kneissl and G. Smoot, 1993, COBE note 5053.
- [92] C. Lineweaver and G. Smoot, 1993, COBE note 5051.
- [93] K. Ganga, E. Cheng, S. Meyer and L. Page, 1993, Ap. J. Lett. **410**, L57 [firs].
- [94] A.C.S. Readhead et al., 1989, Ap. J. Lett. **346**, 556 [ov7].
- [95] P. Meinhold and P. Lubin, 1991, Ap. J. Lett. **370**, 11 [sp89].
- [96] T. Gaier et al., 1992, Ap. J. Lett. **398**, L1; J. Schuster et al., 1993, Ap. J. Lett. **412**, L47 [sp91].
- [97] E.J. Wollack et al., 1993, Ap. J. Lett. **419**, L49 [sk93].
- [98] P. Meinhold et al., 1993, Ap. J. Lett. **409**, L1; J. Gunderson et al., 1993, Ap. J. Lett. **413**, L1 [max3].
- [99] A. Clapp et al., 1994, M. Devlin et al., 1994, Ap. J. **430**, 1 [MAX4, max4].
- [100] S. Tanaka et al., 1996, Ap. J. Lett. **468**, 81 [max5].
- [101] E.S. Cheng et al., 1993, Ap. J. Lett. **420**, L37 [msam2 (g2),msam3 (g3)].
- [102] G.S. Tucker et al., 1993, Ap. J. Lett. **419**, L45 [wd2, wd1].
- [103] S.M. Gutierrez de la Cruz et al., 1995, Ap. J. **442**, 10; S. Hancock et al., 1994, Nature **367**, 333; R. Watson et al., 1992, Nature **357** 660 [ten].
- [104] J.E. Ruhl et al., 1995, Ap. J. Lett. **453**, L1 [py].
- [105] P. de Bernardis et al., 1994, Ap. J. Lett. **422**, L33 [ar].
- [106] M. White and M. Srednicki, 1995, Ap. J. **443**, 6.
- [107] T.N. Gautier et al., 1992, AJ **103**, 1313.

- [108] A. Kogut et al., 1996, Ap. J. **460**, 1.
- [109] E.T. Vishniac, 1987, Ap. J. **322**, 597.
- [110] L. Kofman and A.A. Starobinsky, 1985, Sov. Astron. Lett. **11**, 271.
- [111] B. Chaboyer et al., 1996, Science, in press.
- [112] S. Coles and N. Kaiser, 1987, M.N.R.A.S., 233, 637.
- [113] R. Schaeffer and J. Silk, 1988, Ap. J., 332, 1.
- [114] W.K. Gear and C.R. Cunningham, 1995, ASP conference series **75**, 215, (Multi-feed Systems for Radio Telescopes, ed., D.T. Emerson and J.M. Payne).
- [115] S. Church et al., 1993, M.N.R.A.S., in press.
- [116] E. Kreysa and A. Chini, 1989, Proc. Particle Astrophysics Workshop, Berkeley (World Scientific, Singapore).
- [117] J.R. Bond and S. Myers, 1991, in: D. Cline and R. Peccei, eds., Trends in Astroparticle Physics (World Scientific, Singapore) 262.
- [118] M. Markevitch, G.R. Blumenthal, W. Forman, C. Jones and R.A. Sunyaev, 1992, Ap. J. **395**, 326.
- [119] J.G. Bartlett, A.K. Gooding and D.N. Spergel, 1993, Ap. J. **403**, 1; J.G. Bartlett and J. Silk, 1994, Ap. J. **423**, 12.
- [120] J.R. Bond and S. Myers, 1996, The peak-patch picture of cosmic catalogues III: application to cluster X-ray emission and the SZ effect, Ap. J. Supp. **103**, 1.
- [121] S. Colanfrancesco, P. Mazzotta, Y. Rephaeli and N. Vittorio, 1994 Ap. J. **433**, 454.
- [122] R. Scaramella, R. Cen and J.P. Ostriker et al., 1994, Ap. J. **416**, 399.
- [123] R.B. Partridge and D.T. Wilkinson, 1967, Phys. Rev. Lett. **18**, 557.
- [124] E.K. Conklin and R.N. Bracewell, 1967, Nature **216**, 777.
- [125] P.E. Boynton and R.B. Partridge, 1973, Ap. J. **181**, 243; R.B. Partridge, 1980, Ap. J. **235**, 681.
- [126] J.M. Uson and D.T. Wilkinson, 1984, Ap. J. Lett. **277**, L1.
- [127] F. Melchiorri, B.O. Melchiorri, C. Ceccarelli and L. Pietranera, 1981, Ap. J. Lett. **250**, L1.

- [128] I.A. Strukov, D.P. Skulachev and A.A. Klypin, 1987, in: J. Audouze and A.S. Szalay, eds., *Proceedings I.A.U. Symposium 130* (Reidel, Dordrecht).
- [129] R.D. Davies et al., 1987, *Nature* **326**, 462.
- [130] P.J.E. Peebles and J.T. Yu, 1970, *Ap. J.* **162**, 815.
- [131] A.G. Doroshkevich, Ya.B. Zeldovich and R.A. Sunyaev, 1978, *Sov. Astron.* **22**, 523.
- [132] M.L. Wilson and J. Silk, 1981, *Ap. J.* **243**, 14.
- [133] M.L. Wilson, 1983, *Ap. J.* **273**, 2.
- [134] J.R. Bond and G. Efstathiou, 1984, *Ap. J. Lett.* **285**, L45.
- [135] N. Vittorio and J. Silk, 1984, *Ap. J. Lett.* **285**, L39.
- [136] N. Vittorio and J. Silk, 1992, *Ap. J. Lett.* **385**, 9.
- [137] M. Fukugita, N. Sugiyama and M. Umemura, 1990, *Ap. J.* **358**, 28.
- [138] N. Gouda, N. Sugiyama and M. Sasaki, 1991, *Prog. Theor. Phys.* **85**, 1023; 1991, *Ap. J. Lett.* **327**, 49; N. Gouda, M. Sasaki and Y. Suto, 1989, *Ap. J.* **341**, 557.
- [139] K.M. Gorski, R. Stompor and R. Juskiewicz, 1993, *Ap. J. Lett.* **410**, 1.
- [140] R. Crittenden, J.R. Bond., R.L. Davis., G. Efstathiou and P.J. Steinhardt, 1993, *Phys. Rev. Lett.* **71**, 324.
- [141] R. Crittenden, R. Davis and P. Steinhardt, 1993, *Ap. J. Lett.*, L13.
- [142] S. Dodelson and J. Jubas, 1994, *Phys. Rev. Lett.* **70**, 2224.
- [143] U. Seljak, 1994, *Ap. J. Lett.*, in press.
- [144] J.R. Bond, J.R. Crittenden, R.L. Davis, G. Efstathiou and P.J. Steinhardt, 1994, *Phys. Rev. Lett.* **72**, 13.
- [145] J.R. Bond, R.L. Davis and P.J. Steinhardt, 1995, *Astrophys. Lett. and Comm.*, **32**, 53.
- [146] J.R. Bond, 1994, *Phys. Rev. Lett.* **74**, 4369.
- [147] K. Ganga, L. Page, E. Cheng and S. Meyer, 1994, *Ap. J. Lett.* **432**, L15.
- [148] J.O. Gundersen et al., 1995, *Ap. J. Lett.*, **443**, L57 sp94.
- [149] C.B. Netterfield, N. Jurosik, L. Page and D. Wilkinson, 1995, *Ap. J. Lett.* **455**, L69. sk94.

- [150] C.B. Netterfield, M.J. Devlin, N. Jurosik, L. Page and E.J. Wollack, 1997, Ap. J. **474**, 47 sk95.
- [151] P.F. Scott et al., 1996, Ap. J. Lett. **461**, 1 cat96.
- [152] C. Bennett et al., 1996, MAP experiment home page, <http://map.gsfc.nasa.gov>
- [153] M.A. Janssen et al., 1997, Ap. J., in press, astro-ph/9602009.
- [154] M. Bersanelli et al., 1996, COBRAS/SAMBA, The Phase A Study for an ESA M3 Mission, preprint.
- [155] J.O. Berger 1985, Statistical Decision Theory and Bayesian Analysis (Springer-Verlag, New York).
- [156] C.W. Therrien 1992, Discrete Random Signals in Statistical Signal Processing, ISPN0-13-852112-3 (Prentice Hall).
- [157] E.F. Bunn et al., 1994, Ap. J. Lett. **432**, 75.
- [158] M. White 1996, Phys. Rev. D **53**, 3011.
- [159] L. Knox 1995, Phys. Rev. D **52**, 4307.
- [160] M. Tegmark, and G. Efstathiou 1996, M.N.R.A.S. in press.
- [161] G. Jungman, M. Kamionkowski, A. Kosowsky, and D.N. Spergel, 1996, Phys. Rev. Lett. **76**, 1007.
- [162] Bond, J.R. and Jaffe, A., 1996, CITA preprint, astro-ph/9610091.
- [163] L. Knox 1996, CITA preprint, astro-ph/9606066.
- [164] J.R. Bond, G. Efstathiou and M. Tegmark, 1997, astro-ph/9702100.
- [165] M. Zaldarriaga, D. Spergel, and U. Seljak, 1997, astro-ph/9702157.
- [166] D. Coulson, R.G. Crittenden and N.G. Turok, 1994, Phys. Rev. Lett. **73**, 2390.
- [167] R. Arnowitt, S. Deser and C.W. Misner, 1962, in: L. Witten, ed., Gravitation (Wiley, New York) pp. 227–265.
- [168] J.W. York, 1979, in: L. Smarr, ed., Sources of Gravitational Radiation (Cambridge University Press, Cambridge) p. 83.
- [169] J.W. York, 1983, in: N. Deruelle and T. Piran, eds., *Gravitational Radiation*, 1982 Les Houches Proceedings (North-Holland, Amsterdam) p. 175.
- [170] E.M. Lifshitz, 1946, JETP Letters **16**, 587.

- [171] J.M. Bardeen, 1980, Phys. Rev. D **22**, 1882.
- [172] T. Piran, 1988, in: W.G. Unruh, ed., *The Early Universe*, Proc. NATO Summer School, Vancouver Is., Aug. 1986 (Reidel, Dordrecht).
- [173] H. Kodama and M. Sasaki, 1984, Prog. Theor. Phys. Suppl. **78**, 1.
- [174] H. Kodama and M. Sasaki, 1986, Intern. J. Mod. Phys. **A1**, 265.
- [175] J. Bardeen, P.J. Steinhardt and M.S. Turner, 1983, Phys. Rev. D **28**, 679.
- [176] G. Chibisov and V.F. Mukhanov, 1982, M.N.R.A.S. **200**, 535; V.F. Mukhanov, H.A. Feldman and R.H. Brandenberger, 1992, Phys. Rep. **215**, 203; D.H. Lyth, 1985, Phys. Rev. D **31**, 1792.
- [177] V. Mukhanov, 1988, JETP **94**, 1.
- [178] J.M. Bardeen, 1988, in: A. Zee, ed., Proc. CCAST Symposium on Particle Physics and Cosmology (Gordon and Breach, New York).
- [179] L. Landau and Lifshitz, 1971, Classical Theory of Fields (Addison-Wesley, Reading, MA).
- [180] A.A. Starobinski, 1986, in: H.T. de Vega and N. Sanchez, eds., Current Topics in Field Theory, Quantum Gravity, and Strings, Proc. Meudon and Paris VI, Vol. 246 (Springer Verlag) p. 107.
- [181] D.S. Salopek and J.R. Bond, 1990, Phys. Rev. D **42**, 3936; 1991, Phys. Rev. D **43**, 1005.
- [182] D.S. Salopek and J. Stewart, 1994, preprint.
- [183] E.D. Stewart and D.H. Lyth, 1993, Phys.Lett. **302B**, 171.
- [184] A.R. Liddle and M.S. Turner, 1994, Fermilab preprint, astro-ph/9402021 FERMILAB-Pub-93/399-A; E.W. Kolb and S.L. Vadas, 1994, Fermilab preprint, astro-ph/9403001.
- [185] D.H. Lyth and E.D. Stewart, 1992, Phys. Lett. **274B**, 168.
- [186] D. La and P.J. Steinhardt, 1989, Phys. Rev. Lett. **62**, 376.
- [187] E.W. Kolb, D.S. Salopek and M.S. Turner, 1990, Phys. Rev. D **42**, 3925.
- [188] A.D. Linde, 1983, Phys. Lett. B **129**, 177.
- [189] K. Freese, J.A. Frieman, and A.V. Olinto, 1990, Phys. Rev. Lett. **65**, 3233.
- [190] F.C. Adams, J.R. Bond, K. Freese, J.A. Frieman and A.V. Olinto, 1993, Phys. Rev. D, 47, 426.

- [191] L.A. Kofman and A.D. Linde, 1987, Nucl. Phys. B **282**, 555; L.A. Kofman and D.Yu. Pogosyan, 1988, Phys. Lett. **214** B, 508.
- [192] D.S. Salopek, J.R. Bond and J.M. Bardeen, 1989, Phys. Rev. D **40**, 1753.
- [193] J. Ehlers, 1971, in: R. Sachs, ed., General Relativity and Cosmology (Academic Press, New York).
- [194] J.M. Stewart, 1971, *Non-Equilibrium Relativistic Kinetic Theory*, Lecture notes in Physics **10** (Springer Verlag, Berlin).
- [195] J.R. Bond and A.S. Szalay, 1983, Ap. J. **277**, 443.
- [196] C.W. Misner, K.S. Thorne and J.A. Wheeler, 1973, Gravitation (Freeman, San Francisco).
- [197] G. Baym and L. Kadanoff, 1962, Quantum Statistical Mechanics (Benjamin).
- [198] R.K. Osborn and S. Yip, 1966, The Foundations of Neutron Transport Theory (Gordon and Breach, New York).
- [199] U. Fano, 1954, Phys. Rev. **93**, 121.
- [200] S. Chandrasekhar, 1960, Radiative Transfer (Dover, New York) pp. 1–53.
- [201] J.R. Bond, 1987, in: C. Dyer and B. Tupper, eds., *General Relativity and Astrophysics*, Proc. Second Canadian Conference on General Relativity (World Scientific, Singapore) pp. 310–314.
- [202] M. Jones, 1994, in: F. Durret et al., eds., *Clusters of Galaxies*, Proc. XXIXth Rencontre de Moriond (Edition Frontières).
- [203] Y. Raphaeli 1995, Ann. Rev. Astron. Ap. **33**, 541.
- [204] M. Birkinshaw, 1990, in: N. Mandolesi and N. Vittorio, eds., The Cosmic Microwave Background: 25 Years Later (Kluwer, Dordrecht) p. 77.
- [205] M. Birkinshaw, J.P. Hughes and K.A. Arnaud, 1991, Ap. J. **379**, 466.
- [206] M. Birkinshaw and J.P. Hughes, 1993, Ap. J..
- [207] T. Herbig, C.R. Lawrence, A.C.S. Readhead and S. Gulkis, 1995 Ap. J. Lett. **449**, 5.
- [208] S.T. Myers et al., 1997 Ap. J. **485**, 1.
- [209] M. Jones et al., 1993, Nature **365**, 320; K. Grainge et al., 1994, M.N.R.A.S., **265**, L57.
- [210] T.M. Wilbanks, P.A.R. Ade, M.L. Fischer, W.L. Holzappel and A.E. Lange, 1994, Ap. J. Lett. **427**, 75.

- [211] H. Liang and R. Ekers 1995, Publ. Astron. Soc. Australia **12**, 123.
- [212] J.E. Carlstrom, M. Joy and L. Greggo, 1996 Ap. J. Lett. **456**, 75.
- [213] G. Evrard, 1990, Ap. J. **363**, 349.
- [214] J.R. Bond and J. Wadsley, 1994, unpublished.
- [215] G. Efstathiou and J.R. Bond, 1986, M.N.R.A.S. **218**, 103.
- [216] G. Efstathiou and J.R. Bond, 1987, M.N.R.A.S. **227**, 33P.
- [217] N. Kaiser, 1984, Ap. J. **282**, 374.
- [218] G. Efstathiou, 1988, in: G. Coyne and V. Rubin, eds., *Large Scale Motions in the Universe*, Proc. Pontifical Academy of Sciences Study Week # 27 (Princeton University Press, Princeton).
- [219] W. Hu, E. Bunn and N. Sugiyama, 1995, Ap. J. Lett. **447**, 59.
- [220] S. Dodelson and J. Jubas, 1995, Ap. J. **439**, 503.
- [221] W. Hu, D. Scott, and J. Silk 1994, Phys. Rev. D **49**, 648.
- [222] J.R. Bond, 1990, in: N. Mandolesi and N. Vittorio, eds., *The Cosmic Microwave Background: 25 Years Later*, Proceedings of the L'Aquila Conference (Kluwer, Dordrecht) p. 45.
- [223] N. Kaiser and A. Stebbins, 1984, Nature **310**, 391.
- [224] A.A. Starobinsky, 1979, Pis'ma Zh. Eksp. Teor. Fiz. **30**, 719; 1985, Sov. Astron. Lett. **11**, 133.
- [225] L.F. Abbott and M. Wise, 1984, Nucl. Phys. B **244**, 541.
- [226] D.S. Salopek, 1992, Phys. Rev. Lett. **69**, 3602.
- [227] T. Souradeep and V. Sahni, 1992, Mod. Phys. Lett. **A7**, 3541.
- [228] R.L. Davis et al., 1992, Phys. Rev. Lett. **69**, 1856.
- [229] L. Krauss and M. White, 1992, Phys. Rev. Lett. **69**, 869; F. Lucchin, S. Matarrese and S. Mollerach, 1992, Ap. J. Lett. **401**, 49; A. Liddle and D. Lyth, 1992, Phys. Lett. B **291**, 391; J.E. Lidsey and P. Coles, 1992, M.N.R.A.S. **258**, 57P.
- [230] A.G. Polnarev, 1985, Sov. Astron. **29**, 607.
- [231] G. Efstathiou, J.R. Bond and S.D.M. White, 1992, M.N.R.A.S. **258**, 1P.
- [232] J.M. Bardeen, J.R. Bond, N. Kaiser and A.S. Szalay, 1986, Ap. J. **304**, 15.

- [233] J.M. Bardeen, J.R. Bond and G. Efstathiou, 1987, Ap. J. **321**, 28.
- [234] E. Bertschinger et al., 1990, Ap. J. **364**, 370.
- [235] N. Kaiser et al., 1991, M.N.R.A.S. **252**, 1.
- [236] K.B. Fisher et al., 1992 Ap. J. **402**, 42.
- [237] M.S. Vogeley et al., 1992, Ap. J. Lett. **391**, L5.
- [238] G.D. Dalton et al., 1992, Ap. J. Lett. **390**, L1.
- [239] R.C. Nichol et al., 1992, M.N.R.A.S. **255**, 21p.
- [240] C.M. Baugh and G. Efstathiou, 1993, M.N.R.A.S. **265**, 145.
- [241] S.J. Maddox, G. Efstathiou and W.J. Sutherland, 1990, M.N.R.A.S. **246**, 433.
- [242] C.A. Collins, R.C. Nichol and S.L. Lumsden, 1992, M.N.R.A.S. **254**, 295.
- [243] J.R. Bond, G. Efstathiou, P.M. Lubin and P. Meinhold, 1991, Phys. Rev. Lett. **66**, 2179.
- [244] M. Kamionkowski, B. Ratra, D.N. Spergel and N. Sugiyama, 1994, Ap. J. Lett. **426**, 57.
- [245] B. Ratra and P.J.E. Peebles, 1994, Ap. J. Lett. 432, L5; Phys. Rev. D **50**, 5232; Phys. Rev. D, in press (PUPT-1444).
- [246] J. Yokoyama et al., 1992, Ap. J. Lett. **396**, L13.
- [247] P.J.E. Peebles, 1983, Ap. J. Lett. **263**, L1.
- [248] P.J.E. Peebles, 1987, Ap. J. Lett. **277**, L1.
- [249] R. Cen, N. Gnedin, L. Kofman and J. Ostriker, 1992, Ap. J. **413**, 1.
- [250] J.A. Peacock and S.J. Dodds, 1994, M.N.R.A.S. **267**, 1020.
- [251] N. Sugiyama, 1995, Ap. J. Supp. **100**, 281.
- [252] J.R. Bond and G. Efstathiou, 1991, Phys. Lett. B **379**, 440.
- [253] J.A. Holtzman, 1989, Ap. J. Supp. **71**, 1.
- [254] L. Kofman, N. Gnedin and N. Bahcall, 1993, Ap. J. **413**, 1.
- [255] D.Yu. Pogosyan and A.A. Starobinsky, 1993, M.N.R.A.S. **265**, 507; 1995 Ap. J. **447**, 465.
- [256] E.L. Wright et al., 1994, Ap. J. **436**, 443.

- [257] K. Gorski et al., 1994, Ap. J. Lett. **430**, L89.
- [258] U. Fano, 1954, Phys. Rev. **93**, 121.
- [259] R. Durrer, 1989, Astron. Ap. **208**, 1.
- [260] C.-P. Ma and E. Bertschinger, 1995, Ap. J. **455**, 7.
- [261] J.R. Bond and Y. Lithwick 1995, unpublished.
- [262] S. Dodelson, E. Gates, and A. Stebbins, 1996 Ap. J. **467**, 10.
- [263] A. deShalit and H. Feshbach, 1974, Theoretical Nuclear Physics, Volume 1: Nuclear Structure (Wiley, New York).
- [264] S. Weinberg, 1972, Gravitation and Cosmology: Principles and Applications of the General Theory of Relativity (Wiley, New York).
- [265] S. Weinberg, 1971, Ap. J. **168**, 175.
- [266] N. Kaiser, 1983, M.N.R.A.S. **202**, 1169.
- [267] W.H Press and E.T. Vishniac, 1980, Ap. J. **236**, 323.
- [268] S.A. Bonometto, F. Lucchin and R. Valdarnini, 1984, Astron. Ap. **140**, 27.
- [269] A.G. Doroshkevich, 1988, Pis'ma Zh. Eksp. Teor. Fiz. **14**, 296; F. Atrio-barandela, A.G. Doroshkevich and A.A. Klypin, 1991 Ap. J. **378**, 1.
- [270] A.A. Starobinsky, 1988, Pis'ma Zh. Eksp. Teor. Fiz. **14**, 394.
- [271] W. Hu and N. Sugiyama, 1995, Phys. Rev. D **51**, 2599.
- [272] P. Schneider, J. Ehlers and E.E. Falco, 1992, Gravitational Lenses (Springer Verlag, New York).
- [273] N. Kaiser, 1992, Ap. J. **388**, 272; 1992, in: *New Insights of the Universe*, Proc. Valencia Summer School, Sept. 1991.
- [274] A. Blanchard and J. Schneider, 1987, Astron. Ap. **184**, 1.
- [275] S. Cole and G. Efstathiou, 1989, M.N.R.A.S. **239**, 195.
- [276] K. Tomita and K. Watanabe, 1989, Prog. Theor. Phys. **82**, 563.
- [277] M. Sasaki, 1989, M.N.R.A.S. **240**, 415.
- [278] V.E. Linder, 1990, M.N.R.A.S. **243**, 362.
- [279] L. Cayon, E. Martinez-Gonzalez and J.L. Sanz, 1993, Ap. J. **403**, 471; **413**, 10.

- [280] U. Seljak, 1996, Ap. J. **463**, 1.
- [281] M.B. Hindmarsh and T.W.B. Kibble, 1994, Cosmic Strings, SUSX-TP-94-74, IMPERIAL /TP/94-95/5, hep-ph/9411342.
- [282] D.P. Bennett, A. Stebbins and F.R. Bouchet, 1992, Ap. J. Lett. **399**, 5.
- [283] B. Allen, R.R. Caldwell, E.P.S. Shellard, A. Stebbins and S. Veeraraghavan, 1996, Phys. Rev. Lett. **77**, 3061.
- [284] D.P. Bennett and S.H. Rhie, 1993, Ap. J. Lett. **406**, 7.
- [285] Ue-Li Pen, D.N. Spergel and N. Turok, 1994, Phys. Rev. D **49**, 692.
- [286] D. Coulson, P. Ferreira, P. Graham and N. Turok, 1994, Nature **368**, 27.
- [287] Ue-Li Pen and D.N. Spergel, 1994, Phys. Rev. D, in press.
- [288] M. Tegmark and E.F. Bunn 1995, Ap. J. **455**, 1.
- [289] M. Tegmark **1996**, M.N.R.A.S. in press.
- [290] E.F. Bunn, D. Scott and M. White, 1995, Ap. J. Lett. **441**, L9; M. White and E.F. Bunn, 1996, Ap. J. **460** 1071.
- [291] R. Stompor, K.M. Gorski and A.J. Banday, 1995 M.N.R.A.S. **277**, 1225.
- [292] K.M. Gorski, B. Ratra, N. Sugiyama and A.J. Banday, 1995, Ap. J. Lett. **446**, 67.
- [293] S.D.M. White, G. Efstathiou, and C.S. Frenk, 1995, M.N.R.A.S. **292**, 371
- [294] V.R. Eke, S. Cole, and C.S. Frenk, 1996, M.N.R.A.S. **282**, 263.
- [295] M. Strauss, and J. Willick, J. 1995, *Phys. Rep.* **261**, 271
- [296] J.A. Willick et al.1996, Ap. J.**457**, 460.
- [297] S. Zaroubi et al.1996, preprint, astro-ph/9603068
- [298] T. Kolatt and A. Dekel 1996, preprint, astro-ph/9512132
- [299] S. Zaroubi, I. Zehavi, A. Dekel, Y. Hoffmann and T. Kolatt 1996, preprint, astro-ph/9610226
- [300] Bertschinger, E. 1996, in “Cosmology and Large Scale Structure”, Les Houches Session LX, August 1993, ed. R. Schaeffer, Elsevier Science Press

- [301] E. Bertschinger, P. Bode, J.R. Bond, D. Coulson, R. Crittenden, S. Dodelson, G. Efstathiou, K. Gorski, W. Hu, L. Knox, Y. Lithwick, D. Scott, U. Seljak, A. Stebbins, P. Steinhardt, R. Stompor, T. Souradeep, N. Sugiyama, N. Turok, N. Vittorio, M. White, M. Zaldarriaga 1995, ITP workshop on *Cosmic Radiation Backgrounds and the Formation of Galaxies*, Santa Barbara [COMBA]
- [302] W. Hu, D. Scott, N. Sugiyama and M. White, 1995, Phys. Rev. D **52**, 5498.
- [303] L. Abbott and R.K. Schaeffer, 1985, Ap. J. **308**, 546.
- [304] M. White and D. Scott, 1996, Ap. J. **459**, 415.
- [305] J.R. Bond and T. Souradeep, 1996, preprint.
- [306] U. Seljak and M. Zaldarriaga, 1996, Ap. J. **469**, 437.